European Commission

Study to support an Impact Assessment on enhancing the use of data in Europe

Digital Strategy

Internal identification

Contract number: SMART 2019/0024

VIGIE number: 2020-0694

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology
Directorate G — Data
Unit G.1 — Data Policy & Innovation

Contact: cnect-g1@ec.europa.eu

European Commission

B-1049 Brussels

Study to support an Impact Assessment on enhancing the use of data in Europe

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. The Commission does not guarantee the accuracy of the data included in this study. More information on the European Union is available on the Internet (http://www.europa.eu).

EN PDF   ISBN 978-92-76-41909-9   doi:10.2759/759296   KK-01-21-277-EN-N

The European Commission is not liable for any consequence stemming from the reuse of this publication.

Luxembourg: Publications Office of the European Union, 2022

This study has been carried out for DG CNECT by:

Deloitte.

the Lisbon council
think tank for the 21st century

JIIP
Joint Institute for Innovation Policy

GOVLAB

TIMELEX

open data institute ODI

Authors:

Sebastiaan van der Peijl (Deloitte)

Emily Denny (Deloitte)

Eike-Christian Koring (Deloitte)

Angeliki Papadimitriou (Deloitte)

Franziska Hoerth (Deloitte)

Cleónice León Vargas (Deloitte)

Marc Nikolov (Deloitte)

Marie Eichholtzer (Deloitte)

David Osimo (The Lisbon Council)

Cristina Moise (The Lisbon Council)

Claire Stolwijk (The Joint Institute for Innovation Policy)

Stefaan Verhulst (The GovLab)

Hans Graux (Timelex)

Mahlet Zimeta (The Open Data Institute)

Alan Walker (The Open Data Institute)

Contents

Abstract

In line with the European Commission's communication on the **European strategy for data** of 2020 (COM(2020) 66 final), the following study assesses the key domains that fall under the concern and potential scope of action of the Data Governance Acts and the Data Act.

The **Data Governance Act** outlines four key issues to be tackled, namely: access and reuse of sensitive public-sector data; certification/authorisation schemes for "data altruism"; data sharing through metadata standards across or within sectors; and, certification framework for European data intermediaries or data marketplaces to enable data demand and supply. On the other hand, the **Data Act** focuses on four different issues, namely: 1) Business-to-Government Data Sharing (B2G) for the public interest; 2) citizen empowerment ('human-centric data economy') in the context of data generated through devices; 3) rights to co-generated data in a Business-to-Business (BSB) context; and, 4) conflicts that companies face due to different laws at the international level.

For each of these issues, the study **explores the state of play in Europe and determines the impact of a number of possible policy options** acting as stepping-stones enabling relevant stakeholders to build common data spaces and fully realise the benefits of increased data governance and data sharing and reuse. Following this, a **Multi-Criteria Analysis was performed to determine the preferred policy option** for each domain, analysing the effectiveness, efficiency, coherence and legal and political feasibility and proportionality of each option. Lastly, the study identified three **policy packages** (i.e. sets of policy options) to analyse impacts from a macroeconomic standpoint vis-à-vis the baseline.

The **overall results** of the study point out that for the **measures to enhance data governance** the preferred policy options are low-intensity regulatory options for the domains on access and reuse of sensitive public sector data; data sharing through the establishment of metadata standards across or within sectors; and, certification framework for European data intermediaries or data marketplaces to enable data demand and supply. A higher-intensity regulatory option is preferred for the certification/authorisation schemes for "data altruism". Both a top-down analysis of the policy packages and a bottom-up analysis based on the cost-benefit results of the policy options were performed. They found that by 2028, the value of the data economy could increase from EUR 533.51 billion (in the absence of EU action) to EUR 540.73 billion – 544.43 billion with the mixed regulatory intervention (representing from 3.87% to between 3.92% and 3.95% of the GDP).

Moreover, the **overall results for the measures to foster data re-use** point out that the low-intensity regulatory option is preferred for the measures on B2G data sharing, as well as for the measures developing rights on co-generated data and B2B data sharing. As regards measures supporting citizen empowerment, in the context of fitness trackers, a low-intensity regulatory option is also preferred, while in the context of smart home appliances, a soft option (non-regulatory intervention) is preferred. For the measures supporting companies in cases of conflict of laws and international level, the preferred option is the higher-intensity regulatory option. The assessment of impacts concludes that in 2028, the economic impact of the policy measures as compared to the baseline scenario could imply an increase in GDP with 273 billion EUR in case a mix of the preferred options is implemented (representing an additional 1.98% of GDP).

Abstrait

Conformément à la communication de la Commission européenne sur la **Stratégie européenne pour les données** de 2020 (COM(2020) 66 final), l'étude suivante évalue les domaines clés qui relèvent de la préoccupation et du champ d'action potentiel des lois sur la gouvernance des données et de la loi sur les données.

Le **Loi sur la gouvernance des données** décrit quatre questions clés à aborder, à savoir: l'accès et la réutilisation des données sensibles du secteur public; les régimes de certification/autorisation pour « l'altruisme des données »; le partage de données au moyen de normes de métadonnées entre les secteurs ou au sein de ceux-là; et un cadre de certification pour les intermédiaires européens de données ou les marchés de données afin de permettre la demande et l'offre de données. D'autre part, le **Loi sur les données** se concentre sur quatre questions différentes, à savoir: 1) Partage de données interentreprises (B2G) pour l'intérêt public; 2) l'autonomisation des citoyens (« économie des données centrée sur l'humain ») dans le contexte des données générées par les appareils; 3) droits sur les données co-générées dans un contexte Business-to-Business (BSB); et 4) les conflits auxquels les entreprises sont confrontées en raison de lois différentes au niveau international.

Pour chacune de ces questions, l'étude **explore l'état d'avancement des activités en Europe et détermine l'impact d'un certain nombre d'options politiques possibles** servir de tremplins permettant aux parties prenantes concernées de créer des espaces de données communs et de tirer pleinement parti des avantages d'une gouvernance accrue des données, du partage et de la réutilisation des données. Par la suite, un **Une analyse multicritère a été effectuée pour déterminer l'option de stratégie préférée** pour chaque domaine, analyser l'efficacité, l'efficience, la cohérence, la faisabilité juridique et politique et la proportionnalité de chaque option. Enfin, l'étude a identifié trois **ensembles de stratégies** (c'est-à-dire des ensembles d'options politiques) pour analyser les impacts d'un point de vue macroéconomique par rapport au niveau de référence.

Les **résultats globaux** de l'étude soulignent que pour le **mesures visant à améliorer la gouvernance des données** les options stratégiques privilégiées sont des options réglementaires de faible intensité pour les domaines de l'accès et de la réutilisation des données sensibles du secteur public; le partage de données par l'établissement de normes de métadonnées entre secteurs ou au sein de ceux-là; et un cadre de certification pour les intermédiaires européens de données ou les marchés de données afin de permettre la demande et l'offre de données. Une option réglementaire de plus haute intensité est préférable pour les régimes de certification/autorisation pour « l'altruisme des données ». Une analyse descendante des ensembles de politiques et une analyse ascendante fondée sur les résultats coûts-avantages des options stratégiques ont été effectuées. Ils ont constaté que d'ici 2028, la valeur de l'économie des données pourrait passer de 533,51 milliards d'euros (en l'absence d'action de l'UE) à 540,73 milliards - 544,43 milliards d'euros avec l'intervention réglementaire mixte (représentant de 3,87 % à entre 3,92 % et 3,95 % du PIB).

De plus, les **résultats globaux des mesures visant à favoriser la réutilisation des données** soulignent que l'option réglementaire de faible intensité est préférable pour les mesures sur le partage de données B2G, ainsi que pour les mesures développant les droits sur les données co-générées et le partage de données B2B. En ce qui concerne les mesures soutenant l'autonomisation des citoyens, dans le contexte des trackers de fitness, une option réglementaire de faible intensité est également préférée, tandis que dans le contexte des appareils électroménagers intelligents, une option douce (intervention non réglementaire) est préférée. Pour les mesures de soutien aux

entreprises en cas de conflit de lois et au niveau international, l’option privilégiée est l’option réglementaire de plus haute intensité. L’évaluation des impacts conclut qu’en 2028, l’impact économique des mesures politiques par rapport au scénario de référence pourrait impliquer une augmentation du PIB avec 273 milliards d’euros en cas de mise en œuvre d’une combinaison des options privilégiées (représentant 1.98 % du PIB supplémentaire).

1 Introduction

This chapter illustrates the purpose of the document and briefly explains which data collection tools were used to gather the evidence underpinning the findings and conclusions of this assignment.

1.1 Purpose of the document

The purpose of this report is to share the **final results of the Study to support an Impact Assessment on enhancing the use of data in Europe**.

The document is structured as follows:

- Chapter 1: Introduction on the purpose and the structure of the document
- Chapter 2: Measures to enhance data governance
- Chapter 3: Measures to foster data sharing and re-use
- Chapter 4: Annexes

Part 1 of this assignment comprises four separate domains:

- Measures facilitating secondary use of sensitive data held by the public sector;
- Establishing a certification scheme for data altruism mechanisms;
- Establishing a European structure for governance aspects of data sharing; and
- Establishing a certification framework for data intermediaries.

Part 2 of this assignment comprises four separate domains:

- Business-to-Government (B2G) data sharing for the public interest;
- Measures supporting citizen empowerment (‘human-centric data economy’);
- Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing; and
- Measures supporting companies in cases of conflict of laws at international level.

These chapters are then followed by the below annexes:

- **Annex I** – Measures to enhance data governance: includes the CBA, case studies, one pagers, macroeconomic analysis and references.
- **Annex II** – Measures to foster data sharing and re-use: includes the CBA, Legal analysis of Business-to-government data sharing, case studies, additional information on Withings and Green Button initiative as well as the reciprocity clause in Australia Consumer Data Rights.

1.2 Scope of this study

As mentioned in the Executive Summary, this study sustains the development of policy measures concerning the areas defined by the Communication on a European Strategy for data[1].

Part 1 of the study covers the following topics**:**

1.1. The question of access and **reuse of sensitive public sector data** which are currently not disclosed by public sector bodies and not covered by the Public Sector Information (PSI)/Open data directive[2] (e.g. health data, statistical microdata, company ownership data, microdata from public transport systems and others)[3].

1.2. The possibility of **establishing "data altruism" schemes** in Europe, defined as means of making data available (whether anonymised or non-anonymised) without expecting anything (not even services) in exchange.

1.3. The question of facilitating data sharing through the establishment of **metadata standards** across or within sectors and including both technical and legal standards.

1.4. The relevance of building a **certification framework** for European data intermediaries or data marketplaces which help data demand and supply to match through independent platforms.

From a stakeholder perspective, the study focuses on the relevant stakeholders in the data value chain for each of the topics in scope, meaning on data holders, data intermediaries and data re-users. The table below summarises the main categories of stakeholders involved in the domains' data collection and analysis activities conducted as part of Part 1.

Table 1 - Stakeholder scope for Part 1 (data value chain mapping)

Domain	Data holder	Data re-user	Intermediaries	Personal data?	Purpose
Measures facilitating secondary use of sensitive data held by the public sector	Public sector authorities (e.g. Health institutions, transport authorities, statistical offices)	Public sector authorities, researchers and businesses	Public sector authorities, research organisations, non-for-profit orgs.	Yes (and sensitive)	Research and innovation, public health, increased efficiency
Establishing a certification/authorisation scheme for data altruism mechanisms	Public sector authorities, businesses, NGOs and researchers	Public sector bodies, researchers and non-for profit orgs.	Public sector authorities, businesses, research orgs.	Yes (and sensitive)	Research, innovation, public health and other societal benefits
Establishing a European structure for governance aspects of data sharing	Businesses in traditional sector	Other businesses and researchers from various sectors	Public and private orgs. in charge of data spaces; standardisation initiatives	N/A	Innovation, competitiveness

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM/2020/66 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1582551099377&uri=CELEX:52020DC0066

[2] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[3] In agreement with the Commission, this study focuses on the former two.

Establishing a certification framework for data intermediaries - *Generic Approach*	Businesses, Academia and research orgs., Governmental orgs., NGOs, Citizens	Businesses, Academia and research orgs., Governmental orgs., NGOs, Citizens	Certified data Intermediaries (i.e. Data marketplaces, data brokers, data repositories, PIMS/PDS, industrial data platforms, trusted third parties, data unions, data cooperatives, data collaboratives, data trusts)	Potentially	Business, R&I, Public Good
Establishing a certification framework for data intermediaries - *B2B Approach*	Businesses	Businesses	Certified data intermediaries: Data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts	No	Business, R&I
Establishing a certification framework for data intermediaries - *C2B Approach*	Citizens	Businesses	Certified data intermediaries: PIMS/PDS, data unions, data cooperatives, data collaboratives, data trusts	Yes	Business, R&I, Public Good

Part 2 of the study covers the following topics**:**

- Aspects related to **Business to Government Data Sharing (B2G) for the public interest** (i.e. for the development of better policies and delivery of better public services).
- Possibilities for **empowering citizens** and putting them even more in control of their data, building on the General Data Protection Regulation[4] and establishing a human centric data economy.
- The question of **rights and control over co-generated data** (i.e. in the context of connected and Internet of Things devices) for enabling further business to business (B2B) data sharing.
- Aspects related to **conflict of laws at the international level** and possible obstacles for businesses subject to extra-territorial provisions and foreign jurisdictions.

From a stakeholder perspective, the study focuses on the relevant stakeholders in the data value chain for each of the topics in scope, meaning on data holders, data intermediaries and data re-users. The table below summarises the main categories of stakeholders involved in the domains' data collection and analysis activities conducted as part of Part 2.

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Table 2 - Stakeholder scope for Part 2 (data value chain mapping)

Domain	Data holder	Data (co-) producers	Data re-user (whole dataset)	Data re-user (individual data portability)	Intermediaries	Personal data?
Business-to-Government data sharing for the public interest	Private sector organisations	N/A	Public sector: national executive government (e.g. statistical offices), regional and local government (e.g. municipalities), legislative branch (e.g. parliamentary research services).			Sometimes
Measures supporting citizen empowerment ('human-centric data economy') – Fitness	Device producer	Owner of device (individual)	Researchers Platforms (Google Apple Strava)	App developers Insurance companies Health providers Other device producers	Platforms (Google Apple Strava) PIMS	Yes
Measures supporting citizen empowerment ('human-centric data economy') – Smart home	Producer (Electrolux) Energy companies	Owner of device (family)	Platforms (Google Amazon Apple Samsung IFTTT) Energy companies	Repair shop App developers Insurance companies Platforms Other device producers	Platforms (Google Amazon Apple IFTTT) Produced led platforms (Schneider, Johnson, Siemens, Samsung, Philips) PIMS	Yes
Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing	Private sector companies: IoT Product/ service providers (i.e. OEMs, smart machine or connected vehicle manufacturers)	Private Sector Companies: IoT product/ service users (i.e. farmers, construction companies)	Private Sector Companies: independent service providers (i.e. data analytics companies, data platforms, competitors)		B2B Data Intermediaries (i.e. data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts)	Sometimes

Measures supporting companies in cases of conflict of laws at international level	ICT service providers and their customers	Complementary service providers – data intelligence and analytics	Public sector bodies (law enforcement, national security)	ICT service providers (often but not exclusively cloud based)	Yes (though not exclusively)

1.3 Methodology for the assignment

For **Part 1,** the evidence supporting this analysis comes from a number of different sources:

- Stakeholder mapping;
- Interviews;
- Legal analysis (Measures facilitating secondary use of sensitive data held by the public sector);
- Workshops (for domains: Measures facilitating secondary use of sensitive data held by the public sector and Establishing a European structure for governance aspects of data sharing);
- Case studies (for domains: Measures facilitating secondary use of sensitive data held by the public sector and Establishing a European structure for governance aspects of data sharing);
- Market analysis (Establishing a certification framework for data intermediaries).

The first step of the assignment consisted of **mapping stakeholders** based on desk research to understand the current data economy stakeholders relevant for each of the topics under the four domains. This analysis contributed to the development of the stakeholder map and potential interviewee list for further data collection.

The team carried out **interviews** with public sector entities and private sector organisations to collect primary data for this study. These interviews covered various industry sectors and Member States. The purpose of the interviews was to reach out to European companies and Member States to collect data on the data economy and on the related costs and effects.

Additional research methodologies where used for the various domains relevant for their respective topic. For Measures facilitating secondary use of sensitive data held by the public sector and Establishing a European structure for governance aspects of data sharing, **case studies** and **workshops** were conducted, a **legal analysis** for Establishing a certification scheme for data altruism mechanisms and a **market analysis** for Establishing a certification framework for data intermediaries .

Workshops were conducted to measure facilitating secondary use of data the use of which is subject to the rights of others and for the purpose of reviewing the relevance of establishing a European structure for governance aspects of data sharing with stakeholders. The aim of these workshops was to:

- Discuss the qualitative and quantitative assumptions and findings with regard to the baseline scenario and policy options, and discuss the policy options themselves (Measures facilitating secondary use of sensitive data held by the public sector); and
- Discuss the policy options and validated the consolidated data on costs and benefits (Establishing a European structure for governance aspects of data sharing).

Case studies were also conducted for in-depth and detailed investigation to understand better the state of play or baseline scenario for the topic at hand. The case studies were developed based on desk research and interviews with stakeholders from the public and private sector. Each case study built on the data coming from the stakeholder mapping and provided insights for the baseline scenarios and make hypothesis on the impact of the different policy options.

The **legal analysis** was based on desk research and interviews to understand the current legal status with regard to data altruism and what barriers this might cause for data altruism. The analysis supported the definition of the policy options and overall analysis.

In addition, a **market analysis** was carried out to understand the business environment and data based value chains as well as to identify the key players and key positions in the markets. The market analysis focused on data collection through desk research, triangulation of data and an analysis of the data to answer questions such as what are the main data intermediaries on the EU market. The desk research relied on databases and data marketplaces such as Forrester Research, Gartner Research, IDC, Economist intelligence unit and EMIS Intelligence.

For **Part 2,** the evidence supporting this analysis comes from a number of different sources:

- Stakeholder mapping (All domains);
- Interviews (for domains: Business-to-Government (B2G) data sharing for the public interest, measures supporting citizen empowerment ('human-centric data economy'), and measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing);
- Targeted questionnaire (measures supporting companies in cases of conflict of laws at international level);
- Legal analysis (for domains: Business-to-Government (B2G) data sharing for the public interest, measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing and measures supporting companies in cases of conflict of laws at international level);
- Workshops (for domains: Business-to-Government (B2G) data sharing for the public interest and measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing);
- Case studies (measures supporting citizen empowerment ('human-centric data economy'));
- Market analysis (measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing).

The first step of the assignment consisted of **mapping stakeholders** based on desk research to understand the current data economy stakeholders relevant for each of the topics under the four domains. This analysis contributed to the development of the stakeholder map and potential interviewee list for further data collection.

The team then carried out **interviews** with public sector entities and private sector organisations to collect primary data. These interviews covered various industry sectors and Member States. The purpose of the interviews was to reach out to European companies and Member States to collect data on the data economy and on the related costs and effects.

Additional research methodologies were used for the various domains relevant for their respective topic such as **legal analysis**, **market analysis** , **workshops** and **case studies**. These additional methodologies are described in further detail below.

Workshops were conducted in order to validate interim findings and collect additional data on Business-to-Government Data Sharing (B2G) for the public interest and citizen empowerment and control over data. The aim of these workshops was to discuss the qualitative and quantitative assumptions and findings with regard to the baseline scenario and policy options, and discuss the policy options themselves while also validating the collected data on costs and benefits.

Case studies were also conducted for in-depth and detailed investigation to understand better the state of play or baseline scenario for the topic at hand. The case studies were developed based on desk research and interviews with stakeholders from the public and private sector. Each case study built on the data coming from the stakeholder mapping and provided insights for the baseline scenarios and made hypotheses on the impact of the different policy options.

The **legal analysis** was based on desk research and interviews to understand the current legal status and supported the definition of the policy options and overall analysis.

In addition, a **market analysis** was carried out to understand the business environment and data-based value chains as well as to identify the key players and key positions in the markets. The market analysis focused on data collection through desk research, literature review and interviews, triangulation of data and an analysis of the data to answer questions related to the current state of play of the market and potential market failures linked to the lack of clarity on access and usage rights on co-generated data.

Limitations relating to the findings of this study

As part of this study, evidence was gathered from various sources, including desk research, interviews with businesses and other stakeholders, case studies, market analysis and legal analysis.

The data collection was hampered by the fact that the public and private sector are still relatively new to navigating the data economy and can only share insights into for example costs and benefits to a very limited extent.

This situation poses challenges on the findings of this study. While collect qualitative feedback from the public and private sector was collected on the different policy interventions discussed for each domain, it was more difficult to quantify their costs and benefits, e.g. because case numbers are still small or the data sharing practices are just emerging and stakeholders themselves do not yet know their scale and/or costs of making data available. In addition, the stakeholders consulted do not yet have a final and consolidated perception on for example the potential benefits they could draw from increased data use and availabilities in their respective domain, besides speculative thoughts.

This report should be considered as a first attempt at examining this topic and gathering the existing data on these subjects. This analysis is therefore based on the limited data available and provides a preliminary (mainly qualitative) overview of the costs and benefits for the different topics under scrutiny. The conclusions reached are based on independent judgement and specific to this study.

2 Measures to enhance data governance

This chapter provides the assessment of key issues identified as part of the challenge to enhance data governance in the EU. The problems, its causes and effects are explored, based upon which the policy objectives and options are set out to address these. These options are then assessed along five main criteria as part of a multi-criteria analysis to determine the preferred option in four key areas. Finally, the macro-economic impacts are derived.

2.1 Background and problem assessment

This section contains the problem assessment of issues related to the secondary use of data held by the public sector and the use of which is subject to the rights of others, data altruism, the governance of data sharing, and the role of data intermediaries.

Market developments and policy initiatives of the past decade have set the ground for a European data strategy that could enable the EU to become the world's most attractive, secure and dynamic data-agile economy, improving the lives of its citizens. Europe's technological and digital future depends on whether it seizes this opportunity. Thus, despite the action that the European Commission has taken so far, remaining issues need to be tackled for Europe to reap fully the data economy's benefits.

Data has started to disrupt European economies and markets. The European data market's value will reach EUR 77.8 billion in 2020 employing 8.25 million people, and the overall value of the data economy grew from EUR 247 billion in 2013 to almost EUR 477.3 billion in 2020, worth about 3.2 per cent of total EU GDP.[5] The European Monitoring Tool further predicts that by 2025, the value of the EU data market could reach EUR 93 to 141.6 billion. Likewise, the EU data economy is expected to increase to EUR 1,053 billion with an overall impact of 6.3% on the EU GDP under a high growth scenario. The data suppliers industry would increase from 255,000 companies in 2016 to 294,350 in 2020, and the number of data workers in Europe would increase up to 8.25 million by 2020.

This market development is in line with the perception of businesses. Economic growth and a higher level of competition and innovation in the EU were the key benefits and opportunities identified with regard to the European data economy and data mobility within the EU.[6] Many companies have recognised the potential of data-driven innovation and started to share and re-use data among them to enhance their business opportunities and improve internal efficiency. According to a 2018 study on Business to Business (B2B) data sharing and re-use[7], this trend would grow significantly in the

[5] European Data Market Monitoring Tool, consulted on 7 July 2020, see: http://datalandscape.eu/european-data-market-monitoring-tool-2018

[6] European Commission (2017), Synopsis report: consultation on the 'building a European data economy' initiative, https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-building-european-data-economy

[7] European Commission (2018), Study on data sharing between companies in Europe, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

following five years, while companies not critically investing in data may be missing business opportunities. In addition, a large proportion of SMEs perceive data sharing as important and actively acquire data from other companies.[8] Data also feeds into other new technologies with the potential to foster European economies, such as Artificial Intelligence and the Internet of Things.

Apart from its economic impact, the data transformation will affect European societies and daily lives. The volume of data produced in the world is growing rapidly, from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025, and the amount of data is doubling every 18 months. Data-driven innovation will bring substantial benefits for citizens, for instance through personalisation and enhancements in healthcare and well-being, transport, transparent governance, public services, energy consumption, product, material and food traceability, and could even contribute to the successful implementation of the European Green Deal. While identifying these benefits, the European strategy for data[9] emphasises that transformations could put European core values at risk. To ensure an open, fair, diverse, democratic and confident Europe, it needs to accompany the wide flow and use of data with high levels of privacy, security, safety and ethical standards.

2.1.1 Measures facilitating secondary use of sensitive data held by the public sector

2.1.1.1 Background

2.1.1.1.1 Context

In recent years, there has been a growing trend towards open data – making public data available for reuse by the private sector, civil society and academics in order to enable research or help promote the development of new services. The EU Open Data Directive[10] is one manifestation of this global trend, backed by estimates that the opening up of EU public data could drive economic benefits of EUR 250 billion[11].

The Open Data Directive extended the scope of previous legislation, requiring Member States to make unprecedented amounts of data available for reuse. The new rules recognise, however, that some data held by the public sector is not suitable for entirely open access and reuse. Instead a number of exceptions are carved out, including for personal data, "documents for which third parties hold property rights", and other "sensitive data" protected by national legislation on the grounds including national security, statistical confidentiality, and commercial confidentiality.

Any of these types of data can be categorised as sensitive data in that additional steps are required before it is possible to share them publicly. Indeed, this domain defined 'sensitive data held by the public sector' as "**data the use of which is subject to the conflicting rights of others**". For these reasons, this type of data is less likely to be made available by public administrations. However, in line with the high estimate of the value of public sector data, there is great potential to drive positive economic and social results through opening up some of this data. Health and social data, for example, would generally be classed as "sensitive" data under this categorisation as it is subject to the rights of patients having co-produced the data, and could be used for ends including to

[8] European Commission (2019), SME panel consultation on B2B data-sharing principles and guidance – Report on the results, https://ec.europa.eu/digital-single-market/en/news/sme-panel-consultation-b2b-data-sharing

[9] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, 2020, https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[10] DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2019 on open data and the re-use of public sector information. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

[11] Deloitte, Open Evidence, Wik Consult, timelex, Spark, The Lisbon Council (2018), Study to support the review of Directive 2003/98/EC on the re-use of public sector information. Available at https://ec.europa.eu/digital-single-market/en/news/impact-assessment-support-study-revision-public-sector-information-directive

"develop personalised medicine or advance research to find cures for specific diseases".[12] Other data the use of which is subject to the rights of others could include genetic data, and statistical microdata.[13]

At European level, there is an ongoing discourse within the research community on how data the use of which is subject to the rights of others can best be shared and made available for research purposes. European projects including EUDAT CDI[14] and EOSC-hub[15] have engaged on this issue, producing recommendations[16] on how to enable the secure sharing of this data.

The types of measures considered go well beyond the anonymisation of personal data, for which there are concerns that it will nonetheless be possible to link the data back to the individual concerned.[17] Instead, they focus on points such as:

- Promoting free access to the metadata of data the use of which is subject to the rights of others;
- Providing a "safe haven" - a secure environment for research work on data the use of which is subject to the rights of others;
- The relative merits of central vs distributed storage of data the use of which is subject to the rights of others.

2.1.1.1.2 Ecosystem

This section outlines the types of stakeholders concerned by the sharing of data held by the public sector and the use of which is subject to the rights of others. The table below provides an overview.

Table 3 - Stakeholder scope (data value chain mapping)

Domain	Data holder	Data re-user	Intermediaries	Personal data?	Purpose
Measures facilitating secondary use of sensitive data held by the public sector	Public sector authority (e.g. Health institutions, transport authorities, statistical offices)	Researchers/Public Sector Bodies/ Businesses	Public bodies, research organisations, not for profit organisations, partnerships	X (and sensitive)	Research & Innovation, Public health, increased efficiency

This study focuses on **data holder, data re-users and data intermediaries**. Data (co-) producers in this context are typically citizens and, depending on the case, employees at statistical offices

[12] European Commission (2020), Shaping Europe's digital future – Questions and Answers, Available at https://ec.europa.eu/commission/presscorner/detail/en/qanda_20_264

[13] EOSC-hub (2019), D2.8 First Data policy recommendations, Available at https://documents.egi.eu/public/RetrieveFile?docid=3419&filename=EOSC-hub%20D2.8%20v1%20Approved%20by%20EC%20Public.pdf&version=5

[14] EUDAT Collaborative Data Infrastructure. See https://eudat.eu/eudat-cdi

[15] European Open Science Cloud hub – providing support services for the development of a European Open Science Cloud, and a single point of contact for researchers for resources for advanced data-driven research. See https://www.eosc-hub.eu/about-us

[16] EOSC-hub (2019), D2.8 First Data policy recommendations, Available at https://documents.egi.eu/public/RetrieveFile?docid=3419&filename=EOSC-hub%20D2.8%20v1%20Approved%20by%20EC%20Public.pdf&version=5; EUDAT (2018), EUDAT Conference Outputs and Recommendations, Available at https://b2share.eudat.eu/records/31b4347b771641e791991578b6731aa1

[17] European Commission (2018), Synopsis Report - Consultation: Transformation Health and Care in the Digital Single Market. Available at https://ec.europa.eu/health/sites/health/files/ehealth/docs/2018_consultation_dsm_en.pdf

performing surveys or general practitioners. They are unknown to the public and should remain so. There cannot be costs or benefits for (co-) producers as part of this study.

Data holders are defined by the OECD as "*a party who, according to domestic law, is competent to decide about the contents and use of (personal and non-personal) data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf*".[18] In the context of this domain, data holders are necessarily public sector entities, such as statistical offices, health institutions or transport authorities. They hold the data that re-users want to access, and constitute the **supply side** of the sensitive public data market.

Data re-users, which can be defined as "*generating the social and economic value of data sharing*" through their use of the data the use of which is subject to the rights of others,[19] may be public bodies, researchers, or businesses, and accordingly may use data the use of which is subject to the rights of others for information, research or commercial purposes. They constitute the **demand side** of the sensitive public data market.

Lastly, **data intermediaries** are public sector entities which primarily "*enable data holders to share their data, so it can be re-used by potential data users*". These intermediaries facilitate the processes (such as data permit applications, or the process of searching for the data holder holding the desired dataset) required for re-users to obtain access to sensitive public data held by data holders.

Both **data holders and data intermediaries perform a public service function**, and therefore may not make pecuniary profits from these functions. However, they may charge for that service in order to cover the costs linked to its provision.

2.1.1.1.3 Ongoing initiatives

This section outlines a list of initiatives aiming at facilitating the reuse of data held by the public sector and the use of which is subject to the rights of others. It examines in more depth two of these – Finland's Findata and the German *Rat für Sozial- und Wirtschaftsdaten* (RatSWD).

Several data holders that already make data the use of which is subject to the rights of others available for access and re-use have been identified:

- The **United Kingdom's Department for Education** controls the **National pupil database** (NPD) containing over 21 million pupils (data stored in the NPD is never deleted).[20] Data is collected by state-funded education and higher education institutions in England only: other Devolved Administrations operate different systems.
- The UK National Health Service (NHS) **National Services Scotland (NSS) national safe haven service** – allows data from electronic records to be used to support research when it is not practicable to obtain individual patient consent, while protecting patient identity and privacy. It provides **secure** file transfer and submission services to data providers and additional services (e.g. analytics platforms) to researchers. It is currently operated by the Edinburgh Parallel Computing Centre (EPCC, University of Edinburgh), while the Farr Institute of Health Informatics Research provides the infrastructure.[21]

[18] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[19] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[20] Her Majesty's Government, National pupil database. See: https://www.gov.uk/government/collections/national-pupil-database

[21] EPCC, NHS National Services Scotland (NSS) national safe haven). See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven

- The **Belgian Federal Public Service (FPS) Mobility and Transport** launched a **Smart Mobility call** to support digital mobility initiatives to improve mobility and to boost the use of open data for mobility.
- The **38 officially accredited *Forschungsdatenzentren* – Research Data centres** – in Germany, such as the Research Data Centre of the German Federal Employment Agency, or the German Microdata Lab.[22] These have as mission to make relevant data available for research on labour market, pensions, unemployment benefits, education, vehicles, migration, copyrighted content, and others, but currently do not cover health (although this will be setup in the near future). Of particular interest is the **German Forschungsdatenzentrum** (Research Data Centre) consisting of two Research Data Centres (one at federal level and one at Länder level) that enable access to official statistics microdata to researchers.[23] This enables data from different regional statistical offices to be centralised in one storage system, thus facilitating scientific analysis. It contains data related to health, income and living conditions, agriculture, education, energy, taxation and other statistics. An online application to access the data must be submitted, and data use is subject to a fee depending on *inter alia* the amount of requested data.
- Many statistical offices across the EU, such as for instance:
- **Statistics Denmark** is Denmark's central statistical authority collecting, compiling and publishing statistics.[24] Data can be accessed for a fee covering the costs of development and operation. Research institutions may access Statistic Denmark's collection of register data and to anonymised microdata after having made a request to the Division of Research Services. In addition, Statistics Denmark provides customised services combining a range of statistical records, for a fee calculated based on the time spent on the custom request (with a defined hourly fee).
- **Statbel** is Belgium's statistical office collecting, producing and disseminating figures on the Belgian economy, society and territory.[25] These figures are available at national, regional, provincial, municipal and even more detailed level, as well as within a European context. Apart from the readily available statistics that are often used by policy makers, consumers and businesses, and researchers, pseudonymised study microdata can be made available for public institutions or research institutes, through a standardised procedure in order to comply with the privacy law.
- The **Czech Statistical Office** (CZSO) is Czechia's central statistical authority collecting, processing, and disseminating statistical information.[26] A wide range of data can be easily accessed free of charge. The use of confidential statistical data (including personal data) can only be used for scientific research purposes and must be officially requested.
- The **National Institute of Statistics** is the main producer of official statistical data for **Romania**. They are responsible for the coordination of all activities at national level regarding the development and dissemination of European statistics.[27] Direct access to the anonymised microdata is provided by means of research contracts. Access is in principle restricted to universities, research institutes, national statistical institutes, central banks inside the EU and the EEA countries, and the ECB.As regards data intermediaries, the following initiatives were identified:
- The **eHealth platform Belgium** is a Belgian federal government service that offers an electronic platform where all stakeholders involved in public health (businesses, citizens, care providers,

[22] The full list may be found on the German Data Forum's website. See: https://www.ratswd.de/en/data-infrastructure/rdc

[23] Forschungsdatenzentrum. See: https://www.forschungsdatenzentrum.de/de

[24] Statistics Denmark. See: https://www.dst.dk/en

[25] Statbel. See: https://statbel.fgov.be/en/

[26] Czech Statistical Office. See: https://www.czso.cz/csu/czso/about-czso

[27] National Institute of Statistics. See: https://insse.ro/cms/en/content/about-nis

institutions…) can exchange information, including personal data, in a safe and efficient manner.[28] It offers a range of other services, such as **MyCareNet**,[29] a platform enabling data exchange between care providers and health insurance providers.

- The UK Office for National Statistic's (ONS) **Secure Research Service** is a facility that enables access to restricted data from surveys and other confidential datasets produced by the ONS to Accredited Researchers.[30]
- **Administrative Data Research UK** (ADR UK) is a partnership between ADR Scotland, ADR Wales, ADR Northern Ireland and the ONS, and which links together and anonymises data held by different public bodies and facilitates access thereto for approved researchers.[31]
- **The Health Data Research Hubs** are centres of excellence in the UK facilitating access to data held by the public sector and the use of which is subject to the rights of others. The Hubs bring together data from NHS hospitals and facilitate access for the public sector, academic and industry research.[32]
- The **French Centre d'accès sécurisé aux données (CASD)** is a partnership between several French universities, research centres and the national statistics institute (INSEE). CASD makes available data from INSEE and from different ministries (including Justice, Education, Agriculture and Economics), from hospitals, and even from some private companies.[33] It provides this data through SD-Box, a secure infrastructure (or 'bubble') accessible remotely and where confidential data is 'sanctuarised'.
- The **French Health Data Hub** is a database and service provider gathering health data from various databases with a view to facilitating their reuse by research institutions – in full respect of privacy.[34] It also aims at enabling increasingly personalised medicine and a more efficient public system. In addition, it is tasked with promoting standardisation norms regarding health data exchange and reuse, taking into account European and international standards.
- The French **AVIESAN** (Alliance nationale pour les sciences de la vie et de la santé) brought together representatives from the research and health sectors to develop a plan to advance precision medicine and encourage the emergence of a national and industrial sector for genomic medicine. This Genomic Medicine Plan 2025 takes into account technological progress in storage, analysis, and reporting of big data. A National Centre for Intensive Calculation will gather the large volumes of data generated by twelve sequencing services and provide services for health care practitioners. One of the targets is to create a national framework capable of driving scientific and technological innovation and economic growth in numerous fields including big data processing, semantic web and the Internet of Things, medical devices, and eHealth. The consequences of precision medicine on policy will most likely not be negligible.[35]
- The **Nordic eInfrastructure Collaboration** (NeIC) is hosted by NordForsk, which provides for and facilitates Nordic cooperation on research and research infrastructure across the Nordic region. Among their activities, **Tryggve** is a Nordic collaboration for data the use of which is subject to the rights of others funded by NeIC and by research institutes forming ELIXIR nodes of participating countries.[36] Tryggve develops and facilitates access to secure e-infrastructure for data the use of which is subject to the rights of others, suitable for hosting large-scale cross-border biomedical research studies. It develops state-of-the-art scalable infrastructure for safe,

[28] Belgian Government, eHealth. See: https://www.belgium.be/fr/services_en_ligne/app_be_health
[29] CIN-NIC, Aperçu général de MyCareNet. See : http://fra.mycarenet.be/algemene-beschrijving
[30] Office for National Statistics, Secure microdata. See: https://www.ons.gov.uk/census/2011census/2011censusdata/censusmicrodata/securemicrodata
[31] ADR UK. See: https://www.adruk.org/
[32] Health Data Research UK, The Hubs. See: https://www.hdruk.ac.uk/infrastructure/the-hubs/
[33] CASD. See: https://www.casd.eu/
[34] Health Data Hub. See: https://www.health-data-hub.fr/
[35] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)32467-9/fulltext
[36] NEIC, Tryggve – Collaboration for Sensitive Data. See: https://neic.no/tryggve/ ; and ELIXIR, ELIXIR Nodes. See: https://elixir-europe.org/about-us/who-we-are/nodes

efficient, ethical, and legal storage, analysis and sharing of sensitive personal data for biomedical research between countries. The project supports open and transparent data access processes by engaging with the key stakeholders from each of the Nordic countries, and facilitates automated cross-border data exchange among Nordic countries with the ambition to scale this to Europe and beyond.

- The **Leuven Statistics Research centre** (LStat), created in 1988 as an interfaculty institute providing a coordinating link between all university research centres dealing with statistics. Currently, there are around 25 such centres (spread over 11 different faculties) involved in the activities of the LStat.[37]
- The **Edinburgh Parallel Computing Centre** (EPCC), which works closely with NSS and the Farr Institute to extend and enhance the new NHS Scotland safe haven service beyond its current basic computing capability in order to provide traditional High-performance Computing (HPC) services within the safe haven.[38]

In addition, two intermediaries are of particular interest, as they perform functions that are similar to those outlined in the proposed policy options. Annex I provides additional details on these two initiatives.

2.1.1.1.3.1 RatSWD

The German RatSWD is a **public advisory council to the German federal government** and was founded in 2004. The RatSWD aims at sustainably **improving the research data infrastructure** that underlies empirical research and at contributing to the international competitiveness of said research.

It is made up of an **independent body of researchers and representatives of data holders**, and acts as an institution of exchange and of mediation between the interests of science and data producers. As such, it is an important platform for **communication and coordination**.

RatSWD’s core tasks are the following:

- To issue recommendations on further improving the data infrastructure, specifically:
- Recommendations on how to secure and further improve data access, particularly by establishing and evaluating research data centres and data service centres according to a set of clear standards;
- Recommendations on how to improve data use through the provision of scientific and statistical data (research data portal; metadata) and appropriate documentation;
- Recommendations on research topics and research tasks pertaining to the conceptual development of data infrastructures on the national, European and international level;
- Recommendations on how to optimise the production and provision of research-relevant data;
- To advise science and policy, specifically:
- Advising the Federal Ministry for Research and the Länder governments on the development of the research-based data infrastructure;
- Advising public and private data producers;
- Advising data producers that are institutionally unaffiliated with independent scientific research on how to receive certification as a scientific research institution (certification);
- To monitor legal and technological developments, specifically:
- Monitoring national and legal developments in data provision;
- Monitoring technological developments, e.g. virtual research environments; and

[37] Leuven Statistics Research Centre. See: https://lstat.kuleuven.be/
[38] EPCC. See: https://www.epcc.ed.ac.uk/

- To organise and host the Conference for Social and Economic Data every three years.

Although RatSWD itself does not make data available to re-users, it is an **intermediary responsible for the accreditation of Germany's Research Data Centres** (RDCs) (see the discussion of the German Forschungsdatenzentrum above), which act as data holders and sometimes also as data re-users for research purposes. It coordinates these RDCs via a **Standing Committee Research Data Infrastructure (FDI Committee)** established in 2009. In addition, the RatSWD accredits RDCs when they meet a number of criteria, and monitors their continued compliance with these. The RatSWD also provides a search engine for datasets held by the 38 accredited RDCs in Germany (with a few exceptions).

2.1.1.1.3.2 Findata

Findata is Finland's recently established **one-stop-shop** responsible for streamlining and securing the secondary use of social and health data. It guarantees a flourishing ecosystem around the secondary use of social and health data streamlining the processes for the **issuing of research permits** and data collection and ensuring that data is being used in **secure environments**, thereby maintaining the trust that the general public have in authorities and the public sector.

Findata makes retrieving combined health and social data from different sources **easier, faster and possible with just one permit application**, removing the need to approach each authority and data source separately. These applications are not free for the re-user: for a Finnish or EU/EEA-based re-user, a decision on a data request costs EUR 1,000, excluding an additional data processing fee of EUR 115/hour (for the combining, pre-processing, pseudonymisation and anonymisation of the data).

Findata is also responsible for ensuring the **ethically sustainable use of data**. It makes decisions on data permits concerning data held by other controllers, and is responsible for the collection, combination, pre-processing and disclosure of data for secondary use, in accordance with the Act on Secondary Use of Health and Social Data. Furthermore, the data permit authority maintains a data request management system to forward and process data requests and permit applications.

Findata also maintains a **secure hosting service** for receiving or disclosing personal data and a secure operating environment, in which the permit holder may process the personal data he/she has been disclosed on the basis of data permit. It also supervises compliance with the terms and conditions of the permit it has issued. The data permit may be revoked if the permit holder fails to comply with the law or the terms and conditions of the permit. Lastly, the data permit authority is responsible for the pseudonymisation and the anonymisation of personal data.

2.1.1.2 The problem, its magnitude and the stakeholders affected

Several issues can be identified as part of this domain.

The first concerns **discovery of, and access to, data**. As stated in the European strategy for data, "*the value of data lies in its use and re-use*". The unavailability in some Member States of certain types of data the use of which is subject to the rights of others for re-use, results in a range of problems such as:

- The inability for re-users to access and use the data in order to conduct research and development that may have positive impacts for society, such as improved public services including public transport, healthcare or education;
- The inability for decision-makers to rely on re-users research outputs as input to their decision-making that could result in overall 'better' policies; and

- Barriers in the development of Artificial Intelligence (AI), which requires data to improve continuously.

The second issue is **data interoperability and quality**. Even when given datasets held by the public sector and the use of which is subject to the rights of others are made available for reuse, research may require combining datasets from different data holders. This is hampered by the insufficient data interoperability among datasets from different sources, which may structure their datasets differently – resulting in additional time spent by re-users on combining different datasets rather than on conducting research for the public good. This situation necessarily results in **fragmentation** as regards access to, and combination of, data of sufficient quality, and in an **imbalance** between large re-users (such as multinational pharmaceutical companies) that have the resources to perform this work, and smaller re-users (such as SMEs or some researchers) which do not always have that capacity. This imbalance is reinforced by the growing costs associated with re-use of the data, deterring smaller reusers.

An additional issue regards the **ability of citizens to exercise their rights**, notably under the General Data Protection Regulation, in the absence of transparent and streamlined processes to do so.

When examining the issue's **European dimension, the situation is even more fragmented**, as some Member States have taken steps to facilitate the re-use of data held by the public sector and the use of which is subject to the rights of others (such as establishing one-stop shops or single data permit authorities, as well as cross-border data re-use mechanisms) whereas others have not. This hampers research at a European scale – a scale that would enable both higher quality research, and increased AI development, due to larger datasets being available and interoperable.

While the benefits of reusing data the use of which is subject to the rights of others cannot be accurately estimated, data access and reuse of public sector data (including non-sensitive data) is estimated to bring social and economic benefits equivalent to between 0.1% and 1.5% of GDP, according to the OECD.[39] This therefore impacts society as a whole: data re-users, data holders, data intermediaries and citizens (both in their role as co-producers, and also as members of society more broadly).

Two broad categories of data holders can be differentiated: statistical offices and health- and social-related data holders. As regards statistical offices (and other public authorities responsible for the development, production and dissemination of statistics), the European Statistical System keeps an up-to-date list that currently contains 286 entities,[40] of which 27 are related to health (and therefore excluded from this count to avoid double-counting). As a result, the amount of **data holders in the EU27 when it comes to statistical microdata can be estimated to be around 260**.

As regards **health- and social-related data**, several broad types of data holders can be identified,[41] namely an estimated:

[39] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[40] European Commission, List of National statistical institutes (NSI) and other national authorities. See: https://ec.europa.eu/eurostat/documents/747709/753176/20190607_List_other_national_statistical_authorities_IT.pdf/f3c3bddf-c378-4203-92a2-48d0dd789f3d

[41] This identification is an extrapolation based on the different types of data holders concerned by Findata, on the different public sector partners of the French Health Data Hub, and on the health and social affairs data holders listed in the Centre d'accès sécurisé aux données. It excludes statistical offices to avoid double-counting.

- 55 Ministries of Health (typically one per Member State, except in federal countries, namely Austria with 9 länder, Belgium with three federated governments, and Germany with 16 länder);
- 55 Ministries of Social Affairs (typically one per Member State, except in federal countries);
- 104 Medical insurance authorities (of which it can be assumed there is on average one per each 104 NUTS-1 region);
- 55 Pensions Authorities (typically one per Member State, except in federal countries);
- 104 Social Security Authorities (of which it can be assumed there is on average one per NUTS-1 region);
- 27 Medicine Authorities (typically one per Member State);
- 27 Population registers (typically one per Member State); and
- 104 Hospitals Authorities (of which it can be assumed there is on average one per NUTS-1 region)[42].

Thus, there are roughly **530 data holders in the health and social domains**.

In addition, there are a number of cross-sectoral data holders: public universities and research centres that one the one hand re-use data the use of which is subject to the rights of others (on which see below), but on the other also produce such data (e.g. when conducting surveys). To estimate the number of such public entities, the Eurostat list of recognised research entities listing 666 recognised entities across the EU27 was used.

In total therefore, there are around **1,500 data holders** in total impacted.

With regards to data re-users, an estimated number of re-users of statistical microdata can be derived from Eurostat's list of recognised research entities, which lists a total of **666 recognised research entities** in the EU27. The total number of data re-users for health and social data overlaps with the research entities recognised by Eurostat: 48 of these conduct research in *inter alia* social sciences, while 22 conduct research in a health-related domain. However, it also includes a large number of private companies – that number is estimated to be **147,000 companies**.[43]

Thus, there are roughly **150,000 data re-users** impacted in total.

Lastly, two public data intermediaries can be reasonably assumed to exist in each Member State (except those with a federal structure) – one for health and social data, and another for statistical microdata – this is a total of **110 public data intermediaries** in total.[44]

The table below summarises these figures:

Table 4 - Estimated number of stakeholders impacted in the EU-27

Stakeholder type	Health and social	Statistics	Total

[42] Individual hospitals and doctors are considered to be data co-producers rather than data holders.

[43] This estimation was reached using a) the number of people employed in the healthcare industry (800,000 in 2012 in the EU, see https://ec.europa.eu/growth/sectors/healthcare_en#:~:text=A%20vibrant%20EU%20pharmaceutical%20sector,the%20EU's%20total%20manufacturing%20workforce.); b) the number of active businesses in the EU (27.5 million in 2017), and c) the number of employed persons in the EU (150 million persons in 2017), see https://ec.europa.eu/eurostat/statistics-explained/index.php/Business_demography_statistics. These figures were used to reach an average number of employees per active business (150,000,000/27 500 00=5.45); from which the number of healthcare businesses was derived (800,000/5.45=146,788.99) and rounded-up.

[44] Indeed, it is unlikely that a given Member State would have more than one public data intermediary for the same domain, since the reason behind their existence is to streamline procedures.

Data holders (public sector)	55 Ministries of Health 55 Ministries of Social Affairs 104 Medical insurance authorities 55 Pensions Authorities 104 Social Security Authorities 27 Medicine Authorities 27 Population registers 104 Hospitals Authorities	260 Statistical offices	**Approx. 1,500**
	666 recognised research entities		
Data re-users (private sector and academia)	70 recognised research entities (overlapping with statistics) 147 000 private companies	666 recognised research entities	**Approx. 150,000**
Data intermediaries (public sector)	55	55	**Approx. 110**

These numbers are **non-exhaustive**, as there are many additional stakeholders in sectors other than health and statistics, such as mobility, business registers or financial reporting databases.

2.1.1.3 The causes of the problem

This situation is a result of several key drivers. In many Member States, there is **uncertainty regarding applicable rules and legislation** – both to provide access to data the use of which is subject to the rights of others, and to access it. For instance, it remains unclear in some Member States whether national ministries have a legal obligation to make the data (the use of which is subject to the rights of others) they hold available for re-use.[45]

Second, data the use of which is subject to the rights of others cover a wide **range of different types of datasets** – ranging from health data to statistical microdata covering a variety of topics such as household composition. For example, the French CASD provides access to data categorised in a large number of themes, such as agriculture, businesses characteristics, unemployment, household consumption, housing, life expectancy, living standards, health, immigration, and others.[46]

Lastly, **different methods to enable the re-use of data the use of which is subject to the rights of others** co-exist, as detailed above.

2.1.1.4 The effects of the problem

The current situation results in a range of impacts. First, researchers must spend time and resources finding who holds the datasets they seek, discovering and understanding any specific procedure to request these datasets, filling in several separate applications to access the datasets, and potentially curating the data in order to enable their combination. This may **deter researchers with limited resources** from going through the process altogether, while researchers who decide to go through the process may see their application rejected – or one of their applications rejected (when they apply to more than one data holder). This makes it **more difficult** for them **to conduct their research**.

Similarly, businesses have to navigate through the same issues. While large companies may have the resources to do so, SMEs do not always have such resources, resulting in an **unequal access to data the use of which is subject to the rights of others and therefore reduced innovation**

[45] As was revealed during a stakeholder interview.

[46] CASD, Les sources de données déjà disponibles au CASD. See: https://www.casd.eu/les-sources-de-donnees-disponibles-au-casd/

and business opportunities when such access is limited. This impact is **cumulative**, since in effect larger companies are in a better position than small ones to innovate and to develop new products and services.

Data holders that do not allow access to the data the use of which is subject to the rights of others they hold may have **fewer incentives to ensure the data's high quality**. In addition, they are harmed by the abovementioned hindrances to research, as this research could be used as input for **better-informed policy-making**. This has negative impacts across the board for society, health, and the environment.

2.1.2 Establishing a certification/authorisation scheme for data altruism mechanisms

2.1.2.1 Background

2.1.2.1.1 Context

Building up "databases required for the development of artificial intelligence geared towards public service missions" will be a key development in the near future, as predicted in the recent French Parliamentary Mission Report on Artificial intelligence.[47] Through such databases, citizens or other stakeholders could choose to allow their data to be used for the public benefit.

As of today, data sharing is rapidly rising and the value of the EU27 data economy is already EUR 301 billion or 2.4% of the EU GDP. The predicted volume of data that will be shared will increase fivefold to 175 zettabytes in 2025 as compared to 2018[48]. This data includes personal data held by individuals, private sector and civil society organisations that can greatly benefit society. When focusing on the greater good of society, as opposed to the economic benefit of individual stakeholders, data altruism schemes are an interesting mechanism to consider for the public sector. Data altruism is, to date, a small part of data sharing however one that gained increased attention in the previous months, during the COVID-19 pandemic, in the form of data altruism for public health reasons.

While there is an overall willingness to share for example personal data for the public good[49], to date, wide-spread-data altruism scheme does not exist, even though a shared approach on this topic could provide large gains for society. A data altruism scheme is understood as digital data sharing for public benefit. Where data altruism decisions are based purely on the consent of individuals 'donating' their own personal data, data altruism schemes must allow data subjects to revoke their consent for data processing according to the General Data Protection Regulation (GDPR). An example of this revocation right can be found in the Corona-Datenspende App[50]. It needs to be stressed that 'donating' does not mean that the data holder/owner loses rights to their data, instead access to this data is provided and should, under GDPR, always be able to be revoked.

A key challenge is determining exactly when a data sharing decision is driven by altruism, as opposed to being an economic decision. A potential indicator of the presence of an altruistic motivation is to examine the circumstances of the data sharing: if the data producer receives a direct benefit (such as a new or improved product or service) as a result of sharing data, their motivations are less likely to qualify as altruistic even if a social good also materialises. Note that data altruism schemes are generally understood to comprise data infrastructures or mechanisms that will benefit the greater

[47] Villani (2018), FOR A MEANINGFUL ARTIFICIAL INTELLIGENCE TOWARDS A FRENCH AND EUROPEAN STRATEGY. Available at https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

[48] The European Data Strategy. Shaping Europe's Digital Future Factsheet. February 2020. Available at https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283

[49] Halvorson, Permanente, and Novelli (2014), Data Altruism: Honoring Patients' Expectations for Continuous Learning. Available at https://nam.edu/wp-content/uploads/2015/06/dataaltruism.pdf

[50] Corona-Datenspende App FAQ. Robert Koch Institute, https://corona-datenspende.de/faq/

society as a public good[51] and not for economic benefit of individual stakeholders. A data altruism scheme can take a number of different forms, depending on how it is set-up which again depends on various factors such as policy, legal, technology and organisational.[52,53]

Projects have begun in a number of European countries to explore different possibilities to enable data altruism. In Finland, for example, the MyData project aims to enable a 'paradigm shift in personal data management and processing that seeks to transform the current organization centric system to a human centric system'[54]. This conceptual approach goes well beyond the topic of data altruism, presenting a holistic approach to personal data management and developing a discussion framework that can host many implementations and models, including projects. However, one of the main ways that the infrastructure is described as creating value is 'as a common framework for different kinds of research data banks to easily acquire consumers' consent to collect their data"[55].

Similarly, the OwnYourData project has developed a Semantic Container for Data Mobility, supported by Horizon 2020 funding. The semantic container enables "secure and traceable data exchange between multiple parties"[56], with one of the use-cases named as "data donation" to support studies and research. Another example includes the Valencia.Data project in Spain which maintains a database of people who have chosen to make their personal data available through this project, together with a platform for the management of this data. The purpose of the project is to promote the reuse of data for research. This project was funded by the region of Valencia, public resources, and is still running various projects at the Instituto de Biomecanica de Valencia. The institute runs several research and development projects focused on data altruism with a focus on the public good. These include projects to for example reduce the amount of textile waste in the textile industry by asking data holders, citizens, to donate physical data to improve clothing seizing. While it sounds commercial, this is a project to utilize data sharing to reduce environmental damages from the textile industry. To date, up to 4000 data holders have shared their data for various projects, however the institute must request consent from every data holder for every new project to reuse data. The data holder then has the opportunity to approve or decline consent for the various projects. All projects are funded by the Valencia government and the institute had built every project infrastructure for data sharing independently, for which no exact price estimation could be provided. While the institute finds that data holders are willing to share data, this strongly depends on the time they must make available to share data and what they will receive in return (research results). The less time is required to share data and the clearer they will know what they can except in return, the more wiling they are to share data and to do so repeatedly for future projects.

A current approach of a data altruism ecosystem is the Corona-Datenspende-App in Germany. For the public interest, specifically to safeguard health and to prevent new pandemic outbreaks, the German Robert Koch-Institute (RKI) is collecting via this app health data related to COVID-19 of Corona-Donation-App-users. With a scientific evaluation of the donated data that was made

[51] Skatova A, Goulding J (2019) Psychology of personal data donation. PLoS ONE 14(11): e0224240. https://doi.org/10.1371/journal.pone.022424

[52] Kirkpatrick, R. A New Type of Philanthropy: Donating Data (2013). Harvard Business Review

[53] High-Level Expert Group on Business-to-Government Data Sharing (2020), Towards a European strategy on business-to-government data sharing for the public interest.

[54] Poikola, Kuikkaniemi, Honko (2014), MyData – A Nordic Model for human-centered personal data management and processing. Available at http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1&isAllowed=y

[55] Poikola, Kuikkaniemi, Honko (2014), MyData – A Nordic Model for human-centered personal data management and processing. Available at http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1&isAllowed=y

[56] Ownyourdata, SEMANTIC CONTAINER FOR DATA MOBILITY. See https://www.ownyourdata.eu/en/semcon/

available, the RKI project team calculates a ‛fever map’ with the societal purpose to detect early possible hotspots.[57] The app and its use is voluntary and the data are pseudonymised. Data collection and data processing are subject to strict data protection guidelines.[58] Through a scientific blog (www.corona-datenspende.de) providing relevant and further information about the app and the processing of the data, RKI aims to enhance transparency and reliability.

SelfData Territorial is a project of Fing (Next Generation Internet Foundation), a French organization which aims that individuals take control of their personal data, rather than tacitly allowing it to be exploited by others, in order to recognize their right to use (not own) their data.[59] The concept of Self Data means that people are empowered to reuse their personal data themselves for their own purposes. This also includes their ability to access their personal data in a reusable format, and that they are equipped with tools and services to do so.[60] According to Fing, the hoster of MesInfo, SelfData complies with existing data protection regulations in France and complements it. Currently, three French regions, Nantes Métropole, La Rochelle, and Grand Lyon are participating in the SelfData Territorial project.

There are therefore multiple examples of data altruism schemes and infrastructures that could be applied to this purpose and a potential high level of demand for them from both individuals who want to make their data available under specific constraints, and research and other organisations, which want to re-use and analyse it.

2.1.2.1.2 Ecosystem

A data altruism ecosystem consists of various stakeholders: the data subjects, intermediary, data holder and data re-user. The four stakeholders are described, visualized and explained in a Table 1 below.

The ecosystem stakeholders:

Data subjects are individuals or organisations[61] that actively share their data for an altruistic purpose. The exact level of detail on the purpose of such data sharing can, but must not always be known once sharing the data. Such purposes of data sharing include research and development, public health, public interest, matching of and synergising cross-sectoral data, information about public administrations and regarding the society, economy or the environment, transparency, or improvement of the access to public services. Data sharing must always be compliant with the GDPR regulation and allow the data sharer to revoke the data sharing rights.

Intermediaries are organisations or institutions which act as a link between data subjects and data re-users in order to transmit the data to the latter. Intermittent stakeholders are e.g. patient associations and health insurance schemes which collect patient data, or research organisations which collect personal data for research and study purposes.

A **data holder** is an entity such as public sector bodies, companies or organisations that manages, hosts and provides the shared data. Depending on the specific data scheme, data holders have

[57] Robert Koch-Institut, Blog zur wissenschaftlichen Auswertung der Corona-Datenspende, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende.html (13.05.2020).
[58] See previous footnote.
[59] FING, SelfData Territorial, 2020. See http://mesinfos.fing.org/self-data-territorial/.
[60] FING (The Next Generation Internet Foundation), Understanding Self Data, 2017. See http://mesinfos.fing.org/wp-content/uploads/2017/08/selfdata_FAQ_mydata2017.pdf.
[61] Note that the data subject for the purposes of this note is therefore not necessarily a data subject in the sense of the GDPR, since a data subject under the GDPR must be an individual natural person, not an organisation or other form of legal entity.

respective technical, organisational and/or governance functions in order to make the data accessible to interested re-users by the request of the data subjects.

A **data re-user** is an entity, usually a research body, that re-uses the shared data to create new knowledge insights that contribute to the benefits of the society. Data re-users can be researchers, public sector bodies or non-governmental organisations. The re-use of shared data should be solely for public benefit and not for private economic gains.

A data altruism ecosystem can be approached from the perspective of a public body or private body. Both approaches are visualized below and mapped in the ecosystem mapping in Table 1.

As an example, to illustrate that stakeholders can have multiple roles in the ecosystem, the Corona-Datenspende App ecosystem is also mapped in Table 1 as well as a Business-to-Government example from the maritime sector:

Table 5 –Data altruism scheme ecoystem

Example	Data subjects	Data holder	Intermediaries	Data re-user	Type of data	Purpose
Public sector data holder	-Citizens -Companies	-Public sector bodies -NGOs	-Public sector bodies -Organisations -Companies (hosting a platform/scheme for data altruism)	-Public sector bodies -Researchers (not for economic gains) -Organisations	-Personal sensitive data -Personal non-sensitive data -Other data	-R&D -Public benefit
Private sector data holder	-Citizens -Companies	-Companies	-Public sector bodies -Organisations -Companies (hosting a platform/scheme for data altruism)	-Public sector bodies -Researchers (not for economic gains)	-	-R&D -Public benefit
Corona-Datenspende App	*-Citizens*	*-Public sector bodies (Robert-Koch-Institute)*	*-Robert-Koch-Institute*	*-Researchers (project team of the Robert-Koch-Institute)*	*-Personal sensitive data*	*-R&D* *-Public health (prevent new pandemic outbreaks)* *-Societal benefits*
Vessel Traffic Data PoC	*-Vessels*	*-Statistics Netherlands*	*-MarineTraffic*	*-Statistics Netherlands*	*-Other data*	*-R&D* *-Maritime (improve statistics for maritime related policymaking)*[62]

[62] Towards a European strategy on business-to-government data sharing report. European Commission (2019)

2.1.2.1.3 Ongoing initiatives/Market analysis

This section outlines a list of initiatives aiming at facilitating data altruism schemes and active data altruism schemes in the European Union. The focus in on data intermediaries and data reusers, in particular public research institutes and NGO's. Due to the relative recent development of data altruism schemes, there is a limited amount of examples, nevertheless it has to be noted that the current COVID-19 pandemic has led to an acceleration of national health related data altruism schemes. Germany and Italy have both developed data altruism schemes in the form of contact tracing applications for smartphone and will be presented in-depth.

Data altruism scheme- Data intermediary and data re-user

It has to be noted, as explained in the ecosystem, that the intermediary and data re-user can often be the same actor considering public bodies that manage the application (intermediary) are also reusing the made available data.

- The Instituto de Biomecanica de Valencia (IBV), part of the Universidad Politecnica de Valencia in Spain, is a technology research center that ran the **Valencia.Data** project from 2018-2019[63]. As a data reuser, the IBV established a data altruism scheme by creating an app that enables citizens to share personal data, such as anatomical-physiological data, for research purposes. The IBV is an intermediary and data re-user in this instance.
- The **MESINFOS Project** was a project, run by the City of Lyon, aiming to empower citizens to reuse and share their energy consumption data[64].

2.1.2.1.3.1 The German Corona-App approach

At the beginning of April, the German Robert Koch-Institut (RKI) released the **Corona Data Donation App** (in German: Corona-Datenspende-App) which aims to contribute information about the spreading of SARS-CoV-2 in Germany. With a scientific evaluation of data, the project team calculates a "fever map" in order to detect early possible hotspots.[65]

The app and its applications is voluntary and anonymized. Currently (05.05.2020), 509,532 persons have registered with a total amount of 15,259,595 data sets made available.[66] The data covers information about sex, age in a 5-year-interval, size in 5-cm-intervals, health status and activity data regarding sleeping behavior, heart frequency and temperature, and the regional code.

The algorithms in the app can recognize symptoms, which are correlated with an infection with the Coronavirus.[67] The purpose of the data is exclusively of scientific nature. After a careful analysis, the data contribute to the visualization of a map, which shows the regional distribution of infected persons up to a local level. The RKI has at no time access to personal information such as name or address of the app users. Data collection and data processing are subject to strict data protection guidelines.[68] These standards were verified before launching the app. Persons who choose to make their data available have the possibility to access, administer and delete all of their personal data. The app was edited by the RKI in collaboration with Thryve (mHealth Pioneers GmbH), a digital health

[63] VLC.Data. Instituto de Biomecanica de Valencia. https://www.ibv.org/blog/proyecto/valenciadata-ecosistema-digital-centrado-en-las-personas/

[64] MESINFOS http://mesinfos.fing.org/english/

[65] Robert Koch-Institut, Blog zur wissenschaftlichen Auswertung der Corona-Datenspende, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende.html (13.05.2020).

[66] Corona-Datenspende Blog, Robert Koch-Institut, Der Corona-Datenspende Blog, 2020, https://corona-datenspende.de/science/ (13.05.2020).

[67] Robert Koch-Institut, Corona-Datenspende-App. Hände waschen, Abstand halten, Daten spenden – Ihr Beitrag gegen Corona, 2020, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Corona-Datenspende-allgemein.html (13.05.2020).

[68] See previous footnote.

company. Through a scientific blog (www.corona-datenspende.de) providing relevant and further information about the app and the processing of the data, RKI aims to enhance transparency and reliability.

The release of the Corona Donation App has been accompanied by an intense public discussion about where to store the data and who is ultimately controlling the data that was made available:[69] The two diverging approaches are either a decentralized or a centralized storage of data. Whereas within a decentralized model, a user in case of an infection only sends its own IDs to the server, in the central version the app sends additionally the codes of all contacted persons to the server. This information is sensitive and needs to be protected with special standards. The current version of the app follow the decentral model.

In addition, the Federal Government of Germany launched the **Corona-Warn-App** to "*help fight the coronavirus"*[70] on the 16th of June 2020. This contact tracing app aims to notify [the app] users when they have been in contact with an infected person which could result in a risk of catching the virus. Thereby, the government hopes to interrupt the chains of infection and reduce the spread of the virus in Germany. The application is a form of data altruism because it helps to break chains of infection and provides valuable data insights for the health ministry, and user of the app, who could not track everyone's interaction of the past 14 days. This is also an example that data altruism exists is many different variations that can help the greater public good, including limiting the further spread of a pandemic.

The app operates by relying on Bluetooth technology, which must be active on a user's smartphone, to measure the distance and duration of contact between people that have installed the app. The app is available for free to download on IOS and Android devices and serviced by the Federal Government. The technology is enabled to exchange temporary encrypted random IDs but does not allow connections to be made to a user's identity or location. When a user tests positive for the virus, they can voluntarily inform other users by notifying in the app that they have tested positive. The app then checks who has been in contact with the infected person and notifies them through a warning in the app. To conclude, the Federal Government is the data re-user, the application user the data holder and the application the intermediary[71].

Note, that data security and protection is of high importance in Germany and the Federal Government ensures users that the app meets European and German data protection requirements. To do so, the Federal Commissioner for Data Protection and Freedom of Information (BfDI) and the Federal Office for Information Security (BSI) were involved in the development of the Corona-Warn-App[72]. Users remain anonymous, they do not have to provide e.g. their personal data when they register to use the app, and all data is **encrypted** and **stored exclusively on your own smartphone**. This is significant because thereby the **data holder** provides access to data but does not share or donate the data itself with the data reuser.

[69] Tagesschau, Corona-Tracing-App. Was heißt zentralisiert oder dezentral?, 2020, https://www.tagesschau.de/inland/coronavirus-app-101.html (13.05.2020).

[70] The Coronavius warning app. The Federal Government of Germany (2020): https://www.bundesregierung.de/breg-de/themen/corona-warn-app/corona-warn-app-englisch

[71] Ibid.

[72] Ibid.

To date, Friday the 26th of June 2020, the app has been downloaded by 13.3 Million users[73] and the Federal Government is investigating cross-border collaboration with Switzerland, the Netherlands and France to establish interoperability across national borders to fight the virus.

2.1.2.1.3.2 The Italian Immuni App

At the beginning of June 2020, "Immuni", the Corona App of Italy was released and has to date (26.06.2020) already been downloaded over 3 million times. As other Corona-apps, Immuni aims to contribute to the containment of SARS-CoV-2 and thereby to prevent potential outbreaks of the virus.[74] The functions are not yet activated nationwide, which is expected to happen in the upcoming weeks. As in Germany, the app is based on a source code which is openly available and based on contract tracking via Bluetooth Low Energy technology. The data is stored in a decentralized way; and personal data, such as name, date of birth, mobile phone number, identity of contact person or location is not asked for.[75] Anyone with close contact with a user who has tested positive for COVID-19 will receive a warning from the app regarding a potential risk of infection. In order to increase the number of potential users, altruism was used as a strategy to improve the trust in and acceptance of the app, according to Paolo de Rosa, Chief Technology Officer at the Ministry for Innovation Technology and Digital Transformation.

Data protection was key in developing the Immuni App. Data holder is the Ministry of Health in Italy and other public institutions and are stored locally at servers in Italy. Data and connection of the app to the server are protected.[76] The completely transparent approach was chosen by the government because it should create, according to de Rosa, trust in the app. Before the release of the app, these trust and reputation problems were addressed with communication campaigns and awareness raising. A specific fast track law, which is also GDPR compliant, was made for the Immuni app.[77]

The app was developed in a public-private partnership: the mixed stakeholder-combination consisted of one IT-company and, publicly, people from the academia and ministries. The cost of development, the provision of infrastructure, legal, implementation and other costs is estimated to 10 mio. EUR. However, everyone involved worked pro bono, so no real costs have arisen.[78]

In the meantime more Member States have, or are, developing and releasing Corona tracing applications.

2.1.2.2 The problem, its magnitude and the stakeholders affected

The European Union's aim to create a single market for data, to ensure Europe's global competitiveness and data sovereignty, and to create a data-driven society. This includes data sharing initiatives for the public good, such as in case of the healthcare sector as has been argued for especially during the current COVID-19 pandemic. To date, there is no European, or national, data sharing initiatives that enable data holders, whether private or public, to share data for the public good. There are initiatives, especially for the healthcare sector, however, these are nationally bound.

Data altruism has been highlighted by the German Ethics Council since 2017, which addressed the importance and value of data altruism, however also a foundational problem- the definition of data

[73] Anzahl der Downloads der Corona-Warn-App…im Juni 2020. Statista (26.06.2020) https://de.statista.com/statistik/daten/studie/1125951/umfrage/downloads-der-corona-warn-app/
[74] Immuni ist ein weiteres Instrument im Kampf gegen die Pandemie. Presidenza del Consiglio die Ministri (2020): https://www.immuni.italia.it/?gclid=EAIaIQobChMIso7XkMCf6gIVwZAYCh3c0QDpEAAYASAAEgJwLvD_BwE
[75] Ibid.
[76] Ibid.
[77] Paolo De Rosa, Chief Technology Officer, Ministry for innovation Technology and Digital Transformation, Italy in an interview on the 16th of June 2020.
[78] Ibd.

altruism, how to handle and or grant consent for it and the subsequent reuse of shared data[79]. This is a legal debate to ensure adequate data protection, however the German Ministry of Health found that 79% of German would be willing to share their data for research purposes, highlighting a willingness to participate if such a data altruism mechanism would be in place. This data is from 2019, before the COVID-19 pandemic and the current public willingness to share data appears to be less when looking at the amount of participants that have downloaded the German COVID-19 mobile phone applications.

This highlights a problem within the problem. The first problem is the absence of data altruism mechanisms in the European Union, the second problem is the stakeholders that are essential to a functioning data altruism mechanism such as the data holders, often citizens or private companies, that need to trust the mechanisms to share data with the data re-users.

2.1.2.3 The causes of the problem

One of many reasons why data altruism schemes are created is to resolve, or at least to mitigate, a multitude of legal challenges which can be linked to data sharing. When done by individual persons on a voluntary basis, data altruism will almost inevitably involve the processing of personal data as defined under the GDPR, both through the initial transfer of personal data to the data re-users, and through any subsequent use of the data thereafter.

As a result, the requirements of European data protection law must be observed, including those included in the GDPR. This implies first and foremost that any re-use of the personal data made available must have a clear legal basis. In a typical data altruism scheme, a person will provide their consent to re-use their data, thereby providing a suitable legal basis.

However, reliance on consent is not as trivial as it seems: in order to be legally valid, consent must be freely given, specific, informed and unambiguous. Freely given implies that there may be no element of coercion in the consent, e.g. because consent for re-use must be provided in order to be able to gain access to a product or service. This also implies that the consent cannot be given when there is a relationship of authority between the data subject and the recipient, such as e.g. between an employee and an employer, or between a student and teacher.

More importantly, consent must be specific and informed, which means that the purpose of re-use must be described in a way that allows the data subject to understand at a sufficient level of detail what their data will be used for, and what the potential implications are. A generic description – e.g. making data available "for the public good" or "for scientific research" is in principle considered too generic to meet this requirement[80]. However, there is some flexibility on this point for scientific research, in cases where the purposes for data processing within a scientific research project cannot be specified at the outset. In those cases, recital 33 to the GDPR allows the purpose to be described at a more general level[81]. As a result, attempts have been undertaken – e.g. by the Medizininformatik

[79] "Datenspende"- Bedarf fuer die Forschung, ethische Bewertung, rechtliche, informationstechnologische und organisatiorishce Ramenbedingungen. Bundeministerium fuer Gesundheit. March 2020; https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Ministerium/Berichte/Gutachten_Datenspende.pdf

[80] Recital 33 of the GDPR notes that "It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose."

[81] The European Data Protection Board has affirmed this position. See the Article 29 Working Party Guidelines on consent under Regulation 2016/679, adopted on 28 November 2017, as last Revised and Adopted on 10 April 2018, WP257; https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051

Initiative[82] - to provide a more generic and broad consent template for scientific research, which have been approved by the Conference of Independent Data Protection Commissioners of the Federal Government and the German federal states. A scaled-up version of this initiative at the EU level could possibly provide greater legal certainty on the possibilities and constraints of such consent forms. Finally, consent must also be unambiguous, i.e. expressed through an affirmative action; consent cannot be deduced from circumstances, or induced by e.g. pre-ticking consent boxes and relying on the data subject's inaction.

Reliance on consent, while clearly a key way to ensure that re-use in data altruism is lawful, also has the limitation that it is not possible for children, that a person can only consent to the use of their own data (which precludes e.g. cases where the data describes interactions with other persons, since their personal data would then be made available as well), and that the consent must be revocable – implying that platforms relying on consent must allow persons engaging in altruism to essentially change their minds and put a stop to the use of their data. This has infrastructural implications as well, since consent management must be built into data altruism platforms, along with sufficient controls to allow future processing to cease after the revocation of consent.

Data altruism schemes can be a part of the answer to these problems, not only by offering a shared platform for consent management (essentially through a specialised form of Personal Information Management Systems (PIMS)[83], but also by streamlining the process of informing users appropriately about new re-use cases, and by building a governance mechanism on top of the altruism use cases that enables a degree of control over future re-use, and notably whether the re-use complies with the information provided when the consent was obtained. In this case, the intermediaries in data altruism can play the role of supervisors and enforcers of the scheme, at least to some extent, and notwithstanding the protections afforded by the GDPR.

The legal basis is not the only challenging factor in data altruism. The GDPR also is based on the purpose limitation principle, implying that the purpose of the data processing must be defined upfront, the data subject must be informed of the use that will be made of the data, and that the data may thereafter not be used in a manner which is incompatible with the communicated purpose. In this way, the GDPR ensures predictability for legal subjects, and avoids misuses which are based on overly broad, ambiguous or misleading phrasing. As a result, a data altruism decision requires that a reasonably precise description of the re-use is given, comparable to the informed consent requirement as described above, and that this description accurately describes the limitations to any future use of the data. This also implies that the data subject must be informed in a sufficiently systemic manner when a new purpose of re-use is identified.

The data subject also has rights that are unalienable and continue to apply even after the data has been made available for re-use. Beyond the right to revoke consent (where applicable), the data subject also has the right to restrict future processing of their data if they feel that a specific use is unlawful. Furthermore, the data subject has a data portability right to the data that they made available, implying that they may ask to receive it back in a structured, commonly used, machine-readable and interoperable format. Interestingly, this obligation might be perceived as a barrier for re-users (who must implement a way to support this right), but it is also an enabler for data altruism: the data portability right can only be respected in a relatively user friendly manner if a way is found

[82] See https://www.medizininformatik-initiative.de/en/collaboration/consent-working-group

[83] Personal Information Management Systems (or PIMS) are systems that help give individuals more control over their personal data. PIMS allow individuals to manage their personal data in secure, local or online storage systems and share them when and with whom they choose. See EDPS Opinion 9/2016 on Personal Information Management Systems, https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf

to move it back to the user, or to third parties, thus containing an implicit obligation to provide the interfaces and technical resources to facilitate data migration across a multitude of stakeholders.

The GDPR also contains safeguards against profiling, which can occur in data altruism cases where the personal data of the data subjects is used to firstly evaluate their situation, in order to then make inferences that affect third parties. In health care for instance, data from historic patients and their treatment might be analysed in order to create treatment profiles for future patients – resulting in profiling of future patients, on the basis of profiles created using the data made available by the original patients. This implies that care must be taken that the data made available is not used in a manner that automatically produces significant effects on the person concerned, e.g. by affecting the availability or cost of a service on the basis of the data made available.

Data altruism schemes can mitigate many of these legal problems by creating governance mechanisms that ensure that a homogeneous response can be given to all of these questions, and that the relevant safeguards (consent, revocability, purpose restriction and so forth can be verified and policed to some extent.

2.1.2.4 The effects of the problem

The current state of development on data altruism schemes is very fragmented in the European Union. While the number of examples of such schemes seems to be increasing, the examined schemes seem to struggle to scale up to a European level, or generally to grow beyond a strictly defined and geographically bounded context. Based on currently available data and conducted interviews, to a large extent the legal uncertainty and lack of consensus on how to address the legal challenges with regards to data privacy, permissible reuse, governance and enforcement, act as a barrier. As a result, researchers and project initiators spend much time addressing legal questions to set-up their own schemes in accordance with local rules and sensitivities. This **legal fragmentation**, **lack of awareness** and **lack of consensus on best practices deters** researchers and initiators from establishing data altruism schemes or at least **significantly increases legal costs** to establish such a scheme. The ultimate effect is that **data altruism initiatives are harder and more costly to organize in the EU, resulting in both internal market fragmentation and a competitive disadvantage towards other regions of the world when it comes to using data for the public benefit.**

2.1.3 Establishing a European structure for governance aspects of data sharing

2.1.3.1 Background

Data sharing and data re-use are essential to data innovation. OECD assess that data sharing can "generate social and economic benefits worth between 0.1% and 1.5% of GDP in the case of public-sector data, and between 1% and 2.5% of GDP (in a few studies up to 4% of GDP) when also including private-sector" (OECD, 2019).

Yet a set of conditions needs to be in place so that businesses (and in particular across sectors) can reap the benefits from data-sharing. One of such conditions includes ***the agreement and implementation of data standards, metadata standards, data schemes and interoperability principles***.

2.1.3.1.1 Context

Data sharing and reuse requires data holders' agreements on data standards widely adopted across industries. Those standards are not only difficult to negotiate but also, and most importantly, to implement, being often fundamental to unleash innovation.

Data standards refer to "reusable agreements that make it easier for people and organisations to publish, access, share and use better quality data"84. A data standard is considered *open* when it is accessible to everyone for use or share. The main functionalities of data standards are to[85]:

- Agree on a consistent vocabulary and common attributes for data, which are defined in registers, taxonomies, vocabularies or ontologies.
- Facilitate the exchange of data within and across organizations by employing common data formats and shared rules, which are defined in specifications, schemas or templates.
- Offer guidance for sharing better quality of data and understanding information flows, which are defined in models, protocols or guides.

2.1.3.1.2 Ecosystem

The data ecosystem related to this domain refers to:

The different stakeholders in the data ecosystem considered under this domain include (see Table 2):

- **Data holder**, which refers to companies in traditional sectors that collects, maintains and publishes data, making it available for others to use.
- **Data reusers,** which refers to any companies who use data and extracts benefits from information and insight – including both business in traditional sectors and data companies
- **Intermediaries,** which refers to any organization that facilitates data flows between data holders and data users. Data intermediary can take the role of partnerships, consortium, platform, non-governmental bodies, data standardization associations and any organizational form that facilitates data sharing across organizations.

Table 6 - Stakeholder scope (data value chain mapping)

Domain	Data holder	Data reusers	Intermediaries
Establishing a European structure for governance aspects of data sharing	Business in traditional sector	Other business (competitors)	Public and private organisations in charge of data spaces and standardization initiatives
		Other business in the same sector (downstream/upstream)	
		Business and researchers from different sectors, esp. tech	

In order to obtain an understanding of the current problems in data standardization activities in EU, its causes and effects, we applied purposive sampling to select key informants that can provide us with a comprehensive overview. We use our judgement to choose the key informants based on i) geographical, ii) cross-sector and iii) inclusiveness of data stakeholder type, to inform our sampling, rather than aiming to construct a representative sample.

2.1.3.1.3 Ongoing initiatives/Market analysis

Top-down Standard Development Organizations (SDOs) coordinate the development of compatibility of data standards that ensure technological progress. They are legally mandated processes where data holders come together in a participatory process of consensus building that seeks to enable the development and diffusion of data standards that are democratically agreed and aligned with broader

[84] Open Data Institute, 'Open Standards for Data Handbook', Retrieved from https://standards.theodi.org/.
[85] Ibid.

policy goals. Yet, formal standardization processes often suffer from major drawbacks, due to the lengthy and challenging process of consensus forming and sometimes lack of market orientation.

In response to such problems, ad-hoc and industry-lead SDOs emerged with heterogeneous origins, goals and institutional logics. Industry-lead standards development processes are self-organized and not mandated by law. They include industry consortia, loosely coordinated temporary working groups and task forces, but also not-for-profit organizations that help stakeholders to organize data standard making processes in a more permanent fashion. Such de facto data standards result from market-based standard setting processes, in which data standards are generated through competitive forces.

While top-down SDOs emphasize consensus and social welfare implications, de-facto data standards stress speed, agility and the needs of the industrial constituents. Both standardization efforts co-exist to develop compatibility standards; which implies that the process of standardization evolves within and across multiple SDOs.

While different formal or informal SDOs have emerged to foster data standardization needs, a set of ***intermediaries*** are facilitating the agreement between data holders and users on data standards. As part of the research effort in the present study, a set of data intermediaries have been identified and interviewed (see table 3). We provide a brief description of them below and some estimates about the benefits and costs that they incur according to the sources gathered in the desk research and the evidence provided during the interviews (see detailed analysis in section assessing the four policy options):

2.1.3.1.3.1 International Data Spaces Association[86]

IDSA consists of a trustworthy architecture where more than 101 companies and institutions across industries from more than 20 countries. The goal of IDSA is to guarantee data sovereignty by **reference architecture for peer-to-peer network** providing usage control of data from all domains. Trust and security are the core pillars structuring its work. IDSA aims at providing the architecture that supports sharing data between different endpoints while ensuring data sovereignty. Main components in IDS architecture are the so-called 'IDS Connectors', which are the gateways that ensure control over data sharing at any source and point of use.

One key element of these connectors is the automated enforcement mechanisms of the relevant data policy – such as restrictions, limited persistence, disallowing transfer to third parties and so on, based on the standard policy language XACML (eXtensible Access Control Markup Language) and Information Model's Usage Control module, which provides machine-readable specifications of usage control policies. In other words IDS has enforcement mechanisms built in the architecture to control which data are used, how and by whom.

The governance and control in IDS ecosystem is expressed in the *certification* criteria for tiered security levels, specifically appointed evaluation facilities and certification bodies also at global scale. It enables the secure exchange of data and easy integration and aggregation of data in business ecosystem. By employing certified core components and certified technical and organizational security measures, IDS guarantees to its member that the architecture operates under the principles of trust. The certification of participants and components takes place in two phases:

- IDS_ready Review, which is implemented by members of the certification working group and the Head Office in charge of issuing the IDS_ready statements.

[86] International Data Spaces Association: https://www.internationaldataspaces.org/

- IDS Certification, which consists in an evaluation implemented by evaluation facilities and approved by the IDS Certification Body, which is responsible for issuing the IDS certificates.

Organizations engagement in IDSA requires around 20% dedication of one person and is estimated to generate in average around 15% efficiency savings for the companies. The current 22 use cases of IDSA that span from logistics, defence and manufacturing sectors reflect how a common reference architecture (i.e. technical, procedural, organizational and legal) leads to companies efficiency gains.

2.1.3.1.3.2 CODATA[87]

CODATA is the Committee on Data of the International Science Council (ISC). CODATA exists to promote global collaboration to improve the availability and usability of data for all areas of research. CODATA Works towards fostering scientific data sharing. The principle of CODATA is that research data should be as open as possible and as closed as necessary. CODATA works also to advance the interoperability and the usability of such data: research data should be FAIR (Findable, Accessible, Interoperable and Reusable). Recent studies have estimated that the annual financial cost of not sharing FAIR data to be at least EUR 10.2bn for the European economy; an additional estimate of the impact of FAIR on potential economic growth is EUR 16bn annually[88].

In consequence, one of the current tasks of CODATA is to support scientific data sharing across research domains. CODATA facilitates the dialogue across disciplines to agree on minimum common denominator across research domains about metadata and data structure that facilitates data interoperability across research domains. To achieve such goal, CODATA has set up a number of standing committees and strategic executive led initiatives, and through its Task Groups and Working Groups. It also collaborates on major data conferences like SciDataCon and International Data Week.

The benefits of research data sharing include: to improve reproducibility; to accelerate scientific processes and research velocity; increased scientific quality; to prevent scientific fraud; and to increase scientific productivity by reducing redundancy and innovation gains[89],[90],[91],[92],[93].

Yet the average estimated costs of introducing the metadata and contextual information required for scientists to re-use the data are around 5% of the total research budget. Other sources estimate that such production of metadata and the contextual descriptions of datasets could span an estimated two to three weeks from an average of a two-year research grant application[94]. In a dedicated study to examine high-energy physics practices, the vast majority of respondents (94.3%) thought that "the additional effort needed for preparing data for preservation in a re-usable form is substantial

[87] CODATA: https://codata.org/

[88] European Commission. 2019b. "Cost-Benefit Analysis for FAIR Research Data : Cost of Not Having FAIR Research Data." Website. https://op.europa.eu:443/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1

[89] Borgman, Christine L. 2015. Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, UNITED STATES: MIT Press. http://ebookcentral.proquest.com/lib/georgetown/detail.action?docID=3339930

[90] Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. "Science Friction: Data, Metadata, and Collaboration." Social Studies of Science 41 (5): 667–90. https://doi.org/10.1177/0306312711413314

[91] European Commission. 2019a. "Facts and Figures of Open Research Data." Text. European Commission - European Commission. 2019. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en

[92] OECD. 2015. "Making Open Science a Reality." 25. OECD Science, Technology and Industry Policy Papers. Paris: OECD Publishing. https://wiki.lib.sun.ac.za/images/0/02/Open-science-oecd.pdf

[93] Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." Edited by Peter van den Besselaar. *PLOS ONE* 10 (8): e0134826. https://doi.org/10.1371/journal.pone.0134826

[94] OpenAire, Super. 2019. "RDM Costs." OpenAIRE. 2019. https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs

(more than 1% of the overall effort invested in producing and analysing the data) whereas 43.0% think that the supplementary effort is more than 10%"[95].

2.1.3.1.3.3 iSHARE[96]

iSHARE consists in a stable set of agreements that makes possible that organizations give access to their data to a pool of unknown organizations without requiring bilateral or ad hoc agreements. With the scope in the logistics sector, iSHARE was set up through a collaborative project that started in 2017 lead by Innopay, a consultancy firm in the Netherlands, and soon the benefits of the scheme were acknowledge and lead to the creation of iSHARE foundation.

By giving all organizations the same identification, authentication and authorization methods, companies joining iSHARE scheme gain efficiency and do not incur in costs every time they want to share data with a specific organizations. Overall, organizations can avoid time-consuming integrations when they want to share data. ISHARE set of agreements allow the data owner to remain in full control over their own data at all times. Additionally, data owners decide the terms under which their data will be shared, with whom and for how long.

As such, iSHARE combines functional, technical, operational and legal agreements that organizations adhere. These set of agreements support both Machine to Machine (M2M) or Human to Machine (H2M) interaction. It also supports portable identity(s) for parties and humans; flexible authorizations, applicable in heterogeneous context; facilitates data exchange based on delegations; control over own data through management of consent; and provides a trust framework.

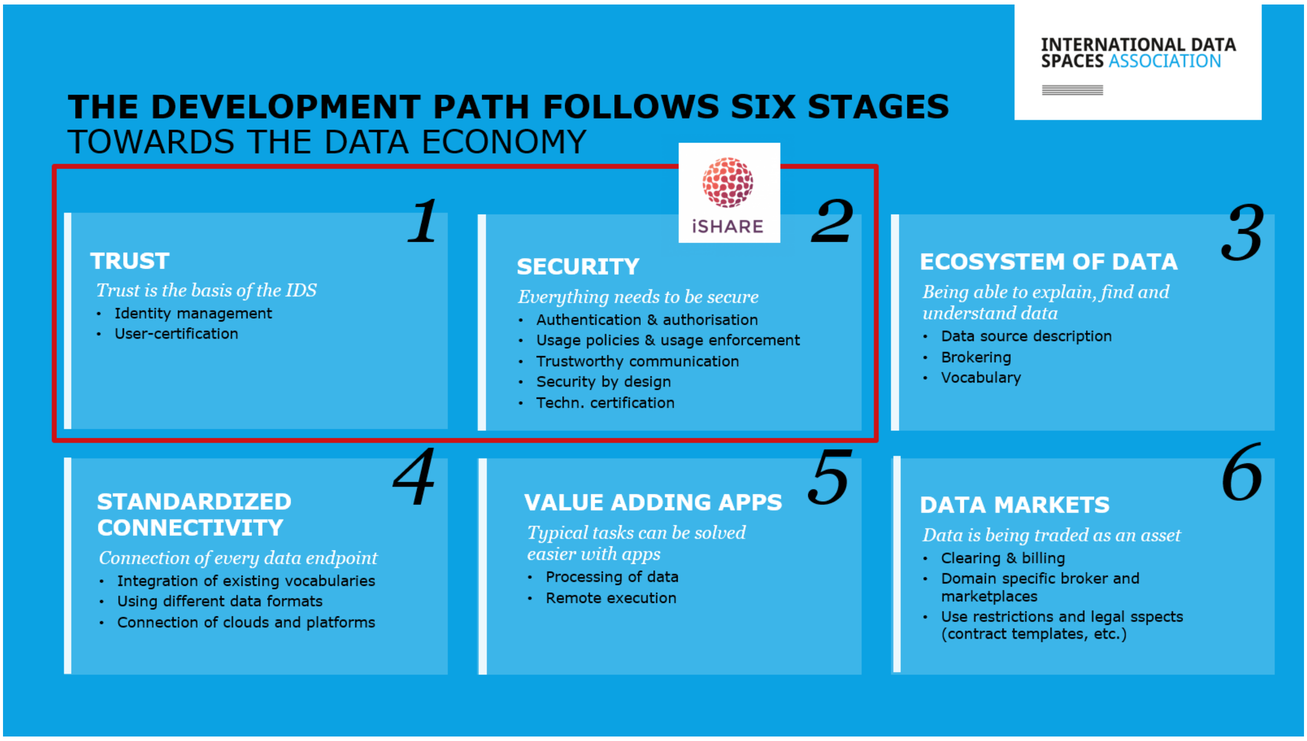
The operational description of what iSHARE offers to the data holders and users is the following: participants sign one contract with the Scheme Owner, which implies having a contract with all participants of iSHARE automatically. While participants remain free to develop additional contracts that do not conflict with iSHARE scheme, by signing the contract with the Scheme Owner, participants are able to share their data amongst them. As part of iSHARE scheme, an important aspect is the trust framework that the scheme designs, where licenses define the conditions under which data can be shared or the services that can be offered or consumed. The trust framework relies on technological solutions that allow organizations to authenticate with the other in a reliable way. Within the iSHARE scheme there is an API (Application Programming Interface) architecture for identification, authentication and authorization, which is based on a modified version of OAuth and OpenID Connect standards. While the setup of iSHARE scheme cost around few million Euros, the maintenance of the scheme is considered to require less than one million operational cost, which can be supported through members fee and transitioning public grants.

iShare and IDSA are working on similar issues and have formalized their collaboration in December 2019. iShare is at a more advanced stage of deployment but it is narrower in scope as it focusses only on the first two stages of trust and security, as illustrated by the chart below. The iShare solution is now aligned with the IDSA reference architecture and can therefore be consider as one solution to implement the IDSA framework.

[95] Holzner, Andre, Peter Igo-Kemenes, and Salvatore Mele. 2009. "First Results from the PARSE.Insight Project: HEP Survey on Data Preservation, Re-Use and (Open) Access." ArXiv:0906.0485 [Hep-Ex, Physics:Physics], June. http://arxiv.org/abs/0906.0485

[96] ISHARE: https://www.ishareworks.org/en/node/6

Figure 1: Relation between iShare and IDSA



INTERNATIONAL DATA SPACES ASSOCIATION

THE DEVELOPMENT PATH FOLLOWS SIX STAGES
TOWARDS THE DATA ECONOMY

iSHARE

1 **TRUST**
Trust is the basis of the IDS
- Identity management
- User-certification

2 **SECURITY**
Everything needs to be secure
- Authentication & authorisation
- Usage policies & usage enforcement
- Trustworthy communication
- Security by design
- Techn. certification

3 **ECOSYSTEM OF DATA**
Being able to explain, find and understand data
- Data source description
- Brokering
- Vocabulary

4 **STANDARDIZED CONNECTIVITY**
Connection of every data endpoint
- Integration of existing vocabularies
- Using different data formats
- Connection of clouds and platforms

5 **VALUE ADDING APPS**
Typical tasks can be solved easier with apps
- Processing of data
- Remote execution

6 **DATA MARKETS**
Data is being traded as an asset
- Clearing & billing
- Domain specific broker and marketplaces
- Use restrictions and legal sspects (contract templates, etc.)

2.1.3.1.3.4 ABOUT ML

Data sharing for machine learning purposes has specific challenges, notably that thee way datasets have been created can lead to bias. Hence the need for full documentation of the origin and purpose of machine learning datasets. Different companies have adopted their own solution for metadata of these datasets, such as Google dataset fiche or Microsoft datasheet for datasets, but the lack of a standard limits the scalability of reuse. [97]

This is the rationale behind ABOUT ML (Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifestyles) is an initiative created by the leading AI industrial consortium representing all the main players, with the intention of establishing, encouraging and promoting novel standards for transparency within machine learning systems by way of documentation. This aims to be done by studying best practices from inception to deployment.

The project is broken down into 8 phases: Understand latest research; understand current practice; combine research theory and results of current practice into testable pilots; run pilot test with PAI partners/ organizations (not individually specified); collect data from pilot transparency practices; iterate on pilots with the latest research and practice; when enough evidence has been collected, elevate it to a best practice; and promote effective practices to establish new industry norms for transparency. The partner organizations include: Facebook, Xbook, IBM, Leverhume Centre, Accenture, Quantumblack, Mckinsey & Co.,Future of Humanity Institute, EFF, Future of Privacy Forum, Deepmind, Berkman Klein Center, Tech Policy Lab (UoW), Google, Policy Link, AI Now, Berggruen Institute, Data & Society, Center for Internet and Society, Sony, BBC, UCL, Microsoft, Intel, Vision and Imaging Processing. The main drivers for industry engagement in AboutML are to avoid misuse and harm that arise from ML systems by creating guidelines for transparency

[97] See Margaret Mitchell and others, 'Model Cards for Model Reporting', Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19, 2019, 220–29 https://doi.org/10.1145/3287560.3287596; Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. "Datasheets for Datasets." ArXiv:1803.09010 [Cs], March. http://arxiv.org/abs/1803.09010 [accessed 29 June 2020].

documentation which, if implemented early on, can prevent future harm. Additionally the partnership provides major corporations to gain legitimacy through being associated with one of the first great push for transparency initiatives. It also seeks to insulate major corporations from future challenge by regulators or public opinion by highlighting the influence and guidance of civil society and social justice organizations in the creation of these standards. The partnership annual revenue has been the following: USD 7.25M (2017); USD 10.53M (2018); USD 8.14M (2019) where USD 3.91M spent on all programs including About ML.

2.1.3.1.3.5 eNewGovernance98

aNewGovernance is a Public Private Partnership launched in 2020 designed to support free flow of data in a human centric approach. The partnership seeks to support start-ups, SMEs, corporates, local authorities and governments alike to develop new services based on data re-use with no trust or liability issues. By supporting themselves in already on-going initiatives such as MyData, Fiware or Gaïa-X, aNewGovernance seeks to govern the data landscape to empower users by allowing their interaction with their personal data via technological tools that enable them to enact their right to data portability as claimed in GDPR.

ANewGovernance has the goal to develop an infrastructure that fosters data sharing and enables interoperable data ecosystem. The partnership seeks to ensure that organizations storing the data are not managing the permissions over the data use. Some of the expected outputs by aNewGovernance are: agreed data models, liability model for data sharing, personal data sharing APIs and a common consent or permission layer.

2.1.3.1.3.6 BDVA- Standardization working group

The Big Data Value Association (BDVA) is an industry-driven international not–for-profit organisation with more than 200 members across Europe, which contributes to the implementation of the Big Data Value PPP program. As part of their activities, the organization fosters a wide range of activities to facilitate data sharing across industries. Under these activities, the organization has a task force devoted to foster data standardization (under task force 6). Data standardization activities are estimated a dedication of 2 to 3 hours per week of a person; around 3 to 5 meetings per year with an average of 3 to 6 days of meeting and the correspondent (and usually continental) travel and accommodation costs of such 3-5 meetings. Organizations dedication can go from 1 to 7 people dedicated in the participation in the standardization process. As a result, data standardization activities face incentive problems for companies (in particular SMEs) who need to have a clear business case before engaging in such high-effort-intensive tasks.

2.1.3.2 The problem, its magnitude and the stakeholders affected

The overall problem analysis rests on a set of causal relationship summarized in the table below.

In this specific context, the problem is represented by the suboptimal adoption of data sharing by companies, which leads to lower innovation and productivity in traditional sector as well as in data business. Two of the main barriers to data sharing lie in the limited standardisation of data and metadata, lack of interoperability and trust. While fear and perceived risks of sharing data, reduces the likelihood in data sharing, other factors such as limited standardisation of data and metadata come into place when an organization negotiates access to data. Standardization is in such context a cost reduction strategy. The ongoing initiatives are therefore not designed at stimulating data sharing *per se*, but the setting up of governance mechanisms to support and accelerate

[98] ANewGovernance: https://www.anewgovernance.org/

standardisation, within sectors and across them. In particular, the goal is to facilitate the speeding up and scaling up of the standardisation activities which fall fully under the European strategy for ICT standardisation.[99]

Table 7 - overview of the problem analysis

Ongoing initiatives	Causes	Problem	Effects
Standardisation and coordination initiatives	Lack of data and metadata standards , data schemes within sectors Lack of technical interoperability across sector	Lack of data sharing within/across sector	Lower productivity and innovation

Data sharing among business is increasing but remains below optimal leading to missed economic opportunities. In a 2018 report by Everis, 60% of companies do not engage in b2b data sharing. Deloitte estimates that the vast majority of the benefits expected from IoT data in different sectors by 2027 stems from data sharing, but that data sharing has reached only a minor part of its potential: 32% for horizontal (between competitors) data sharing, 47% for vertical (business in the same value chain) and 31% for data sharing across sectors.[100] In particular, the opportunities stemming from data sharing across sector are remarkable but clearly smaller than data sharing within sector – from one third in the case of manufacturing to about 60% in the case of automotive.

More data are available on scientific data sharing to illustrate the problem. Only 14% of researchers deposit their data on trusted scientific repositories which gather less than 20% of overall scientific data.[101]

This is particularly important for advanced, data intensive machine learning applications. In fact, access to data is the second most frequently mentioned barrier for artificial intelligence in Europe.[102]

The stakeholders affected by the problem are of three types:

- The data holders are any data generating company in Europe. There are 22 million companies in the EU. However, of these 22 million, only around 700,000 are considered genuine "data users" by the EU data market study because of their intensive use of data, increasing to 844,000 by 2025 in the most favourable scenario.[103]
- The data reusers are any other company in Europe, since by nature data spaces allow for peer to peer data sharing between companies. In addition, specific benefits will be drawn by a subgroup of reusers, the technology companies. According to the same study, there are an estimated 280,000 data companies in the EU.
- Intermediaries are composed by dedicated intervention (market or government led) to facilitate data sharing and data standardization. They are difficult to quantify but can be considered in the order of 10 to 100 if we limit ourselves at intervention with visible footprint at EU level.

[99] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. Digitising European Industry. Reaping the full benefits of a Digital Single Market. COM/2016/0180

[100] Deloitte, Realizing the economic potential of machine-generated, non- personal data in the EU.

[101] The Lisbon Council and others, 2019. STUDY ON OPEN SCIENCE: MONITORING TRENDS AND DRIVERS. European Commission

[102] Claire Beatty. The global AI agenda: Europe. MIT Technology Review Insight, 2020.

[103] These data come from the EU data market study. See www.datalandscape.eu

Table 8 - Estimated number of stakeholders impacted in the EU-27

Stakeholder type	Business in traditional sectors	Data companies	Intermediaries	Total
Data holders	22 million companies of which, 700,000 use data extensively			**Approx. 700,000**
Data re-users	The same 22 million companies of which, 700,000 use data extensive	280,000 data companies		**Approx. 1,000,000**
Intermediaries (data sharing and standardisation)			10-100	**Approx. 10-100**

2.1.3.3 The causes of the problem

Barriers to data sharing include the use of varying and non-compatible data standards that make difficult to integrate, aggregate and combine different data from diverse data holders in the ecosystem. These varying data standards are a bottleneck for data reuse. The same study points out that the "lack of common sharing protocols and standards" is one of the main barrier in manufacturing and implies the loss of about 40% of valuable data sharing – mainly when it comes to vertical data sharing among players in the same value chain. On the same tone, a second barrier is that "the cost of normalising data to be shared is high". On the other hand, when standards are implemented in cases such as the OpenActive standards for physical activities, this can result in a visible increase in data sharing – 200,000 new activities were shared after the introduction of the standard, resulting in 150 to 500,000 new activities carried out by consumers per month. The case of OpenActive also shows that the development of standards was only part of the problem solved, but the OpenActive also coordinated work across the sector to communicate the benefits of standard adoption. The case reflects that standardisation involved more than just technical work, but also engagement towards adoption in order to unlock such benefits.

Hence, interoperability sits at the core of data sharing goals and it implies engaging in data standardisation processes, whether bottom-up (industry-led) or top-down via *de jure* Standards Development Organizations (SDOs) mandates.

The challenges of making industries to agree and widely adopt data standards to achieve the desired interoperability for data re-use are substantial. While data standards can potentially ensure that industry sectors are more competitive and support a vibrant ecosystem of innovative new business, the lengthiness, time-consuming efforts by businesses, lack of incentives of for-profit to engage in such standardization process when no clear concrete business case on data re-use is clear, and the complexity to achieve practical and wide consensus, make data standards a real barrier for data sharing. Hence, standardisation needs to be driven by either regulatory intervention that addresses market failures, or to address specific goals across a business ecosystem. If standards are not connected to the goals of the sector then they are less likely to be adopted. In other words: generic standards for data sharing and metadata will be more challenging to develop and adopt than those that support specific use cases. Yet at the same time, there is a clear economic case for greater findability and interoperability of data across sector – this is where most of the benefits will come in sectors such as automotive.[104] It can be argued then that the more generic the data standard, the less likely it is to be entirely market driven. It is worth noting that many of the initiatives for generic,

[104] Deloitte study for Vodafone group, Realising the economic potential of machine-generated, nonpersonal data in the EU, see: https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

scalable data sharing are backed up by government agencies (as it is the case for iShare and SITRA IHAN as well as FAIR for research data).

A second major bottleneck is trust. Companies are reluctant to share data because of the loss of control it implies. Data reuse can harm them from a competitive perspective, by letting other companies understand valuable information about the processes. And it can expose them to legal problems, for instance in relation to GDPR or commercial secrets. In the same study, Deloitte points out that for manufacturing the remaining top barriers are: "Exposing machines to attack and/or inadvertently disclosing commercial secrets" and "Legal procedures will need to replicated for every data-sharing partner, which is time consuming". In the 2018 report by Everis, the 60% of companies which do not engage in b2b data sharing attribute this to privacy concerns, trade secrets, and fear of misappropriation. In another report by Accenture, data breach lead to almost 10 percent-decline in revenue for up to six months after the breach compared to companies who did not suffer a breach. [105]

Another trust related issue refers to the potential risks of reusing datasets for machine learning without a full understanding of their limitation, leading to possible mistakes in decisions as well as discrimination. Datasets useful for some purposes are not for others, because of their inherent limitations. Companies have started to work on "datasheets" similar to those of the electronic devices. Just as a transistor provides information on the range of temperature for its use, the datasheet would provide information on the key limitations of the sample of the datasets.[106] However, we are still at an early stage where each company is starting to provide this information in different ways.

The further away the data travel from the original data holder, the more the need to provide valid documentation to support reuse – mainly through metadata describing the data and additional aspects including the method of collection, the purpose of reuse, the consent of the various data holders (including personal data) but also additional documentation is required beyond metadata, as the cases such as Datasheets show. The more investment in good metadata and documentation, the more reused the data will be, as reflected in the current European work on high value datasets. Yet there is an issue of effort optimization: While datasets that are less likely to be reused, e.g. from small studies or experiments, may not need the same level of documentation, standardisation, data that is expected to be widely re-used needs to appropriate investment on metadata and documentation. A clear example of the relevance of metadata and documentation more broadly comes from scientific data, where the FAIR principles require documentation and formatting that enables the widest reuse by different communities. Based on the experience of the interviewee, it is widely recognized that providing high quality metadata and documentation for scientific datasets requires 5 to 10% of the total project budget – a very substantial expenditure.

In other words, both interoperability and trust issues are solvable. There is a long history of standardization, metadata, interoperability and definition of sharing agreements. But while those activities require time, consensus and coordination efforts across stakeholders, in the meantime present needs for data sharing are mostly solved with bilateral contractual arrangements which do not scale adequately and entail excessive cost and most importantly an opportunity loss for most organizations.

[105] See Everis, 2018, Study on data sharing between companies in Europe, European Commission and Accenture, 2019. Maximize collaboration through secure data sharing

[106] Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2020. "Datasheets for Datasets." ArXiv:1803.09010 [Cs], March. http://arxiv.org/abs/1803.09010 [accessed 29 June 2020]

There is a need of a set of high-level guiding principles to support data standardization efforts that can cut across any data governance attempt. As FAIR principles are currently guiding the data standardization efforts in research and the activities providing the appropriate metadata, data structure and descriptions, some complementary guiding principles could support the spread and fragmented efforts towards data standardization across the different industries. While there is no such a one-size-fits-all approach for data standardization, grounding efforts in general underlying principles can provide a source of clarity across the diverse verticals and application areas.

While there are diverse data intermediaries that have emerged to support or even coordinate some data standardization efforts, yet there co-exist many conflicting and non-interoperable data standards. The cause of the problem is that standards are usually developed locally and vertically to facilitate data sharing. For instance, where Pistoia Alliance or Global Alliance for Genomics and Health set up data standards to support scientific data sharing in biomedical research, when organizations need to aggregate health patient data we move towards what is considered another sector (health) were data standards are agreed by parallel standardization efforts in different SDOs. If we need to move further and for epidemiological purpose aggregate data about the environment then the data standards will be again different and conform to the consensus of 'another' community. In sum, while data standards agreed locally or in a sectorial basis work, they lack the ability to travel beyond and far from its origins, in the absence of some minimal and generic prescriptions of how the structure and data descriptions should look like. The absence of such standardization effort that allows data to travel across sectors impedes innovation and supposes high-cost.

In sum, there is a need to invest in standards development around specific use cases e.g. by sector or broader challenges but also to bridge across environments and sectors. While data standards agreed locally or in a sectorial basis work, we need also standards that make possible for data to travel across sectors far from its origins. The adoption of common principles for standards development, data access and data governance could help make that achievable. Overall, the absence of such standardization effort within but also across sectors impedes innovation and supposes high-cost.

2.1.3.4 The effects of the problem

The effects of the fragmented efforts towards data standardization and the lack of incentives of for-profit to actively engage in such processes, in particular cross-sector, to facilitate data sharing across industries has an effect not only in restricting data-driven innovation but also in incapacity to reap the benefits of data in terms of efficiency and productivity.

In manufacturing alone, data sharing of IoT data is expected to generate - if fully implemented – 1.4 trillion euros in increased productivity by 2027. Another study by the World Economic Forum estimates at 100bn the current opportunity for data sharing in manufacturing.[107]

In the broader context, the historical contributions to Germany economic growth rate attributed to standardization is 0.9%, and for the period from 2002 to 2006 the total economic benefit of standardization averages about 16.77 billion Euros per year.[108]

Additionally, widely adopted data standards contribute to more transparent competition: Not only individuals can benefit by having better products and services at lower prices, but also organizations can benefit by driving up their profits. The collaboration of governments and for-profits is required in such standardization activities to ensure a minimum stack of standards for authentication, consent

[107] WEF, 2020. Share to Gain: Unlocking Data Value in Manufacturing.

[108] Manchester Institute of Innovation Research, 2013. The Impact of Standardization and Standards on Innovation, NESTA.

interoperability, sector-specific API available. Scalable portability and interoperable consent management is important to enable data sharing and some type of standardization or agreement on data formats, descriptions and workflows is needed. Ultimately, by providing general guidelines that inform data standards across sectors may expect to increase the gross value added to the economy through increased productivity.

2.1.4 Establishing a certification framework for data intermediaries

2.1.4.1 Background

2.1.4.1.1 Context

Data intermediaries play an increasingly pivotal role in the thriving data sharing and selling market. The OECD report 'Enhancing Access to and Sharing of Data' defines **'data intermediaries'** as organizations that 'enable data holders to share their data' which 'may also provide additional added-value services such as data processing services, payment and clearing services and legal services, including the provision of standard-license schemes".[109] Additionally, the EC Communication "Building a European Data Economy" of 2017 and its accompanying SWD present data market places as organizations that facilitate data use and exchange and identifies three types of data intermediaries (i.e data marketplaces, industrial data platforms, personal information management services),[110] while the above-mentioned OECD report identifies five types of data intermediaries (i.e data repositories, data brokers, data marketplaces, Personal information management systems/personal data stores, trusted third parties). Other types of data intermediaries include data unions, data cooperatives, data collaboratives and data trusts. The Open Data Institute (ODI) has published a visualisation of the different concepts in use with a proposed clustering.[111] This study, aims to examine the intermediaries that are neutral in the sense that they are at least functionally/organisationally separate from both data holders and data users.

This overarching and very broad definition of data intermediaries encompasses many different types of organisations which can have very different characteristics:

- **They can be focused on personal or non personal data (or both)**: the Staff Working Document[112] accompanying the Communication on Building a EU Data economy[113] for instance distinguishes between 1) non-personal data Industrial Data Platforms (which can be vertical an sectorial like AutoSar[114] or community led and horizontal like FiWare[115]) and 2) Personal Information management services (such as generic solution providers like Mydex[116], digi.me[117], Meeco.me,[118] Polypoly[119] or sector specific solution providers like MiData Cooperative[120] etc.). However, the line between personal and non-personal data intermediaries is liked to become more blurred in the future and for certain specific sectors (i.e. health and automotive).

[109] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2
[110] COM(2017)9 and SWD(2017)2
[111] https://theodi.org/project/the-data-access-map/
[112] COMMISSION STAFF WORKING DOCUMENT on the free flow of data and emerging issues of the European data economy, Accompanying the document Communication, Building a European data economy, 2017, https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy
[113] Communication on Building a European Data Economy, 2017, https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy
[114] https://www.autosar.org/
[115] https://www.fiware.org/about-us/
[116] https://mydex.org/
[117] https://digi.me/
[118] https://meeco.me/
[119] https://polypoly.eu/en/home
[120] https://www.midata.coop/en/home/

- **They can be completely independent from data holders or stem from data holders' initiatives**: in some cases, data intermediaries are established by data holders in order to enhance the access to their data, control how their data is being used and gain access to third parties' data. These types of organisations can be found within many different value chains. For instance, in the railway sector, a specific "data space" has recently been established by railway infrastructure service providers, original equipment manufacturers, railway operators and other stakeholders to pool together everybody's data and securing data exchange while maintaining data sovereignty.[121] Other examples of industrial data platforms developed by big industrial players include among others Mindsphere (Siemens), Skywise (Airbus), RIO (Traton Group), Predix (GE Digital), FieldView, Xarvio, as well as the Data Intelligence Hub of Deutsche Telekom and Radianz of BT Group.
- **They can provide only access to data or services on top of the data**: as the OECD suggests, some data intermediaries provide services on top of the data and they specialise in offering data storage or access management features to their clients. This is the case for instance of data intermediaries like Nallian[122] which provides standard license schemes for sharing the data uploaded on the platform as well as the possibility of plugging in applications for smart billing and smart auditing.[123]
- **They can be well-established players with a long history of providing data or start-ups and newly established businesses:** some industries and especially the financial industry are used to the existence of big data brokers such as Bloomberg, Thomson Reuters, etc. and some of these players date back of several decades. For other industries, such as the aerospace or automotive industries to name two, these players are new and respond to the changes brought by the data and internet of things economy.
- **They can be profit driven or not**: the OECD mentions that, on top of business driven data intermediaries, public data repositories such as those set up by public libraries or scientific communities can also be considered as data intermediaries[124].

These and other differences make data intermediaries a very heterogeneous category of players. However, their common characteristic lies in their role of **matchmakers between demand and supply of data**. For this reason, they are also sometimes called "data marketplaces" or even "data brokers" although these terms are also used to identify more specifically certain types of data intermediaries. According to the Summary report of the open public consultation on the European strategy for data, almost 60% of the 772 respondents to this section considered that emerging novel intermediaries, such as 'data marketplaces' or 'data brokers', are useful enablers to the data economy, while almost 22% don't know or remain neutral to the question.[125]

2.1.4.1.2 Ecosystem

The data-sharing ecosystem includes various types of stakeholders involved in the value chain of the data intermediaries, including in particular data holders, data re-users and (certified) data intermediaries.

The **data holders** in this value chain are the data providers sharing their data with the data users through the certified intermediaries. The **data intermediaries** in this value chain are the enablers

[121] https://www.internationaldataspaces.org/knorr-bremse-establishing-data-sovereignty-and-data-ecosystems-in-the-rail-industry/

[122] https://www.nallian.com/solution/how

[124] OECD, Enhancing Access to and Sharing of Data, Reconciling Risks and Benefits for Data Re-use across Societies, 2019, http://www.oecd.org/going-digital/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[125] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

of data sharing between data holders and data re-users. The **data re-users** will be the clients of the data intermediaries.

The table below provides three different approaches of the main stakeholders identified for this domain, as potentially affected from the establishment of a certification framework for data intermediaries: a) a generic approach, b) the B2B data-sharing scenario and c) the C2B data-sharing scenario.

Table 9 - Stakeholder scope (data value chain mapping)

Domain	Approach	Data holder	Data (re)user (whole dataset)	Intermediaries	Personal data involvement	Purpose
Establishing a certification framework for data intermediaries	Generic Approach	Data Providers: Businesses/ Academia and research organisations /Governmental Organisations / NGOs/ Citizens	Intermediaries' Clients: Businesses/ Academia and research organisations /Governmental Organisations / NGOs/ Citizens	Certified data Intermediaries (i.e. Data marketplaces, data brokers, data repositories, PIMS/PDS, industrial data platforms, trusted third parties, data unions, data cooperatives, data collaboratives, data trusts)	Potentially	Business, R&I, Public Good
	B2B Approach	Data providers: Businesses	Intermediaries' Clients: Businesses	Certified data intermediaries: Data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts	No	Business, R&I
	C2B Approach	Data providers: Citizens	Intermediaries' Clients: Businesses	Certified data intermediaries: PIMS/PDS, data unions, data cooperatives, data collaboratives, data trusts	Yes	Business, R&I, Public Good

2.1.4.1.3 Data holders

For the domain on establishing a certification framework for data intermediaries, data holders will be the data providers who will be enabled to share their data through the data intermediaries. In a generic approach, the types of data providers might vary depending on data intermediary category and could include businesses, academia and research organisations, NGOs or citizens. **This study**

focuses on two different data sharing scenarios: a) Business to Business data sharing (B2B) and Consumer to Business data sharing (C2B).

In the B2B scenario, businesses is the most common source of data for certified data intermediaries like data marketplaces, industrial data platforms, trusted third parties, data trusts and data collaboratives. In the C2B scenario, individuals are the data providers for certified data intermediaries like personal information management service (PIMS) or personal data stores (PDS), data unions, data cooperatives, data collaboratives and data trusts.

2.1.4.1.4 Data Intermediaries

- Depending on the type of data sharing **Data marketplaces**: There is no uniform definition of what a data marketplace is. The Commission in its 2017 Staff Working Document *(SWD(2017)2)* followed the definition of Stahl et al. describing data marketplaces as electronic marketplaces where data is traded as a commodity, an electronic marketplace being "the concrete agency or infrastructure that allows participants to meet and perform the market transactions, translated into an electronic medium",[126] while the OECD report 'Enhancing Access to and Sharing of Data' defines them as online platforms that host data from various publishers and offer the (possibly enriched) data to interested parties.[127] Finally, a Forrester research report entitled "The Insights Professional's Guide to External Data Sourcing, Beginner Level: Data Practices For Insights-Driven Businesses" defines data marketplaces are data exchanges that enable sellers to offer data products and services and enable buyers to find and acquire data, often as a self-service, transactional model.[128]
- **Industrial Data platforms** are defined in SWD(2017)2 as virtual environments facilitating the exchange and connection of data among different companies and organisations through a shared reference architecture, common governance rules and within a secure business ecosystem.[129]
- **Trusted third parties / Data intermediary acting as a third-party certification authority:** According to the OECD report 'Enhancing Access to and Sharing of Data' in some cases, data intermediaries can act as a certification authority as in the case of the Industrial Data Space (IDS). The certification authorities of the IDS certifies all participants based on standards defined by the IDS regarding, for example, security, privacy, and terms of use. Data owners define terms of use and the fees of data use, which data brokers use to match with other data owners and users.[130] Participants and core components shall provide a sufficiently high degree of security regarding the integrity, confidentiality and availability of information exchanged in the Industrial Data Space. Therefore, an evaluation and certification of the core components as well as of the technical and organizational security measures is mandatory for participating in the Industrial Data Space. This applies to both organizations that develop software components intended to be deployed within the Industrial Data Space (i.e., prospective software providers) and to organizations that intend to become participants in the Industrial Data Space. During the

[126] SWD(2017) 2, p. 17
F. Stahl, F. Schomm, G. Vossen, & L Vomfell, A Classification Framework for Data Marketplaces, Vietnam J Comput Sci, 2016, p. 137.
[127] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2 (p.36)
Dumbill, E. (2012), Microsoft's plan for Hadoop and big data, http://radar.oreilly.com/2012/01/microsoft-big-data.html
[128] Forrester research, The Insights Professional's Guide to External Data Sourcing, Beginner Level: Data Practices For Insights-Driven Businesses, May 2019
[129] SWD(2017) 2, p. 18
IDC and Open Evidence, European Data Market Study, 2016, publication forthcoming, https://docs.google.com/a/open-evidence.com/viewer?a=v&pid=sites&srcid=b3Blbi1ldmlkZW5jZS5jb218ZG93bmxvYWR8Z3g6NjJiZTQ1NTYyZjdlOGNhNg
[130] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2

certification process, the primary focus of the evaluation will be either on the product or on the organization itself.[131]

- **Data collaboratives** are defined as a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors—including in particular private companies, research institutions and government agencies- exchange their data to solve public problems and create public value.[132]
- **Personal information management services (PIMS) and personal data spaces[133]** : The OECD report 'Enhancing Access to and Sharing of Data' defines PIMS/PDS as platforms to give data subjects (consumers) more control over their personal data and thus to restore user agency, including in the context of the Internet of Things.[134] The SWD(2017)2 further defines (PIMS) as "a developing set of technical means, currently in its infancy, for individuals to manage control over their personal data. While considerable conceptual differences exist, PIMS can be summarised as technical means which individuals can use in order to exercise their right to data portability under article 20 GDPR. PIMS in this respect can serve as a means to receive back personal data from data controllers (within the limits of the right under article 20 GDPR). PIMS would then also give individuals the means to provide personal data through a web or mobile application for processing by others on the basis of one of the legal bases of the GDPR (e.g. consent, performance of a contract)".[135]
- **Data unions:** A Data Union is a framework, currently being built on the Streamr Marketplace that allows people to easily bundle and sell their real-time data and earn revenue. On its own, our data does not hold much value, but when combined in a Data Union, it aggregates into an attractive product for buyers to extract insights. This is crowdselling, and has the potential to generate unique data sets by incentivising trade directly from the data producers.[136]
- **Data cooperatives:** Similarly to the above mentioned data unions, data cooperatives can be defined as structures that enable the creation of open data and personal data stores for mutual benefit; they could rebalance what many perceive as asymmetric relationship between data subjects (people with personal data) and data users (people who use data to develop services and products)[137]
- **Data trusts:** The ODI defines data trusts as legal structures that provide independent, fiduciary stewardship of data. Data trusts are an approach to looking after and making decisions about data in a similar way that trusts have been used to look after and make decisions about other forms of asset in the past, such as land trusts that steward land on behalf of local communities. They involve one party authorising another to make decisions about data on their behalf, for the benefit of a wider group of stakeholders. With data trusts, the independent person, group or entity stewarding the data takes on a fiduciary duty. In law, a fiduciary duty is considered the

[131] https://www.internationaldataspaces.org/publications/whitepaper-certification/

[132] https://ec.europa.eu/knowledge4policy/online-resource/data-collaboratives_en & https://oecd-opsi.org/toolkits/data-collaboratives-canvas/ & http://thegovlab.org/the-emergence-of-data-collaboratives-in-numbers/

[133] Also defined by other terms, including among others personal data stores (PDS)/vaults/wallets/clouds or infomediaries, vendor relationship management tools, life management platforms, personal information management systems; information fiduciaries, mediators of individual data - MID, information banks. [source Understanding MyData Operators White paper, https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf]

[134] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2
Urquhart, L., N. Sailaja and D. Mcauley (2017), "Realising the right to data portability for the domestic Internet of things", Personal and Ubiquitous Computing, http://dx.doi.org/10.1007/s00779-017-1069-2.

[135] SWD(2017) 2, p. 19

[136] https://medium.com/streamrblog/what-are-data-unions-how-do-they-work-which-ones-can-i-use-887e67fb7716

[137] http://opendatamanchester.org.uk/2015/04/14/open-data-cooperation-building-a-data-cooperative/ & https://medium.com/@opendatamcr/open-data-cooperation-building-a-data-cooperative-264eef373b63

highest level of obligation that one party can owe to another – a fiduciary duty in this context involves stewarding data with impartiality, prudence, transparency and undivided loyalty.[138]

2.1.4.1.5 Data re-user

For the domain on establishing a certification framework for data intermediaries, the data (re-)users will be the data intermediaries' clients. In a generic approach of the value chain, the client base of a data intermediary could entail various categories of organizations including businesses (e.g. buyers and suppliers), academia and research organisations, NGOs, public sector organisations and citizens. The category of data re-users vary according to the type of data intermediary and the services provided. In the B2B and C2B scenarios in the context of this study, the client base of the intermediaries will be mainly comprised of businesses. In particular, for industrial data platforms (B2B data platforms) the client base will most likely be comprised of businesses such as Original Equipment Manufacturers (OEMs) first and second tier buyers and suppliers.

2.1.4.1.6 Ongoing initiatives/Market analysis

Non-exhaustive listing, providing examples of B2B and C2B data intermediaries, active in the European market is presented in the tables below.

Table 10 - Data Intermediaries - B2B Data Sharing European Market Overview

Data Marketplaces	Industrial Data Platforms	Trusted Third Parties	Data Collaboratives (B2B Data Sharing)	Data Trusts (B2B Data Sharing)	Other B2B Data Sharing Operators
Dawex	Mindsphere (Siemens)	International Data Spaces Association	Industrial Data Space Project (German Federal Ministry of Education and Research-BMBF) - International Data Spaces Association	OpenCorporates	Ocean Protocol
DataPace	Skywise (Airbus)	Smart Connected Supplier Network (SCSN)	Amsterdam Data Exchange (AMDEX)	Truata	Refinitiv
Streamr	RIO (Traton Group)		DeepMind & NHS Machine Learning for Health		Meeco.me
OpenDataSoft	Predix (GE Digital)		Data and Analytics Facility for National Infrastructure (DAFNI)		
Databroker DAO	FieldView		Smart Connected Supplier Network (SCSN)		
Rocketgraph	Xarvio		SmartFactoryKL		

[138] https://theodi.org/article/what-is-a-data-trust/

Smart Jobs S.L	Data Intelligence Hub (Deutsche Telekom)
Spaziodati	Radianz (BT Group)
WhoApi	Nallian
City Context Open Data API	AutoSar
Datalayer	FiWare
DataScouts	Far-edge
dmi.io	Arrowhead
GLOBMOD	
Helix Nebula Science Cloud	
Open Corporates	
qDatum	
Advaneo	
Caruso	
The IOTA Foundation	
Kasabi	
Datafairplay	

Table 11 - Data Intermediaries - C2B Data Sharing European Market Overview

PIMS/PDS	Data Unions	Data Cooperatives	Data Trusts (C2B Data Sharing)	Data Collaboratives (C2B Data Sharing)	Other Personal Data Operators
Digi.me	Streamr	MiData	UK Biobank	SalusCoop	Meeco.me
Mydex	The Data Union	SalusCoop		Copenhagen's City Data Exchange	Vastuu Group
CitizenMe	Swash	Holland Health Data Cooperative		Grampian Data Safe Haven (DaSH)	Peercraft
Datawallet	Tracey Project - TX/ WWF Philippines/ UnionBank/ Streamr Partnership	The Good Data Cooperative		Consumer Data Research Centre	Criteo
Schluss		Polypoly		Decode	Worker Info Exchange
Qiy Foundation		Healthbank Cooperative		Next Generation Internet - Engineroom	Digita

Polypoly	Datavillage
Solid Inrupt	Happy-Dev
BitsaboutMe	Ontola
Coelition	1001 Lakes
Comuny GmbH	Business Finland
Cozy Cloud	Caelum Labs
Datafund	City Of Oulu
DataYogi	Conseils Oy
esatus AG	de Volksbank
Ockto B.V.	Fair & Smart
OwnYourData	MyLife Digital
iGrant.io	Diabetes Services ApS
	Demos Helsinki
	Enfuce
	Electronic Frontier Finland

2.1.4.2 The problem, its magnitude and the stakeholders affected

The lack of a certification framework for data intermediaries, or more generally of mechanisms to differentiate neutral data intermediaries from the others leads to two main clusters of problem that are coming to prominence. The first cluster of problems involve misuse and overuse of data; the second set of problems involves underuse of data.[139] In both cases, this further leads to a generic lack of trust between the actors involved in the data intermediaries' ecosystem. As a final consequence, a fair and well-functioning market level playing field at European level is not ensured.

Additionally, according to the summary report of the open public consultation on the European strategy for data, almost 80% of the 512 respondents to the question have encountered difficulties in using data from other companies. These difficulties relate to technical aspects (data interoperability and transfer mechanisms), denied data access, and prohibitive prices or other conditions considered unfair or prohibitive. A very large share of respondents (87.7%) supported the idea that the EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate. Around the same proportion of respondents considered that the development of common European data spaces should be supported by the EU in strategic industrial sectors and domains of public interest. [140]

2.1.4.2.1 Estimation of Stakeholders affected

The wide definition of data intermediaries used for this study, and their several different categories constitute difficult the calculation of the total number of stakeholders affected. An estimation of the total number of data intermediaries active in the European market could include an average number of150 organisations, while the number of data users or data holders affected could entail any European company or individual wishing to buy or sell data through the intermediaries.

[139] https://medium.com/@vincejstraub/the-new-ecosystem-of-data-trusts-36901fc59010

[140] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

The companies present big differences in the scale of client base. In particular, Siemens' Mindsphere counted more than 6.100 customers in March 2020; the client base of the late-stage Dawex includes approximately 10,000 organisations; the example of the data trust UK Biobank holds data from about 0.5m people and it includes the number of 946 researchers using its data in its annual accounts of 2018. This would therefore give a ratio of roughly 50000:1:1000 (data holders : data intermediary : data re-users). At the same time, there are several data intermediaries of early or growth stage in the European market with a client base of less than 100 clients.

2.1.4.3 The causes of the problem

Currently, different rules and legislation might apply to data intermediaries in Europe, depending on their category, country of establishment, sector of activity, functionalities offered and use cases and type of data handled. This might often create legal uncertainties and generate burdens to the cross-border activities of data intermediaries. Furthermore, given that the appearance of the majority of data intermediaries has recently happened within the last decade, such companies, being still in early or growth stage, might lack incentives to align on best practices. Finally, there is also a lack of mechanisms for data intermediaries to assess the quality and neutrality of intermediaries' operations, creating a further lack of trust within the market. These barriers might create difficulties in the establishment a common certification framework of data intermediaries, covering all types and market needs.

2.1.4.4 The effects of the problem

The lack of a certification framework for data intermediaries and its interwoven lack of trust between the actors involved in this ecosystem presents various effects and impacts for the stakeholders affected. The intermediaries respecting already certain neutrality requirements present no competitive advantage in the market compared to the others due to the lack of mechanisms for their clients to assess their neutrality. Furthermore, there are currently no mechanisms that could support such data intermediaries to scale up, as many of them are in early or grow stage. As a broader impact, the economic and societal value of data is not maximised in the European market.

An overview of the above-mentioned intervention logic is presented in the following table.

Table 12 - Intervention logic for Establishing a certification framework for data intermediaries

Measure	Barriers	Problem	Broader Impact
Establishing a certification framework for data intermediaries	Different rules applying to data intermediaries depending on sectors and types of data handled (creating legal uncertainty)	A fair and well-functioning market level playing field is not ensured, due to the lack of trust between the actors involved in the ecosystem, which does not allow data intermediaries to scale up.	Economic and societal value of data is not maximised
	Lack of mechanisms for data intermediaries clients to assess the quality of intermediaries' operations (i.e. in terms of respect of GDPR or other legislation)		
	Lack of incentives for data intermediaries to align on best practices		
	Different categories, business models, functionalities offered and use cases of data intermediaries active in the		

Measure	Barriers	Problem	Broader Impact
	European market, creating difficulties in the establishment a common framework covering all types.		

2.2 Policy objectives and policy options

This section contains a description of the policy objectives, which could be pursued in relation to the barriers, problems and effects identified above. It also presents a list of relevant policy actions.

2.2.1 Policy objectives

The general objective of this initiative is to **set the foundations of a Single Market for Data**. This will contribute to **maximising the potential of data for the EU economy and society**, in particular through the empowerment of the individuals and businesses with respect to the use data they generate and create value for society. This vision will be implemented through the creation of **common European data spaces** in strategic sectors and domains of public interest, and will contribute to a more **rapid recovery** from the current economic crisis.

Across spaces, companies, public sector bodies, researchers and individuals themselves should be able to use data, personal as well as non-personal data, irrespective of the sector, domain or Member State, **in line with EU rules and fundamental values**, in particular personal data protection, consumer protection legislation and competition law.

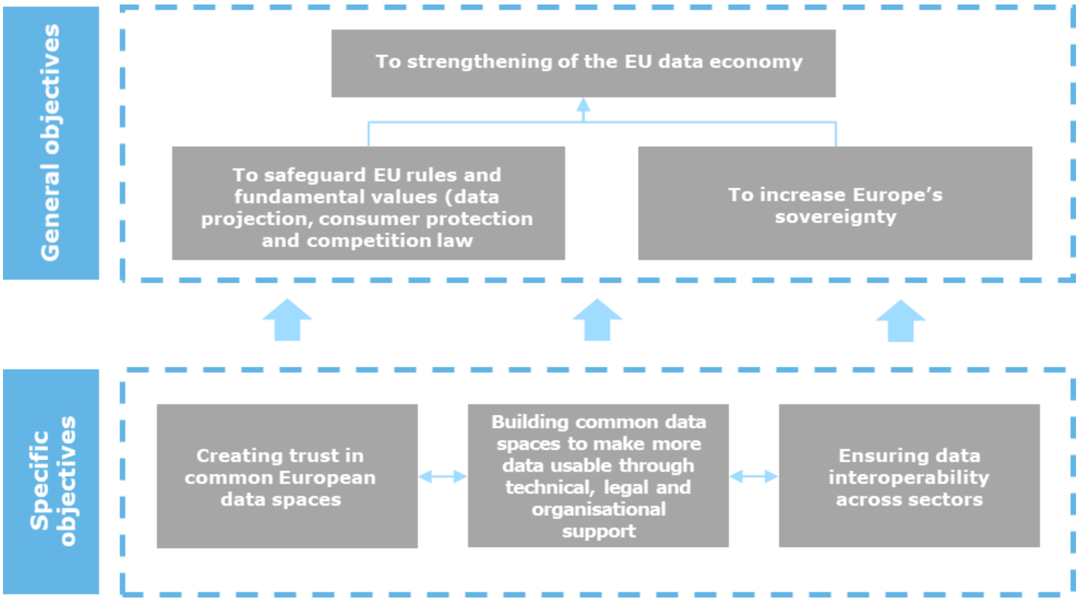
On the global stage, the Single Market for Data will **increase Europe's sovereignty** on data and on all the key enabling technologies and infrastructures that are essential for the data economy. It will underpin a new **European approach** to data as an alternative to the platform model.

This Single Market for Data is an important element that will **complete the EU internal market**, increasing growth and jobs, modernising public services, empowering citizens to exercise their rights, and **accelerating innovation** as data is more widely used for the **common good**.

To reach this general objective, this initiative has **three specific objectives**:

- Creating **trust** in common European data spaces;
- Building **common data spaces**, making more data usable where data holders could agree to it through technical, legal and organisational support; and
- Ensuring data **interoperability** across sectors.

Figure 2 – Objective tree


General objectives
To strengthening of the EU data economy
To safeguard EU rules and fundamental values (data projection, consumer protection and competition law
To increase Europe's sovereignty
Specific objectives
Creating trust in common European data spaces
Building common data spaces to make more data usable through technical, legal and organisational support
Ensuring data interoperability across sectors

In line with these objectives, there are three levels of data governance addressed by the first phase of this study:

Table 13 – Levels of data governance

Levels	Measures facilitating secondary use of sensitive data held by the public sector	Establishing a certification scheme for data altruism mechanisms	Establishing a European structure for governance aspects of data sharing	Establishing a certification framework for data intermediaries
Trust in common data spaces	X	XX	X	XX
Reusable data (technical/legal)	XX	X	XX	X (certification of neutrality of intermediaries)
Cross data space interoperability	X	X	XX	X

These layers can be considered subsequent levels of enhanced abstraction and scalability of providing an enabling environment for data use in the data economy.

Trust in common data spaces deals with ensuring data is available for reuse. This mainly deals with ensuring that the appropriate management of the rights of different stakeholders. Whether data is held by public sector, citizens or businesses, this layer deals with the appropriate rights of data processing. For personal data, including sensitive data, it concerns making sure that consent and other forms of legitimate data access and reuse are ensured.

The second level concerns scalable, reusable tools for data sharing. That means, in addition to the availability of data, ensuring that there are common rules and methods for accessing and reusing. This covers standards for data schemes, for metadata, for architectures, for consent sharing, for the certification of data intermediaries. There are many such examples within different sectors.

The third level is more general and abstract. It refers to ensuring interoperability rules and standards for reuse among a wide variety of actors and use cases, across different sectors.

2.2.2 Policy options

Policy options have been developed in close cooperation between the consortium and DG CNECT. The policy options are listed per domain below:

2.2.2.1 Policy options: Measures facilitating secondary use of sensitive data held by the public sector

The table below lists the policy options developed for the domain on use of sensitive data held by the public sector.

Table 14 - Policy options Measures facilitating secondary use of sensitive data held by the public sector

Option	Description
Option 0	Baseline scenario - No horizontal action at EU level
Option 1	Coordination at EU level and soft regulatory measures only: Guidelines
Option 2	Regulatory intervention with low intensity: One-stop shop
Option 3	Regulatory intervention with high intensity: Single data authorisation body

2.2.2.1.1 Policy Option 0: Baseline scenario- No horizontal action at EU level

In the baseline scenario, **no horizontal action is taken at EU level** on data governance and interoperability of common European data spaces. However, action may be taken at sectoral or Member State level as announced in the European Strategy for Data. This will lead to further interoperability issues and regulatory fragmentation in the internal market. Only certain cross sector data sharing will happen in limited cases between those common European data spaces that have compatible sectoral legislation, standards and infrastructures. Ultimately, there will be less data available for reuse across sectors. This will prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general. In addition, this situation could result in unnecessary duplication of efforts (and costs) among, for instance, data holders in different Member States in the setting up of data sharing infrastructures.

2.2.2.1.2 Policy Option 1: Guidelines

The first policy option would consist of issuing **non-binding Recommendations or guidelines** encouraging Member States to set up structures enhancing the re-use of publicly held data subject to the rights of others such as rights under GDPR, but also intellectual property rights, and legitimate interests to keep commercially sensitive information private. These Recommendations or guidelines would identify best practices (for instance, FinData, the French Health Data Hub or the German Forschungsdatenzentren) and promote their emulation by Member States. In addition, a network of data sharing experts would be set up as an informal Commission Expert Group issuing technical guidance for cross-border and cross-sectoral data sharing – for instance on interoperability issues, generic standards or metadata descriptions.

This option would contribute to some of the action’s objectives: setting up structures enhancing the reuse of data held by the public sector and the use of which is subject to the rights of others may facilitate (and therefore likely increase) the re-use of data by companies, and researchers in line with applicable legislation. In addition, an Expert Group working on cross-border and cross-sectoral data sharing would be a first step towards interoperability across sectors.

2.2.2.1.3 Policy Option 2: One-stop shop

The second policy option would consist of a Directive or a Regulation requiring Member States to set up and/or maintain **capacity and services to facilitate the re-use of publicly held data** subject to the rights of others. These services would notably include a **one-stop shop** that would:

- Orientate re-users to the relevant data holders (i.e. provide information and guidance to re-users on whom to speak to); and
- Provide technical and legal advice to data holders on the permissible uses of such data and on de-identification of data.

Member States would be required to set up **secure processing environments** for the reuse of data the use of which is subject to the rights of others. Member States would have the possibility to either set up a single data processing environment, or to mandate each data holder to maintain its own.

This policy option would **not entail a right for re-users to access publicly held data** subject to the rights of others. However, there would be a best effort obligation to support innovative uses of sensitive public sector data. The use of these services and of that data would be limited to **entities established in the EU,** and potentially to entities located in third countries offering comparable mechanisms (whether this is the case in a third country would be determined by the Commission). Lastly, exclusive arrangements for data not covered by the Open Data Directive would be prohibited to ensure a **level playing field** among re-users.

2.2.2.1.4 Policy Option 3: Single data authorisation body

The third policy option would consist of a Directive or a Regulation requiring Member States to set up a **single data authorisation body** that would:

- Assess and grant (or reject) data re-use requests on behalf of data holders (although this may legally require the centralisation of different registers);
- Provide a secure data processing environment and data analytics tools for the re-use of publicly held data subject to the rights of others;
- Put re-users in contact with data holders (i.e. provide information and guidance to re-users on whom to speak to); and
- Provide advice to data re-users on the procedures to request a data re-use permit and on the likelihood of success of such requests.

These data authorisation bodies should not diverge excessively across the EU, for instance in terms of operational procedures and basic rules on accessibility of data. In addition, the re-use of data would not be limited in terms of purpose, and so **commercial purposes would be allowed** also. Lastly, exclusive arrangements for data not covered by the Open Data Directive would be prohibited to ensure a **level playing field** among re-users.

This policy option would neither, however, require Member States to reorganise competences internally among different data holders, nor make legislative changes to rules on secondary use of data. Member States would remain free to organise their registers, responsibilities and the grounds on which data can be reused as they see fit.

This policy option would contribute to the policy objectives of this action: as for PO2, it would facilitate (and therefore likely increase) the re-use of data by companies, and researchers in line with applicable legislation, and may contribute – through the support and advice it would provide – to creating trust between re-users and holders. In addition, by providing a secure data processing environment and data analytics tools, it would ensure interoperability of data across sectors, thus making data more usable and contributing to building common data spaces.

2.2.2.2 Policy options: Establishing a certification/authorisation scheme for data altruism mechanisms

The table below lists the policy options developed for the domain on data altruism mechanisms.

Table 15 - Policy options Establishing a certification/authorisation scheme for data altruism mechanisms

Option	Description
Option 0	Baseline scenario- no horizontal action at EU level
Option 1	Coordination at EU level and soft regulatory measures only
Option 2	Regulatory intervention with low intensity
Option 3	Regulatory intervention with high intensity

2.2.2.2.1 Policy Option 0: Baseline scenario- No horizontal action at EU level

The baseline scenario assumes costs and benefits of the future if the situation was to remain the same as it is today. In the current situation, there is no European data altruism scheme. This implies that each Member State may explore different possibilities to enable data altruism individually or in cooperation with other Member States. This includes different infrastructures and approaches including legal and governance aspects in Member States. The discussion around the future role of data donors in the management of their data to support data exchange between different parties for a variety of purposes, among them research, public policy usage, public access to official documents or generally to increase efficiency and save transaction costs in one or more specified contexts, is fragmented as well. The European Data Strategy issued in February 2020 by the European Commission highlights the importance of a unified and coherent approach towards a shared data economy. There are several problems resulting of the status quo. First, the lack of European alignment on data altruism scheme leading to multiple and independently developed schemes that could face interoperability issues in the future. Second, a fragmented approach will lead to regulatory fragmentation in the internal market, where data sharing for altruistic motives will be limited to a multitude of separated silos, each acting as isolated data spaces with compatible legislation, standards and infrastructure, but without any realistic option for data to break out of that silo. Third, the rapid growth of data production and sharing which – without a common approach – cannot benefit the public sector. Fourth, the lack of a data donations for research purposes that could hamper innovation, including the development of for example AI, and impact other sectors in the European Union with, consequently, , a negative impact on the EU competitive advantage. Overall, this could prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general[41].

2.2.2.2.2 Policy Option 1: Coordination at EU level and soft regulatory measures only

The European Commission and Member States could explore mechanisms that encourage Member States to collaborate on efforts to sharing of personal data[42]. To facilitate this, the European Commission could adopt a Recommendation or guidelines, with no binding power, to address coordination and cooperation issues with regards to data altruism schemes and ethical guidelines on data use, considering (where applicable) the *Ethics guideline for trustworthy AI,* the “Ethics of information and communication technologies” opinion of the European Group on Ethics in Science and New Technologies, or the European Code of Conduct for Research Integrity, and the GDPR (among other authoritative sources). The recommendations could address MS to establish structures to increase reuse of publicly held data, to support data altruism, and to create processes to lower the transaction cost of data sharing. As a supporting initiative, the European Commission can also set up an informal data sharing expert group, coordinated by the European Commission. This group would be tasked with issuing more detailed guidance on multiple topics (e.g. metadata, semantics, pseudonymisation techniques, equal and non-discriminatory access, the role and rights of the individual, compatible further processing, etc.); all of these guidelines could increase cross-border data sharing on a voluntary basis. Participation in this expert group would be voluntary and open for Member States and subject experts. The expert group could also assess if Member States are interested in trainings and funding, and provide proposals for such further support mechanisms.

Voluntary coordination could be organised at a general and horizontal level – i.e. focusing on the definition of universal principles for data altruism that would be valid independent of the sector or nature of the data – or could be integrated at a vertical sector, thus taking into account the specificities and sensitivities that may be present in individual situations.

2.2.2.2.3 Policy Option 2: Regulatory intervention with low intensity

The European Commission is actively engaging with the European community to advance the data economy and define a path towards a data market. To achieve this the European Commission could pursue regulatory intervention with low intensity such as mandating Member States to establish legislation and or administrative processes that allow data altruism within the Member State, without constraining them too much with respect to the practical approach to be followed. Furthermore, the Commission could (i) oblige Member States to set up certification schemes for data altruism mechanisms and/or organisations offering such mechanisms, (ii) such certification could be issued by private certification bodies under a specific Member State mandate and (iii) encourage voluntary certification of data altruism schemes. These measures would provide for a stronger and more homogeneous governance layer over the related data altruism schemes, thereby providing a more forceful and consistent response to some of the challenges described above. The responsibility to oversee this certification process would be of the Member States.

2.2.2.2.4 Policy Option 3: Regulatory intervention with high intensity

In 2016, the European Commission implemented the General Data Protection Regulation (GDPR) to protect, amongst others, citizens' data against unlawful (re)use. It contains safeguards against re-use of personal data for different purposes than those which were originally communicated to the data subject (i.e. the citizen); this principle can act as a complicating factor for data altruism, as has already been explained above.

The European Commission has increased the availability and re-usability of public and publicly funded data into the scope of the Open Data Directive[43] and while this Directive is likely conducive to supporting data altruism, it leaves the safeguards and constraints of the GDPR intact. A directive or regulation regarding data sharing (including but not necessarily limited to personal data) could facilitate data altruism, in several ways. One approach could be to introduce a tightly limited mandatory European authorisation mechanism for altruism schemes in relation to certain types of data (e.g. data generated or collected using government funding) or for certain purposes (e.g. donation of certain medical records to academic research institutions under specific constraints). Such authorisation would be issued under the auspices of a dedicated national authority, with mutual recognition mechanism between Member States. In some circumstances, it could be made compulsory to certify data altruism mechanisms and/or organisations operating such mechanisms. The responsibility to oversee this certification process would be on the Member States.

Alternatively, a more open approach could be considered, e.g. by establishing a governance structure at the national or EU level that would generically permit data altruism schemes to be established provided that certain safeguards are met. These safeguards could include the establishment of independent supervisory bodies and/or monitoring bodies to oversee compliance with the schemes and their use in practice; certification of schemes and/or technologies, platforms or infrastructures which would be used for data donations; codes of conduct that beneficiaries of such schemes (i.e. the recipients of donated data) or intermediaries in charge of a scheme or a technology would need to sign up to in order to become eligible for data altruism; and/or the establishment of auditing and verification mechanisms accompanied by credible sanctions in case of violations of the framework. Such a framework could facilitate data altruism by providing greater clarity and legal certainty on the conditions for lawful data altruism, including the role and rights of the donors and other stakeholders. Scoping is however critical in legislative interventions: without suitable constraints, mandatory data donation would likely lead to citizen objection. The Commission should therefore consider how to protect citizens' personal data in altruism schemes by defining data ethics requirements.

2.2.2.3 Policy options: Establishing a European structure for certain governance aspects of data sharing

The table below lists the policy options developed for domain on governance aspects of data sharing.

Table 16 - Policy options Establishing a European structure for certain governance aspects of data sharing

Option	Description
Option 0	Baseline Scenario.
Option 1	Coordination at EU level and soft regulatory measures only – Informal Expert Group
Option 2	Regulatory intervention with low intensity: Formal Expert Group - European Data Innovation Board
Option 3	Regulatory intervention with high intensity: Independent European body - European Data Innovation Board

The main barrier to business data sharing lies in the lack of interoperability and scalable trust mechanisms. Simply put, companies are reluctant to share data because of the risks and the lack of control that it entails. And even when they are willing to do it, they often lack interoperable schemes and protocols to exchange data with other companies.

Solving the barriers of interoperability and trust at scale is the objective of a wide range of measures with different degrees of maturity, as illustrated below.

Option 0 includes no action and maintaining the baseline scenario. Option 1, 2 and 3 refer to the creation of an entity (informal, formal or with legal personality) to support data standardisation. The difference lies in the formal requirements and level of engagement, but the type of initiatives covered are similar in the different options:

1.3.0. Sector based data standards, such as those developed within different sectors. This is a mature area and falls outside of the scope of the policy intervention, but is added here to clearly distinguish it from the following three points

1.3.1. Metadata standards on findability and data quality for machine learning, such as those bring developed by AboutML

1.3.2. Data sharing schema such as iShare, IDSA and IHAN

1.3.3. Prioritisation of standardization for data use across sectors/data spaces (interoperability), which at the moment is fairly limited but could be similar to the FAIR principles applied beyond science and potentially include data standards, metadata and data sharing schemes.

In particular, **sectoral data standards** address how data should be formatted and made available to third parties in order to be fully reusable. This is the most basic issue, related to technological interoperability. It is designed to assure that when data are shared, the reuser can immediately make use of them without additional effort and, in the case of open standards, independently from the hardware and software used. Data standards have been in place for decades in different sectors, they have well defined process for standard setting and as such they fall beyond the scope of this initiative, with the exception of cross-sector standards which is part of the last point.

Metadata standards refers to metadata that allow data reuse, both in terms of standardised schema for metadata and standardised conventions on how to describe individual metadata elements (field values), and to the proper reuse, namely in order to understand the limitations and the origin of the data, as well as their purpose of use. This is particularly important when it comes to machine learning, where the data used as an input directly affect the judgements performed by the algorithm. There are many ways to provide this information, from Microsoft datasheet for dataset to Google descriptive fiche which includes many criterion precisely designed to make the dataset bias transparent and manageable. The AboutML initiative is precisely designed "to develop, test, and implement machine learning system documentation practices at scale".

Data sharing schemes are more complex arrangements around data sharing. If data and metadata standards are designed to allow data sharing to happen when companies decide to do it, data sharing schemes aim to increase their propensity to share. They do so by ensuring data sovereignty of the business, reducing the mistrust and reassuring companies about the risks of data sharing – as well as the consent of the individuals. Concretely, this takes the shape of a series of **technical arrangements** and **legal protocols** on top of the data that describe "how to use the data", in terms of provenance, management of consent of the different parties, purpose and limitations of use, as well as tracking of who accessed the datasets for what purposes. Sector-based initiatives such as IHAN, iShare and IDS reference architecture aim to provide a frictionless and scalable way to create multilateral agreements among companies to reuse the data. They basically set up standards on protocols to document, manage and track consent (IHAN), on legal agreements about what data are owned, accessed and shared by whom (iShare), on technical architectures that ensure data sovereignty at every stage of the data value chain and at every data sharing point, for instance through the IDS connector.

Interoperability across sectors aims to define principles and frameworks for interoperability that allow data sharing across sector, including all of the three points above. In this case, the technological and legal challenges are far greater, hence the need for more abstract principles that can ensure the interoperability between the standardization initiatives taking place within the sector, so that they do not constitute de facto barriers to cross sectoral data sharing.

All these initiatives are addressing long standing and well known problems. Data sharing is one of the oldest issues in computer sciences. To develop a technical format, a metadata scheme or to define legal agreements between two companies for sharing data is time consuming but ultimately a matter of costs. But the major difference is that to grasp present opportunities **data sharing has to happen at a scale and speed never seen before**. It is entirely another matter when the companies involved are tens, hundreds or thousands. These metadata standards, data sharing schemes and interoperability are designed to enable the scaling up of data sharing beyond bilateral relations, allowing for data reuse for the widest set of purposes and for serendipitous innovation.

Schemes such as IHAN, iShare and IDSA reference architecture are designed to make these legal and technical agreement as much "plug and play" as possible to facilitate deployment at scale and reduce transaction costs.

In other words, such standards and schemes allow for "many-to-many" network effects in data reuse. IDSA refers explicitly in their white paper to the analogy of peer-to-peer communication. Following the analogy, the beneficial effects of the wide adoption of such protocols could be compared to the benefits of TCP/IP or HTTP.

With regard to the stated activities, the entity should aim:

- To work with data users to capture, understand and address current and emerging standards requirements, and share best practices
- To facilitate an effective method of forming and running collaborative special interest groups and new standards initiatives.
- To work with data holders and intermediaries to develop consensus and facilitate interoperability, to evolve and integrate data specifications
- To offer a set of guiding principles and guidelines to enhance operational efficiency towards data interoperability
- To raise awareness about successful data sharing schemes that can eventually scale-up and widely facilitate data sharing

2.2.2.3.1 Policy Option 0: Baseline scenario - No horizontal action at EU level

In the baseline scenario, no horizontal action is taken at European level on data governance and interoperability of common European data spaces and data standardization. Yet, actions may be undertaken at national and sectorial level as announced in the European Strategy for Data.[141]

This policy option would rely on industry led initiatives such as iShare, AboutML and IDSA, on national or sectoral initiatives without any guidance or orchestration at European level. As stated before, the traction of these initiatives is only emerging and the level of data reuse today remains far below optimal.

2.2.2.3.2 Policy Option 1: Coordination at EU level and soft regulatory measures only

The first policy option would consist of EU coordination and soft measures, which have been used in the area of data sharing over the past decade. Until present, it is estimated that the impact of coordination and soft policy measures is limited. Under this first policy scenario, the European Commission would adopt a recommendation or guidelines with no binding power to address the different problems identified in section 2.1.

The recommendation would suggest to the Member States to set structures in place to support processes that can help lower the transaction costs of data sharing. This scenario would also create a network of data sharing experts as an ***informal Expert Group*** of the European Commission. This group would be tasked with issuing technical guidance for cross-border and cross-sectoral data sharing such as on interoperability issues, generic standards or metadata descriptions.

2.2.2.3.3 Policy Option 2: Regulatory intervention with low intensity

The second policy option would consist of creating a European Data Innovation Board. The board would be a coordination mechanism at European level that would take the form of a ***formal Expert Group or a scientific committee*** set by legislation, hosted by the European Commission. The

[141] COM Strategy data

functions of the Expert Group would be limited to the general technical guidance on issues related to data standards, data specifications, metadata, ontologies or findability.

2.2.2.3.4 Policy Option 3: Regulatory intervention with high intensity

The third policy option would consist of a European Data Innovation Board. However, under this option the board would be an ***independent European body*** with legal personality and supported by a secretariat. This body would be inspired by the structure and operational characteristics of European Data Protection Board (EDPB). The functions of such board would be of low intensity and specific mandate for the accreditation of certification schemes for data intermediaries.

2.2.2.4 Policy options: Establishing a certification framework for data intermediaries

The table below lists the policy options developed for the domain on certification framework for data intermediaries.

Table 17 - Policy options Establishing a certification framework for data intermediaries

Option	Description
Option 0	Baseline scenario- No horizontal action at EU level
Option 1	Coordination at EU level (industry driven self-regulatory certification framework)
Option 2	Regulatory intervention with low intensity (voluntary certification framework)
Option 3	Regulatory intervention with high intensity (compulsory certification framework)

2.2.2.4.1 Policy Option 0: Baseline scenario- No horizontal action at EU level

In the baseline scenario, **no horizontal action is taken at EU level** on regulation of European data sharing platforms and interoperability of common European data spaces. In the current situation, there are no specific regulatory and non-regulatory actions taken at the EU level targeting data intermediaries or data sharing platforms, and therefore no certification framework established for data intermediaries in the European market. However, action may be taken at sectoral or Member State level. This might lead to further interoperability issues and regulatory fragmentation in the internal market. Cross-sector data sharing will happen only in limited cases between those common European data spaces that have compatible sectoral legislation, standards and infrastructures. This will prevent the EU from reaping the full benefits of cross-sector data sharing, which account for 20% of the benefits of data sharing in general[142].

2.2.2.4.2 Policy Option 1: Coordination at EU level (industry driven self-regulatory certification framework)

As a policy option that aims to promote coordination at EU level, the European Commission could ask private operators, such as representatives of data intermediaries (providers of data sharing services) active in the European market, industry associations and certification bodies to create a network of data sharing experts as an informal expert group or stakeholder forum of the European Commission. This would enable the stakeholders involved to coordinate, exchange and present their views and experience on the topic, aiming to align on best practices and the way forward. An outcome of these discussions could be the creation of an **industry-driven, self-regulatory code of conduct** by the stakeholders.[143] This would not be a compulsory regulatory measure but it would remain at

[142] Realising the economic potential of machine-generated, non-personal data in the EU, Deloitte Report for Vodafone Group, July 2018

[143] Similar to the self-regulatory code of practice on Disinformation, agreed by online data platforms in 2018

the choice of the data intermediaries to decide whether they would like to sign it or not. The code of conduct could further lead to the development of **self-regulatory certification scheme-** by the data intermediaries. The certification scheme would include parameters agreed by the industry that help bring trust to data intermediaries offering data sharing services in B2B contexts and/or personal data spaces, by ensuring that the certified intermediaries function as "neutral intermediaries". These could include, among others, rules regarding specified sources of data, the nature of the intermediary, its business model and the service offered, compliance with legislation, cybersecurity measures, transparency and non-discrimination in data sharing (non-discrimination might not apply for certain type of data, i.e. in cases of criminal activities or poor data quality, but it will be important to be transparent on that). This industry-driven approach to establish a self-regulatory certification framework could entail the finance of a private certification agency, while the government role would be limited by participating as an observer or providing guidance. Finance from the public sector actors for the self-regulatory certification might also be available in the case that it is deemed needed.

Similar efforts for the establishment of a self-regulatory certification framework are already in place at the European and international level, within the MyData Community[144] and the NYU GovLab. In particular, a "self-description" process has been initiated within the MyData Community, targeting organisations that have signed the operator interoperability MoU. This is a voluntary, self-description process to allow operators how show their services meet the MyData human-centric criteria as described in the Understanding MyData Operators white paper[145]:

- to demonstrate alignment with the MyData principles. In the future, the development seems to be towards governed ecosystems and thus more neutral operators;
- to describe the systems for personal data management with respect to the MyData operator reference model;
- to show that they follow the two criteria of transparency and the person as the primary beneficiary.

As a result, in July 2019 16 organisations from 12 countries, who are working for human-centric approaches to personal data, were awarded the inaugural status of MyData Operator 2020.[146] Furthermore, the NYU GovLab has also developed a list of "Trusted Intermediaries" for data collaboration, for third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia.[147] Finally, a certified data pool list is available in the frame of Global Data Synchronization Network (GDSN) for computer systems exchanging information through data pools, enabling collaborators to operate based on standards that support live data sharing and trading updates.[148]

2.2.2.4.3 Policy Option 2: Regulatory intervention with low intensity (voluntary labelling framework)

A policy option for the low intensity regulatory scenario could entail the adoption of a legislative/regulatory measure establishing a **voluntary labelling framework** for novel data intermediaries which would allow them to function as organisers/orchestrators of data sharing or pooling within such spaces and to obtain a label/kitemark/seal. This could be implemented by the means of a legislative act adoption (regulation or directive) and further developed by a delegated act, defining in detail the core criteria and certification requirements, that should be met by all labelled intermediaries in order to demonstrate their neutrality and absence of conflict of interest, in

[144] https://mydata.org/about/
[145] https://docs.google.com/document/d/1e3hvYSqsNas8ZWW3HXvq5V9a_H6r0AlpvaCC9HWvvQw/edit
[146] https://mydata.org/2020/07/29/press-release-mydata-operator-2020-status-awarded-to-16-organisations-from-around-the-world/
[147] https://datacollaboratives.org/explorer.html?#trusted-intermediary
[148] https://www.gs1.org/services/gdsn/certified-data-pools-list

particular the absence of competition with data users (providers of services seeking to use data shared by data holders). The adoption of the legislative act would not alter –in substantive law- the rights and obligations of persons and organisations on data, but would establish a structural enabler encouraging data sharing through data intermediaries. The aim would be to lower data sharing transaction costs, bring trust among stakeholders in the data sharing market or pooling within the common European data spaces, in light of the current distrust in platform business models and the limited brand recognition of the novel services providers that are emerging. The certification criteria might entail softer "neutrality" requirement for B2B data intermediaries providing data-sharing services addressing business users and handling exclusively industrial data (e.g industrial data platforms, data marketplaces, trusted third parties, data collaboratives, data trusts, data trusts), compared to C2B data intermediaries or "personal data spaces", addressing individuals. The certification criteria would be stricter for data intermediaries dealing with consumer's personal data (e.g PIMS/PDS, data unions, data trusts, data cooperatives, data collaboratives), as neutral operators of personal data spaces should limit themselves to data sharing services only and consent management (not added value services based on the data) and have fiduciary duties towards the individuals using them. The ambitions for this policy option would entail: a) a quick applicability process after the adoption of the legal instrument setting the criteria, b) verification by the data permit authorities and c) strict deadlines for receiving the results of the verification process. The potential role of Member States's governments to set up the process will have to be examined, in line also with the policy options developed under the other domains of this study, particularly under Measures facilitating secondary use of sensitive data held by the public sector, as the handling of the application process and the awarding of the labels/certification would be done by the one-stop shop mechanisms set up by Member States which would also handle requests regarding the reuse of public sector data. The certification criteria could include among others:

- Strict notion of function structural separation: Structural separation of data intermediation services from both data holders and potential data users: Data intermediation services may not propose any service building on the data transacted [alternatives possible: merely legal or functional separation];
- Questions of data dominance and ownership as well as fair and non-discriminatory access to the data intermediation service for both data holders and data users.
- Data intermediaries' establishment in Europe: Providers of data sharing services offering services to business users shall have their principle place of business within the European Union.

In terms of international data flows, it is proposed that providers of data sharing services need to take adequate organisational and legal measures to prevent that jurisdictional decisions of third countries that would require access to data relating to European companies and individuals would take effect without making recourse to mutual legal assistance request that would ensure European jurisdictional control over these decisions. In practice this may mean that global players need to create legal entities in Europe that are entirely separated from the corporate structure in the third country, including at the level of ownership (cf. previous collaboration between Microsoft and Deutsche Telekom). In addition to the above, for C2B data intermediaries it is proposed not to add elements to the existing adequacy decision regime of the GDPR.

2.2.2.4.4 Policy Option 3: Regulatory intervention with high intensity (compulsory certification framework)

A policy option for the high-intensity regulatory scenario could entail the establishment of a European **mandatory certification framework** for all types of data intermediaries. Similarly to the previous policy option, this could be implemented by means of a legislative act adoption (regulation or directive) defining hard neutrality requirements and criteria to be respected by both B2B and C2B data sharing platforms. In this case, the certification would be compulsory for all the data sharing platforms in order to ensure the compliance of their activities with the specific provisions defined by the regulatory measure. Neutral operators of personal data spaces should only offer data sharing services and consent management (and no added-value services based on the data), while B2B data intermediaries may offer additional data sharing services, but subject to conditions of structurally separating data intermediation services from other services. Certification would be awarded by private conformity assessment bodies, based on criteria developed at the European level. Such bodies would be accredited by the European Data Innovation Board. This policy option might also require, without being necessary, a level of Member States' governmental involvement and responsibilities (e.g MS to financially support the private conformity assessment bodies, provide guidance and overseeing of the certification process).

Similar efforts to regulate data sharing platforms have been conducted also at the international level with particular examples in the US with the adoption of the Data Broker List Act of 2019,[149] in Japan with the Release of the Guidelines of Certification Schemes Concerning Functions of Information Trust ver. 1.0,[150] only for C2B data intermediaries in both cases, as well as in India. At European level, other types of certification frameworks, (including i.a. GDPR certification and cybersecurity certification) have been used in the past to ensure trust in certain markets and provide an added-value to the companies.

2.3 Assessment of the policy options

This section presents the assessment of the policy options per domains identified in the previous section with regard to their effectiveness, efficiency and coherence and who will be affected.

This section presents our draft assessment of the impacts of all the options, including the baseline scenario.

The following assessment criteria were agreed on for the assessment of the impacts of the options:

- Effectiveness in achieving the policy objectives:
- Achievement of specific objectives;
- Achievement of general objectives;
- Efficiency:
- Costs of the option;
- Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders;
- Coherence of the option.
- Proportionality and legal/political feasibility criteria will be also considered when comparing the policy options.

To the extent possible, the assessment is built on **quantitative and qualitative information, including costs and benefits**. For this purpose, we took various data sources into account for the assessment of the impacts, including:

- Desk research, including a legal analysis;
- Interviews;

[149] https://www.govtrack.us/congress/bills/116/s2342
[150] https://www.meti.go.jp/english/press/2018/0626_002.html

- Workshops.

The aim was to collect as comprehensive quantitative data as possible. However, consulted stakeholders and pre-existing studies only provided data for some types of costs and benefits. In this section, illustrative examples of quantitative and qualitative feedback from stakeholders with regards to costs and benefits of each policy options per domains have been included.

2.3.1 Measures facilitating secondary use of sensitive data held by the public sector

This section assesses the baseline and three policy options for Measures facilitating secondary use of sensitive data held by the public sector.

2.3.1.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 18 – Overview of stakeholders affected by policy options on Measures facilitating secondary use of sensitive data held by the public sector

Who?	How?
Data holders	Data holders would in essence have a reduced range of tasks to perform when it comes to sensitive data reuse. Indeed, a number of functions currently performed by (most) data holders would be centralised under both policy options 2 and 3 (the latter entailing a larger number of such tasks that would be centralised). As a result, data holders should see a significant reduction in their running costs. Indirectly, data holders which also serve as decision-making bodies (such as national ministries) or which offer public services to citizens (such as in the public healthcare sector) would benefit from new insights generated by research reusing sensitive data. This could lead to more effective and/or efficient policy-making, and concrete benefits in health, such as lower costs, higher efficiency, better treatments, and lives saved.
Data intermediary	In all likelihood, public sector data intermediaries would be the actors taking on the role of one-stop shop under policy option 2 or data authorisation body under policy option 3. This will result in increased costs linked to these additional tasks – although ultimately, these costs would be borne by data re-users and/or by the taxpayer.
Data (re)users	Data re-users would in essence see their activities facilitated under policy options 2 and 3, since many tasks currently performed by a range of different actors would be centralised. As a result, transaction costs associated with having to deal with a range of actors would be greatly reduced, resulting in time savings. In parallel, data re-users would in all likelihood pay for the service performed by either the one-stop shop under policy option 2 or the data authorisation body under policy option 3. Whether these costs will be lower than costs currently incurred will depend on the specific case.
Society	Overall, society would benefit from greater access and re-use of sensitive data. Indeed, new insights generated from research would in theory lead to more effective and/or efficient decision-making in a range of domains, including health, social affairs, transport, and the environment. In addition, individual citizens will have greater control over the re-use of their data through increased transparency.

2.3.1.2 Policy option 0: Baseline

2.3.1.2.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

2.3.1.2.1.1 Achievement of specific objectives

In the absence of EU action, **Member States would remain free to take their own approach** with regards to the re-use of data held by public bodies and the use of which is subject to the rights of others. Uncertainty with regards to applicable rules and legislation would likely continue in some Member States, and only some Member States would likely take steps towards interoperability of data cross sectors. Yet, over 75% of respondents to the Open Public Consultation on the European strategy for data believe that public authorities should do more to make a broader range of sensitive data available for research.[151]

As a result, it is **uncertain whether data held by the public sector and the use of which is subject to the rights of others would be generally more available for reuse**. Reusers would therefore be unable to increase their use of such data for research and development or new business opportunities, while policy-makers would not benefit from improved input to guide their decisions. The development of Artificial Intelligence (AI) would not benefit from improved access to data the use of which is subject to the rights of others, and would therefore be impeded.

Likewise, **interoperability issues** across sectors and Member States **would likely persist**, causing reusers to continue spending unnecessary time pre-processing (i.e. pseudonymising and anonymising) the data in order to combine it. In the absence of further reuse of their data, data holders would have **no incentive to ensure their data is of the highest possible quality** and accuracy. Fragmentation as regards access to, and combination of data of sufficient quality would continue. Thus, imbalances would persist between reusers with the resources to overcome these issues and reusers without such resources.

Citizens wishing to **exercise their rights under the GDPR** – for instance, retracting their consent for their data to be reused – would continue facing **opaque and/or cumbersome** procedures for doing so in some Member States, with potentially **negative consequences for fundamental rights and for trust** in reuse of data the use of which is subject to the rights of others (and thus in common European data spaces).

This all results in a **limited positive economic impact** overall, particularly in terms of:

- Time and resources spent by data re-users as a result of these issues;
- Duplication of time and resources spent by some data holders to provide data the use of which is subject to the rights of others ; and
- Absence of gains due to increased re-use of data, increased innovation and new business opportunities (and therefore growth and competitiveness), and potentially better economic policies.

Furthermore, the absence of better policies resulting from better information would **limit the potential for positive social and environmental impacts**. This would be compounded by the duplication of re-use mechanisms for data the use of which is subject to the rights of others existing within a single Member State – resulting in a duplication of energy-intensive IT infrastructures enabling such reuse.

2.3.1.2.1.2 Achievement of general objectives

Absence of EU action would not contribute to setting the foundations of a Single Market for Data, since Member States would set their own policies. While the data economy of some Member States

[151] European Commission, Summary report of the public consultation on the European strategy for data. See: https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

would likely be strengthened as a result of national policy, this would not be the case across the EU. This fragmentation would not contribute to increasing the EU's sovereignty on data and on the key enabling technologies and infrastructures, and neither would it contribute to completing the internal market.

2.3.1.2.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

2.3.1.2.2.1 Costs of the option

Data holders and intermediaries would continue to bear a number of costs in some Member States, namely:

- The costs of establishing and maintaining specific reuse mechanisms for data the use of which is subject to the rights of others when applicable, including the associated IT infrastructure;
- The costs of pseudonymising and anonymising data prior to making it available;
- The costs of examining applications for data access;
- The costs linked to training sufficient staff to perform these tasks;
- The opportunity costs, for data holders, linked to not accessing more research made available by data reuse; and
- Potential costs linked to data breaches.

For instance, the statistical office of a mid-sized EU Member State has approximately two FTEs working on pseudonymising and anonymising data, representing a cost of at least EUR 140,000 per year (for 70-100 requests a year); while the statistical office of a large Member State has approximately 10.6 FTEs working on tasks related to making data the use of which is subject to the rights of others available for re-users (including handling applications and pre-processing data, excluding IT costs that represent 20% of staffing costs).

Data reusers would continue to bear a number of costs in some Member States, namely:

- Time and resources spent on identifying the data holder holding the desired data;
- Time and resources, including of a pecuniary nature, spent on producing and submitting different (and not always successful) applications to access data from different holders;
- Time and resources spent combining data which is not necessarily interoperable.

For instance, a private sector data re-user in a large Member State estimates that the time spent on one data request application is equivalent to roughly five to 15 days of effective work depending on the complexity of the application.

2.3.1.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

In the absence of EU action, no particular benefits were identified for stakeholders across the EU.[152]

2.3.1.2.3 Coherence of the option

This policy option does not entail any piece of legislation which might be incoherent with other policy options. However, no action would be incoherent with the policy priorities identified.

[152] There would however be benefits for stakeholders in the Member States that have domestically implemented something similar to policy options 2 or 3.

2.3.1.3 Policy option 1: Guidelines

This section assesses the first policy option for Measures facilitating secondary use of sensitive data held by the public sector.

2.3.1.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

2.3.1.3.1.1 Achievement of specific objectives

The **extent to which this policy option contributes to the specific objectives of the action is contingent on the degree to which Member States decide to follow the Commission's Recommendations or guidelines**. At any rate, this policy option is highly unlikely to be detrimental to the objectives. Indeed, recommendations to set up structures enhancing the re-use of publicly held data, if followed, would likely result in an increase in access and re-use of data the use of which is subject to the rights of others – contributing to **building common data spaces**.

These recommendations could have **positive economic, social and environmental impacts** (due to increased research and improved decision-making, on which more below), and **enhance trust** as a result of increased transparency, if:

- They recommend best practices such as a one-stop shop (policy option 2) or a single data authorisation body (policy option 3), and
- They are followed by Member States.

In addition, setting up an informal Commission Expert Group issuing technical guidance for cross-border and cross-sectoral data sharing – for instance on interoperability issues, generic standards or metadata descriptions – **may on the long term lead to increased interoperability** across sectors and across Member States. This would depend on the extent to which this guidance is implemented.

2.3.1.3.1.2 Achievement of general objectives

The extent to which this policy option contributes to the general objective of the action is contingent on the degree to which Member States decide to follow the Commission's Recommendations or guidelines. In the event that they are, the increased data access and re-use, trust and interoperability would contribute to setting the foundations of a Single Market for Data through the creation of common data spaces.

2.3.1.3.2 Efficiency: Costs and benefits of the option

The non-binding recommendations/guidelines encouraging Member States to set up structures enhancing the re-use of publicly held data subject to the rights of others were discarded from the CBA and subsequent macroeconomic analysis for several reasons.

- First, several stakeholders expressed **doubts as to their overall effectiveness**, noting that recommendations and guidelines on data sharing and reuse abound but are **not always followed**. At the validation workshop organized on 8 July, stakeholders present indicated they **expect a third to half of Member States to implement such guidelines**.
- Second, stakeholders interviewed and present at the workshop indicated that the level of ambition of such guidelines or recommendations would likely be inversely proportional to the number of Member States adopting them. This is because the more ambitious the guidelines, the more effort and resources would be needed for their implementation.

As a result, it is estimated that policy option 1 would have a limited effectiveness, while any measure of its efficiency would be over reliant on assumptions linked to the content and uptake of such recommendations or guidelines.

2.3.1.3.3 Coherence of the option

No incoherence of this option with existing legislation was identified.

2.3.1.4 Policy option 2: One-stop shop

This section assesses the second policy option for Measures facilitating secondary use of sensitive data held by the public sector.

2.3.1.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the policy objectives.

2.3.1.4.1.1 Achievement of specific objectives

The establishment of a once-stop shop would contribute to achieving the specific objectives. Indeed, providing information and guidance to data holders may incentivise them to make more data the use of which is subject to the rights of others available for access and re-use, and may increase demand for data the use of which is subject to the rights of others due to increased transparency, particularly for smaller re-users. This potential increase in available data the use of which is subject to the rights of others would **contribute to building common data spaces**, while increased fairness resulting from equal access to information and guidance would be a **positive social impact**.

This potentially increased demand for, and re-use of, data the use of which is subject to the rights of others would translate into improved and increased research, and therefore into better policy-making resulting in positive **economic, social and environmental impacts**. This would be in line with the results of the OPC, in which 91,5% of respondents agreed that "more data should be available for the common good, for example for improving mobility, delivering personalised medicine, reducing energy consumption and making our society greener."[153]

Furthermore, this option would foster **trust through transparency** between data re-users and data holders, as well as trust among the general public – particularly if the one-stop shop provides legal guidance to citizens on how to exercise their rights under data protection laws. This is consistent with the results of the OPC: 84,6% of respondents believe that it should be made easier for individuals to give access to existing data held on them, in line with the GDPR.[154]

2.3.1.4.1.2 Achievement of general objectives

An increase in the re-use of data held by the public sector and the use of which is subject to the rights of others, contributing to building national common data spaces, complemented by greater trust in how sensitive (and particularly personal) data will lead to **more integrated national markets for data**. This is a significant **first step setting the foundations of a Single Market for Data**.

2.3.1.4.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

[153] European Commission, Summary report of the public consultation on the European strategy for data. See: https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

[154] *Ibid*.

2.3.1.4.2.1 Costs of the option

The establishment of a one-stop shop **would entail costs for data re-users**, in the form of potential fees to fund the one-stop shop's operations – resulting in benefits neutering the costs for the one-stop shop.[155] These costs for data reusers are expected to be neutered – or potentially outweighed – by the benefits incurred by access to new data.

It would entail **costs for data holders**, in terms of additional staff required to cope with potential additional demand for data the use of which is subject to the rights of others (as a result of it being more easily accessible). In addition, coordinating and liaising with the one-stop shop on a regular basis would cost time and resources to data holders. For instance, the statistical institute of a large Member States estimates that each data holder in that country spends about two weeks of effective work per year coordinating with the country's one-stop shop. At the same time however, data holders would gain time from the amount of work (in, for instance, answering queries from data reusers) that would performed by the one-stop shop instead of them.

Establishing a one-stop shop **may also entail costs for society** – specifically, for taxpayers – if fees charged to data re-user do not cover all the costs associated with the one-stop shop. For instance, the one-stop shop of a large Member State employs 8 FTEs.

The figures in the table below originate from interviews with several stakeholders, including the German Data Forum (RatSWD) that acts as a one-stop shop in Germany, one of Germany's accredited Research Data Centres (acting as a data holder), and Statistics Denmark (for the cost linked to maintaining a secure data processing environment). Specifically, as regards costs associated with the one-stop shop, RatSWD's running costs are currently EUR 900,000 per year,[156] while a secure data processing environment costs on average roughly EUR 610,000 per year – thus amounting to an overall annual cost of approximately EUR 1,510,000. On the other hand, the German Forschungsdatenzentrum estimates that each data holder spends roughly 2 weeks of work annually preparing and attending meetings at RatSWD. One hour of work is assumed to cost EUR 45 on average, while two weeks of work are assumed to correspond to 80 hours, which corresponds to a total of EUR 3,600 per annum.

These figures are conservative estimates: participants to the workshop held on 8 July 2020 indicated that the running costs of a one-stop shop are likely to be higher than EUR 900,000, and that data holders would likely spend a little more than 2 weeks per year coordinating with the body.

In addition to these, the costs for reusers (in the form of fees payable to the one-stop shops) was assumed to be EUR 500 per application. This assumption is based on the current fees charged by Findata (Finland's single data authorisation body). Since the one-stop shops would perform fewer tasks than a data authorisation body – in particular, they would not handle data access applications – these fees are assumed to approximate half of the fees charged by Findata. The actual amount however would vary depending on the Member State.

The extrapolation of these figures to the EU level can be found in the table below, while the full overview of costs (and benefits, including the cost benefit analysis) is in Annex I.

[155] In addition, Member States could establish mechanisms for distributing these benefits among data holders. This however would be at the discretion of Member States.

[156] While some of these costs may be covered by user fees, this is currently not the case in Germany with RatSWD and may not be the case in all Member States.

Table 19 – Overview of costs for Measures facilitating secondary use of sensitive data held by the public sector| PO 2

Overview of costs (EUR) – PO 2							
		Data holders		**Once-stop shop (data intermediary)**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures facilitating secondary use of sensitive data held by the public sector	*Direct costs*	-	7.6 million p.a.	286.4 million	16.5 million p.a.	-	41.8 million p.a.
	Indirect costs	-	-	-	-	-	-

2.3.1.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The establishment of a one-stop shop in each Member State would result in a variety of **benefits for data reusers**:

- Time and resources saved when identifying the data holder holding the desired data (these are assumed to be equal to 20 hours of work per application, or EUR 590);
- Increased fairness in access of data held by the public sector and the use of which is subject to the rights of others, i.e. all re-users would have equal access to valuable information on how to access that data;
- This improved access to data the use of which is subject to the rights of others would likely result in an increase in re-use of the data, since potential reusers currently not reusing data the use of which is subject to the rights of others may start doing so once it is easier;
- Access to legal guidance potentially resulting in saving time and resources related to legal training; and
- Time and resources saved by accessing already interoperable data across sectors.

It would also entail potential **benefits for data holders**, in terms of:

- Access to legal guidance potentially resulting in saving time and resources related to legal training;
- Access to technical guidance on how to allow data reuse, resulting in a decreased risk of data breach and the associated costs;
- Time and resources saved by not providing, and maintaining, a secure data processing environment;
- As noted above, such an environment costs on average EUR 610,000 p.a.
- Access to an increased amount of research resulting from an increased demand for data the use of which is subject to the rights of others – leading to better policy-making;
- Increased visibility; and
- Decreased difficulty in cross-border access to data.

This policy option would entail **benefits for society** more broadly:

- A one-stop shop would increase transparency regarding re-use of data the use of which is subject to the rights of others, which may contribute to increased trust;
- This increased transparency – along with additional research resulting from additional data re-use – would likely improve decision-making, with economic, social and environmental benefits for society.

Lastly, the one-stop shop is assumed to charge EUR 500 per application to cover (part of) its running costs.

These figures are extrapolated to the EU level in the table below, while the full overview of benefits (and costs, including a cost-benefit analysis) is in Annex I.

Table 20 – Overview of benefits – Measures facilitating secondary use of sensitive data held by the public sector| PO 2

Type of action	Description	Amount (EUR)	Stakeholders
Measures facilitating secondary use of sensitive data held by the public sector	**Direct benefits**		
	Resources saved as a result of not maintaining secure processing environment and analytics tools[157]	*684 million* p.a.	**Data holders**
	Direct revenues (fees)[158]	*41.8 million p.a.*	**Data intermediary**
	Time/resources saved as a result of easier data discovery[159]	*49.2 million p.a.*	**Data re-users**
	Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders and (re)users**
	Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
	New economic base	*Not quantifiable due to lack of data*	**Data intermediaries**
	Cost savings and revenue generation from results created by data use. Possibility to enter new business sectors, research fields, generation of new correlation of data, which enables new insights.	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		
	Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders**

[157] This is based on the assumption that 20% of data holders would relinquish their data processing environment to use the environment established as part of the policy option, and that 30% of the data pre-processing and provision work would no longer be done by data holders but by the one-stop shop.

[158] This is based on the assumption that the one-stop shops would charge on average EUR 500 per application. This would cover part of the one-stop shops' running costs.

[159] This is based on the assumption that each data re-user would save about 20 hours of work per application.

New insights	*Not quantifiable due to lack of data*	**Data intermediaries**

2.3.1.4.2.3 Findings of the Cost-Benefit Analysis

This policy option, although it carries a large one-off cost for the establishment of the one-stop shops, brings equally large – and recurrent – benefits that greatly outweigh the recurrent costs associated with the one-stop shop's operation. The Cost-Benefit Analysis conducted as part of this study (and which can be found in Annex I) indeed finds a benefit-cost ratio (BCR) of 9.2. In other words, benefits incurred by this policy option are expected to be over nine times larger than costs.

2.3.1.4.3 Coherence of the option

A one-stop shop performing discovery and advisory services is coherent with existing legislation as well as with ongoing efforts by some Member States to set up such one-stop shops.

2.3.1.5 Policy option 3: Single data authorisation body

This section assesses the third policy option for Measures facilitating secondary use of sensitive data held by the public sector.

2.3.1.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the policy objectives.

2.3.1.5.1.1 Achievement of specific objectives

The establishment of a single data authorisation body would contribute to achieving the specific objectives. Indeed, guaranteeing a single, streamlined application process for accessing data the use of which is subject to the rights of others, and providing a secure data processing environment to reusers, would consist of the technical, legal and organisational support needed for data holders to make more data usable – contributing to **building common data spaces**. This would be reinforced by the likely increase in demand for data the use of which is subject to the rights of others resulting from the process of accessing it being easier. This would also ensure more fairness in access to data the use of which is subject to the rights of others, since all re-users, small and large, would have equal access to information on how to access data and to data analytics tools – a net **positive social impact**.

This increased demand for, and re-use of, data the use of which is subject to the rights of others would translate into improved and increased research, and therefore into better policy-making resulting in positive economic, social and **environmental impacts**. In addition, **positive economic impacts** would be reinforced by the non-duplication of efforts by public sector data holders to make data available for re-use.

The single data processing environment and the upstream work conducted by the single data authorisation body would **ensure interoperability** of the data across sectors within a given Member State, and could be a stepping stone towards interoperability across Member States.

This option would foster **trust through transparency** between data re-users and data holders, as well as trust among the general public that their data is re-used following a single, streamlined procedure managed by a publicly **accountable** body. However, this impact may be counterbalanced by reduced trust from the public as a result of sensitive (and particularly personal) data being available for reuse for strictly commercial purposes.

The single data permit authority of a mid-sized Member State expressed doubts as to the feasibility of this option in all Member States, pointing to different levels of governance centralisation and to the centrality of trust among citizens that their data will be re-used in a secure way and for the purposes to which they have agreed. Trust related issues from citizens have already prevented the establishment of a single data permit authority in one Member State.

2.3.1.5.1.2 Achievement of general objectives

An increase in the re-use of data held by the public sector and the use of which is subject to the rights of others, contributing to building national common data spaces, complemented by greater interoperability within Member States and greater trust in how sensitive (and particularly personal) data will lead to **more integrated national markets for data**. This is a significant **first step setting the foundations of a Single Market for Data**.

2.3.1.5.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

2.3.1.5.2.1 Costs of the option

The establishment of a single data authorisation body **would entail costs for data re-users**, in the form of potential fees to fund the single data authorisation body's operations.

It would entail **costs for data holders**, in terms of staff required to coordinate and liaise with the single data authorisation body on a regular basis.

Establishing a single data authorisation body **may also entail costs for society** – specifically, for taxpayers – if the fees it charges do not cover the body's running costs.[160] For instance, the single data authority of a mid-sized EU Member States has an overall budget of EUR 5.2 million this year (expected to decrease), following one-time R&D costs of approximately EUR 10 million (corresponding to a pilot project). It will employ an estimated 25 FTEs once fully running, which each FTE costing approximately EUR 75,000 and requiring one to two weeks of training. This single data authority has had an approximate 7 FTEs specifically working on 38 data access applications since early May 2020.

In addition, enabling the re-use of data the use of which is subject to the rights of others, and especially of personal data, for **strictly commercial purposes would raise ethical questions and may undermine citizens' trust** – which might lead to a reduced amount of personal data available to reuse (due to potentially fewer citizens agreeing to the reuse of their data), in turn negating the positive impacts of increased reuse of data the use of which is subject to the rights of others. In addition, it may contravene national laws of several Member States where for instance statistical microdata may only be reused for research purposes.

The table below provides an overview of annual costs associated with this policy option, based on Finland's single data authorisation body for health and social data, Findata. Prior to Findata's establishment, the Isaacus pilot project was launched with a EUR 14 million budget, of which approximately 75% (i.e. EUR 10.5 million) were directly linked to Findata's establishment. The running costs of Findata, once it is fully up and running (i.e. in a couple of years), are estimated to range between EUR 4 and 5 million per annum. Currently, Findata charges reusers a fixed fee of EUR 1,000 per application for a new permit, and an additional EUR 115 per hour worked on pre-processing

[160] The legislative design would suggest that the fees charged to reusers could cover entirely the cost of operating the single data authorization body. However, the distribution of costs linked to the operation of the single data authorisation body, between re-users (in the form of fees) and the taxpayer (in the form of taxes), would be left at the discretion of Member States.

and combining datasets from different holders. Findata expects to handle an average of 600 applications per year, while one Finland-based re-user interviewed indicated they submit applications regarding around 30 projects per year (i.e. 30 applications via Findata, assuming they all concern data from multiple holders). It is currently expected that on the medium term, Findata will receive approximately **EUR 1 million each year from public funds**, with the remaining costs being covered by user fees. To avoid double-counting the running costs of Findata that will be covered by user fees, the EUR 1 million figure is used as an estimate of running costs while the remaining costs are counted as costs for reusers (in the form of the user fees).

These estimates may be conservative, as participants to the workshop organised on 8 July estimate that both the one-off R&D costs and the recurrent running costs of a single data authorisation body may be somewhat higher. However, Findata's figures themselves may be higher than they will be in the future, as Findata was established very recently and has only begun its operations. Indeed, Findata representatives indicated that the figures are based only on a few months of operation, and that more accurate numbers should be available by the end of 2020 at the earliest. Nevertheless, the figures are extrapolated in the table below, while a full overview of costs (and benefits, including a cost benefit analysis) is in Annex I.

Table 21 – Overview of costs for Measures facilitating secondary use of sensitive data held by the public sector| PO 3

Overview of costs (in EUR) – PO 3

		Data holders		**Single data authorisation body (data intermediary)**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures facilitating secondary use of sensitive data held by the public sector	*Direct costs*	-	-	572.7 million	329.7 million p.a.	-	212.7 million p.a.[161]
	Indirect costs	-	-	-	-	-	-

2.3.1.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The establishment of a single data authorisation body in each Member State would result in a variety of **benefits for data reusers**:

- Increased fairness in access of data held by the public sector and the use of which is subject to the rights of others, i.e. all re-users would have equal access to valuable information on how to access that data, as well as equal access to data processing tools;

[161] This corresponds to the fees payable to the single data processing authority. However, this fee replaces pre-existing fees payable to the different data holders currently processing requests. As a result, according to stakeholders, the costs for data re-users would be lower under PO3 than in the baseline – but current fees could not be quantified.

- This improved access to data the use of which is subject to the rights of others would likely result in an increase in re-use of the data, since potential reusers currently not reusing data the use of which is subject to the rights of others may start doing so once it is easier;
- Access to legal guidance potentially resulting in saving time and resources related to legal training;
- Time and resources saved when identifying the data holder holding the desired data and by handing in one data access request application for access to data from more than one holder, instead of having to apply several times for one project;
- Time and resources saved by virtue of the application process being accelerated – since centralising this process would result in economies of scale; and
- Time and resources saved by accessing already interoperable data across sectors

For instance, three data reusers estimate that having to deal with a single body as opposed to multiple data holders would save about half the time spent on a typical application (down from between 50 to 90 hours to 25 to 45 hours, with each hour assumed to be worth EUR 45, i.e. cost savings in the order of EUR 1.125 to EUR 2.025 for each application). During the 8 July workshop however, participants estimated that these savings may be even higher.

One stakeholder estimates that not having to pre-process data from different holders (since the single data permit authority will have done so) would save them several days of work each time[162] – although this varies and could not be confirmed with other stakeholders (one of which prefers continuing to perform this task to avoid missing potential insights from the raw data).

It would also entail potential **benefits for data holders** resulting from the centralisation of services in the data authorisation body and the above-mentioned economies of scale. Specifically, these would be:

- Time and resources gained as a result of:
- Not processing data access applications;
- This cost is estimated to approximate EUR 400,000 per annum for Statistics Denmark, although participants to the 8 July workshop believe this figure would be lower;
- Not pre-processing the data and providing, and maintaining, a secure data processing environment and data analytics tools;
- These costs are estimated to approximate EUR 1,200,000 (pre-processing and combining data) annually for Statistics Denmark, although participants to the 8 July workshop estimate that the figure would in reality be lower. The costs of maintaining a secure processing environment are estimated to approximate EUR 610,000.
- Access to an increased amount of research resulting from an increased demand for data the use of which is subject to the rights of others – leading to better policy-making.

Lastly, this policy option would entail **benefits for society** more broadly:

- A data authorisation body would increase transparency regarding re-use of data the use of which is subject to the rights of others, which may contribute to increased trust;
- This increased transparency – along with additional research resulting from additional data re-use – would likely improve decision-making, with economic, social and environmental benefits for society; and

[162] The stakeholder was unable to give a precise number.

- The benefits for data holders would translate into public savings, and therefore either reduced tax or increased spending in other areas.

These figures are summarised in the table below, while a full overview of benefits (and costs, including a cost-benefit analysis) is in Annex I.

Table 22 – Overview of benefits – Measures facilitating secondary use of sensitive data held by the public sector| PO 3

Type of action	Description	Amount (EUR)	Stakeholders
Measures facilitating secondary use of sensitive data held by the public sector	**Direct benefits**		
	Resources saved as a result of not processing data access applications	*569.4 million* p.a.	**Data holders**
	Resources saved as a result of not pre-processing and combining datasets[163]	*512.5 million* p.a.	**Data holders**
	Resources saved as a result of not maintaining secure processing environment and analytics tools[164]	*171.5 million* p.a.	**Data holders**
	Resources saved as a result of not having to submit separate applications for one research project	*167.1 million p.a.*	**Data re-users**
	Direct revenues (fees)	*212.7 million p.a.*	**Data intermediary**
	Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders and (re)users**
	Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
	New economic base	*Not quantifiable due to lack of data*	**Data intermediaries**
	Cost savings and revenue generation from results created by data use. Possibility to enter new business sectors, research fields, generation of new correlation of data, which enables new insights.	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		
	Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders**
	New insights	*Not quantifiable due to lack of data*	**Data intermediaries**

[163] This is based on the assumption that 30% of the data pre-processing and combination work would no longer be done by data holders but by the data authorisation body.

[164] This is based on the assumption that 20% of data holders would relinquish their data processing environment to use the environment established as part of the policy option.

2.3.1.5.2.3 Findings of the Cost-Benefit Analysis

This policy option carries benefits that are over twice larger than those associated with Policy Option 2 – in line with its greater ambition and with its higher potential to realise the policy objectives of this intervention. At the same time, it incurs much higher costs – over eight times higher – than Policy Option 2. This is due to the much higher recurrent costs associated with operating a single data authorisation body compared to a one-stop shop, and also reflects the higher level of ambition of this policy option. On balance however, the BCR for this policy option is 2.4 (see Annex I for the full CBA). While positive, this ratio is significantly lower than for Policy Option 2.

2.3.1.5.3 Coherence of the option

Setting up a single data authorisation body is coherent with existing EU policies, and complements Directive (EU) 2019/1024 on open data and the re-use of public sector information (PSI Directive) as it would cover sensitive data currently excluded from the PSI Directive. It is coherent with the GDPR and indeed facilitates the exercise of one's rights under it. However, opening up the reuse of data the use of which is subject to the rights of others to private companies for exclusively commercial purposes would contradict national law in Member States when only researchers are allowed to access and re-use for instance statistical microdata. In addition, a single data authorisation body similar to Findata, which would give access to data the use of which is subject to the rights of others on behalf of data holders, may be incoherent which national laws in at least two Member States stating that only the data holder may provide access to the data it holds for re-use.[165]

2.3.1.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 23 – Summary of impacts for Measures facilitating secondary use of sensitive data held by the public sector

Economic impacts	• Costs for public authorities (one-stop shops/data authorisation bodies) ○ Direct (R&D, staff, premises, equipment etc.) ○ Indirect (coordination with data holders) • Benefits for public authorities (data holders) ○ Direct (time and resources saved) ○ Indirect (new insights from research leading to better decision-making) • Costs for data re-users ○ Direct (fee for the use of one-stop shop/data authorisation body) • Benefits for data re-users ○ Direct (time and resources saved, more equal access) ○ Indirect (increased re-use, better quality data) • Benefits for society • Public savings resulting in better spent taxpayer money, or reduced tax
Social impacts	• Fairer access to, and re-use of, data held by the public sector and the re-use of which is subject to the rights of others • Increased trust among society about the re-use of their data • Better policy-making due to new insights
Environmental impacts	• Better policy-making due to new insights potentially leading to more effective environmental policies

[165] Furthermore, one stakeholder insisted that moving away supporting structures from the underlying data will come with a loss in the capacity to provide targeted expertise on how the specific data assets in question can be used.

Fundamental rights impacts	• Increased transparency as regards reuse of personal data (e.g. related to health) and how to exercise one's rights

2.3.2 Establishing a certification/authorisation scheme for data altruism mechanisms

2.3.2.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 24 – Overview of stakeholders affected by policy options on Establishing a certification scheme for data altruism mechanisms

Who?	How?
Data holders	Data subjects would have more transparency on how their data is being processed and for e.g. companies or organisations that hold data and plan to make it available through data altruism schemes, a certification scheme would be an added benefit. Overall, data altruism schemes, including certifications, would decrease privacy concerns and increase transparency as well as have an added authentication value. It could also enable an equal playing field for also small or medium-sized companies that were previously excluded from data altruism due to high legal or ICT costs.
Data intermediary	Data intermediaries are public sector bodies, organisations and companies. In specific cases this could be research institutes who are hosting data altruism mechanisms for specific research purposes directly, however often a third party-intermediary- will be involved. Data intermediaries would benefit from reduced costs to adapt to various data altruism schemes.
Data (re)users	Data (re)users would greatly benefit from a data altruism scheme. The more data is received, the better data (re)users can make a positive impact on the public good through for example enhanced policy making. The benefits are difficult to quantify but could be tremendous.

2.3.2.2 Policy option 0: Baseline scenario- No horizontal action at EU level

2.3.2.2.1 Effectiveness in achieving the policy objectives

This subsection examines the baseline policy option for Establishing a certification scheme for data altruism mechanisms.

2.3.2.2.1.1 Achievement of specific objectives

In the absence of EU action, Member States and private entities would continue to take their own approach to data altruism. If relevant, they would potentially collaborate bilaterally or initially seek a national approach in federal Member States. Uncertainty with regards to legislation and data handling, think of interoperability and data guidelines, could lead to further fragmentation in the European Union. This would **increase costs** for Member States, companies and organisations such as NGO's because they would have to negotiate data altruism schemes with each Member State individually.

Legislative issues and **fragmentation** is costly and time consuming. Large corporations that want to engage in B2G data altruism schemes have the financial and manpower capabilities to address these issues, however SMEs and possible also NGOS and research organisations could consider this a roadblock to participate in data altruism schemes **due to uncertainty of data security** issues.

2.3.2.2.1.2 Achievement of general objectives

The general objective of creating a Single Market for Data would likely not be achieved very efficiently or effectively considering that every Member State would continue developing data altruism schemes according to its respective political agenda, which greatly varies across the EU27.

2.3.2.2.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

2.3.2.2.2.1 Costs of the option

Costs for data re-users:

- Varying degrees of (macroeconomic) competitive advantages between Member States depending on the advancement of data altruism schemes and legal frameworks
- Loss of additional GDP for the European Union due to lack of coherent data altruism schemes to benefit from
- Unwillingness to share personal data by citizens and private entities companies due to lack of transparency and security issues resulting in increased awareness raising campaigns
- Political ambition, or lack thereof, to prioritize data altruism schemes could limit the creation of a EU27 data economy
- Technical infrastructure costs and interoperability issues for public bodies due to lack of guidelines for data altruism schemes
- Costs related to incorrect data management due to legislative fragmentation
- Data collection at regional level and local levels could be disrupted and additionally complicated when it is a Federal state (such as Germany or Italy) due to different laws within a single Member State which could increase costs for the public body re-using data

Costs for data holders:

- Increased time and resources spent on legislative questions, possible excluding certain Member States and SMEs from data altruism because of high costs, due to legislative fragmentation
- Increase awareness raising and transparency costs to gain consumer trust for the permission to participate in data altruism schemes with the public sector

2.3.2.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data-reusers

- Member States would have more time to define their data altruism scheme needs according to their national standards
- Member States would have more time to prepare citizens for data altruism schemes and increase digital skills to share data
- Lack of a common data altruism regulation and strategy could provide short-term innovation benefits to certain Member States

2.3.2.2.3 Coherence of the option

This policy option does not entail any piece of legislation, which might be incoherent with other policy options. However, no action would be incoherent with the policy priorities identified. Considering that data portability and privacy rights already have national legislative differences which is an issue and could become bigger through no-action.

2.3.2.3 Policy option 1: Coordination at EU level and soft regulatory measures only

This section assesses the first policy option for Establishing a certification scheme for data altruism mechanisms.

2.3.2.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

2.3.2.3.1.1 Achievement of specific objectives

Coordination at EU level and introduction soft regulatory measures could contribute to the specific objectives of creating trust, building a common data space and ensuring data interoperability.

However, this would be to a limited extent since the policy option relies on coordination and soft regulatory measures that could be voluntary or for example an Expert Group. An Expert Group would have to be established, which takes time, and then have 18 months to work on and submit a report on a European data altruism scheme. Until these recommendation are then implemented, it could take many more months. Therefore this would be a very timely exercise. A similar length process would apply to coordination meetings to define soft regulatory measures such as guidelines of metadata or cross-border data sharing.

Most importantly, a lack of incentive or perception of necessity and benefits could deter many data re-users and holders to participate and later implement *soft measures.*

2.3.2.3.1.2 Achievement of general objectives

This policy option would be a step towards establishing the foundation of a Single Market for Data and establishing a common European data space. Nevertheless, due to its voluntary or soft measures it would not immediately accelerate the creation of a Single Market for Data since not all Member States would be bound to these guidelines or it would simply take time to implement them.

2.3.2.3.2 Efficiency: Costs and benefits of the option

Policy Option 1 was not further included in the CBA and subsequent macroeconomic analysis for the following reasons:

- Coordination at EU level would **require willingness from Member States to participate**. Some interviewed Member States stated that they are already involved in bilateral talks with other Member States, and while coordination at EU level would possibly reduce their workload, discussions around table instead of several, this would not necessarily accelerate the discussions. In addition, only the Member States that are already actively pursuing data altruism mechanisms would likely participate according to the interviewed Member States.
- The private sector raised further concerns that coordination at EU level **could take very long and yet not result in concrete action**. Considering that data altruism, and the data economy, are considered a priority, it should be addressed as such with concrete actions.
- Furthermore, adoption of the measures would be voluntary and **could widen the data altruism gap between different Member States and companies** by widening the fragmentation between those that are actively pursuing this and those that are not.

To conclude, this policy option was considered to have very limited effects, possible even **discourage certain stakeholders to participate** and coordinate due to fear of inefficiency. Indicative categories of costs and benefits are nonetheless provided below.

2.3.2.3.2.1 Costs of the option

Costs for data re-users:

- Costs are difficult to predict due to a lack of clarity on *soft measures*
- Could still create legislative fragmentation is certain Member States move faster or beyond the *soft guidelines*
- Willingness of Member States to participate in EU level coordination efforts
- Organisational, legal and technical costs to implement any soft measures
- Interoperability issues could persist limiting cross-border data sharing
- Transparency and trust could still be low among data holder, limiting the amount of data data re-users can reuse
- New, even soft, regulation could create additional costs if it is not build on existing regulations

Costs for data holders:

- Willingness of data holders to participate in EU level coordination efforts
- Organisational, legal and technical costs to implement any soft measures
- Costs by still facing legal and technical fragmentation in Member States
- Transparency and trust issues could persist among customers and individuals holding data

Table 25 – Overview of costs for Establishing a certification scheme for data altruism mechanisms | PO 1

Overview of costs – PO 1

		Data holders		**Data intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
A certification scheme for data altruism mechanisms	*Direct costs*		Time spent on understanding various data altruism schemes to make data available[166]	-	-	-	-Set-up scheme -Raise awareness of scheme -Gain trust of data holders[167]
	Indirect costs	-	-	-	-	-	-

2.3.2.3.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Coordination among Member States would ensure a multi-lateral discussion about data altruism schemes at 'one' table instead of various bilateral discussion, savings time and costs
- Coordination to align on e.g. open source code and a legal basis could increase transparency and trust in data altruism. For example, Germany agreed on a coordinated agreement to request 'consent' from data holders to share their data

Table 26 – Overview of benefits – Establishing a certification scheme for data altruism mechanisms |PO 1

Type of action	Description	Amount	Stakeholders
	Direct benefits		

[166] Further data collection will be performed to provide an estimate.
[167] Further data collection will be performed to provide an estimate.

A certification scheme for data altruism mechanisms	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		
	R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
	Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
	Value of personal data, assumption: the higher the amount of data, the greater the benefit Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	*Not quantifiable due to lack of data*	**Data (re)users**

2.3.2.3.3 Coherence of the option

No incoherence with any existing EU or national legislation was identified, largely due to the absence of any data altruism legislation.

2.3.2.4 Policy option 2: Regulatory intervention with low intensity

This section assess the second policy option of Establishing a certification scheme for data altruism mechanisms.

2.3.2.4.1 Effectiveness in achieving the policy objectives

This subsection assess the effectiveness of achieving policy option 2.

2.3.2.4.1.1 Achievement of specific objectives

By creating a certification scheme for data altruism mechanisms and obligating Member States to implement it, the Commission would achieve to build a common data space with regard to data altruism.

Considering that the certification would be issues by a private organisation would not limit trust and the fact that it is voluntary still ensure data interoperability across sectors.

To facilitate the effectiveness of this policy option, the Commission could additionally **create working sessions** on for example data portability rights, harmonisation of standards and how to set-up and manage personal data spaces in addition to setting-up a certification scheme. This would help Member States to understand the benefits of certification schemes for data altruism mechanisms.

2.3.2.4.1.2 Achievement of general objectives

Considering the current crisis and focus on data altruism schemes to contain the COVID-19 pandemic, the Commission should take this opportunity to make urgent changes and align across the EU27 on data altruism schemes. Thereby it would also take a great step toward building the **Single Market for Data.**

2.3.2.4.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

2.3.2.4.2.1 Costs of the option

Costs for re-users:

- The obligation to set up certification schemes will most likely be carried by the data re-users because most often the data will she made available to the public sector, who then also re-uses the data for analysis. Therefore the obligation to set-up a scheme under option 2 is carried by the public sector, who simultaneously will also re-use the data. However, it has to be noted that the costs and re-use of data could be, and most likely, will be carried by different parties in the public sector. For instance, the ministry of interior affairs could be responsible for setting-up such a scheme, whereas various research institutes then re-use the data for their analysis or research.
- Organisational, legal and technical costs would be associated with such a certification scheme

Costs for data holders and intermediaries:

- Certification will likely have associated costs for the data holders and intermediaries (if relevant), however it would be voluntary

Table 27 – Overview of costs for Establishing a certification scheme for data altruism mechanisms | PO 2

Overview of costs – PO 2[168]

		Data holders		**Data intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
A certification scheme for data altruism mechanisms	*Direct costs*	-Giving consent to make data available - Negotiating data altruism conditions	- Giving consent to make data available (could be recurrent if it is revoked)	- Establish infrastructure to facilitate data altruism 30,000 EUR - Becoming certified	- Maintain data altruism certification	- Establishing scheme -	- Maintaining the scheme/certification - Reviewing public authorities authorized to certify schemes
	Indirect costs	-	-	-	-	-	-

[168] Further data collection will be performed to provide an estimate.

2.3.2.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Could choose to only reuse data that has been shared through certified data altruism schemes, thereby being secured that it is legally compliant to reuse the data
- Save costs because a private company would issue the certification
- The EU would be a first-mover and possibly set a global certification standard considering that this does not yet exist and non-EU27 might copy the EU, as has happened with GDPR
- Increase transparency and trust among data holders, which in return provides data re-users with larger data sets and or insights
- Reduce organisational, technical and legal costs in the long-run because one certification would eliminate the create a new process for every new data altruism project

Benefits for data holders and intermediaries:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- A privately-run certification scheme will likely operate more efficiently thus make is more attractive for data holders and intermediaries to become certified
- In the long-term, the data altruism scheme could possible run on the voluntary scheme due to its success
- Save on legal costs to due to absence of legal fragmentation in various Member States, the certification would provide an automatic and straightforward process for data altruism schemes

Table 28 – Overview of benefits – Establishing a certification scheme for data altruism mechanisms | PO 2

Type of action	Description	Amount	Stakeholders
A certification scheme for data altruism mechanisms	**Direct benefits**		
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*22 Eur million*	**Data (re)users**
	Indirect benefits		
	R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
	Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
	Value of personal data, assumption: the higher the amount of data, the greater the benefit	*Not quantifiable due to lack of data*	**Data (re)users**

	Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	

2.3.2.4.2.3 Findings of the Cost-Benefit Analysis

This policy option carries voluntary costs for the data intermediaries and brings a relative amount of benefits for those, EUR 22 million. The Cost-Benefit Analysis for PO2 finds that the voluntarily certification scheme has a cost-benefit ratio of only 2.3.

2.3.2.4.3 Coherence of the option

No relevant data altruism certification legislation has been identified, however the importance to align with or build on any other existing certification mechanisms to avoid conflicting mechanisms that would create additional costs for Member States and society. In addition, the Commission should consider incentives for voluntary certification. What would encourage the private sector to voluntarily become certified? The Commission could encourage authorities to provide funding for certification or bring experts together to encourage the community to become certified. An example from the banking sector could be the PCI-Payment Card Certification- which is a voluntarily credit card verification certification that has been adopted voluntarily globally and even become contractual requirements because the financial ecosystem took responsibility to ensure consumer trust and greater transparency.

Overall, this option would be aligned with the Digital Single Market and help build the European Data Economy, while still giving Member States the liberty to control the certification process and not overburdening the private sector and providing additional security measures for personal data.

2.3.2.5 Policy option 3: Regulatory intervention with high intensity

This section assesses the third policy option of Establishing a certification scheme for data altruism mechanisms.

2.3.2.5.1 Effectiveness in achieving the policy objectives

This subsection assess the effectiveness of this policy option.

2.3.2.5.1.1 Achievement of specific objectives

While an administrative approval scheme would increase **trust** of the society in data altruism schemes, it could slow down the building of a common data space in case the public sector does not have sufficient resources to authorise data altruism schemes. This could create a negative impact on the data economy because it could discourage data holders to seek authorisation to participate in data altruism schemes. This could then limit the **social impact** of data altruism for public good, because if less data is available for the data re-users, the lesser the impact will be.

2.3.2.5.1.2 Achievement of general objectives

A strong and regulated data altruism scheme could either accelerate the creation of a **Single Market for Data,** or could completely put it at a halt should Member States not have the capacity to set-up and manage an authorisation authority.

2.3.2.5.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

2.3.2.5.2.1 Costs of the option

Costs for data (re)users

- High cost to create a national authorisation body
- High organisational, legal and technical costs
- High human resources costs to find and or skill people to manage data altruism schemes in the Member States
- Could lead to increased costs if the public body does not have the capacity to handle and process the certification requests
- Could limit the amount of data for re-users

Costs for data holders and intermediaries:

- Increased costs due to mandatory authorisation
- Increased costs if the data altruism authorisation process is slow or poorly managed by the public sector
- In the worst case scenario, this could deter data holder and intermediaries to contribute to data altruism schemes which would lead to less data being shared

Table 29 – Overview of costs for Establishing a certification scheme for data altruism mechanisms | PO 3

Overview of costs – PO 3[169]

		Data holders		**Data intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
An authorisation scheme for data altruism mechanisms	*Direct costs*	-Giving consent to make data available	- Giving consent to make data available (could be recurrent if it is revoked)	Becoming authorized (if applicable 3,800-10,500 EUR depending on the size of the organisation Establish scheme/authorisation process and national oversight body Non-quantifiable, however every EU27 state has a data authority (or equivalent) that could implement this.	- Maintain data altruism authorisation (if relevant) 5,000 EUR	- Establishing scheme/certification and national oversight body N.a. because this will be covered by the established authority under Measures facilitating secondary use of sensitive data held by the public sector	- Maintaining the scheme/authorisation

[169] Further data collection will be performed to provide an estimate.

	Indirect costs	-	-	-	-	-	-

2.3.2.5.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Benefits for data re-users:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- Possibly increased data sharing by SMEs, NGOs and private citizens, although this could be limited if the authorisation process is slow

Benefits for data holders and intermediaries:

- Increase of trustworthiness, security and possibly awareness of data altruism schemes
- Possibly increased data sharing by SMEs, NGOs and private citizens

Table 30 – Overview of benefits – Establishing a certification scheme for data altruism mechanisms | PO 3

Type of action	Description	Amount	Stakeholders
An authorisation scheme for data altruism mechanisms	**Direct benefits**		
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	**Data holders**
	New business relationships with other stakeholders (e.g. data holders)	*Not quantifiable due to lack of data*	**Data intermediaries**
	Easy and transparent way to access data of various fields, contributing to research and development as well as improved decision-making	*300 EUR million*	**Data (re)users**
	Indirect benefits		
	R&I and competition advancement Impact on policy- and decision-making	*Not quantifiable due to lack of data*	**Data holders, Data producers**
	Contribution to a societal goal If donated to research, development of new scientific insights (including medical research, leading to lives saved)	*Not quantifiable due to lack of data*	**Data donors**
	Value of personal data, assumption: the higher the amount of data, the greater the benefit Possibility to enter new business sectors, research fields, generation of new correlation of data -> new insights	*Not quantifiable due to lack of data*	**Data (re)users**

The high intensity option, compulsory authorisation, would lead to benefits of EUR 300 million because a compulsory certification will lead to more data altruism, due to increased trust in the certified mechanisms, and therefore more shared data. In this calculation all 500 companies are certified and all 5 million data holders (of the ones willing to participate in data altruism) provide their data altruistically. In the first year this already leads to a benefit of EUR 50 million, the value of a data set remains EUR 10, and increases annually by 10% as was assumed for PO2 as well. This increase is again due the impact that more companies will enter the market, seeing the benefits, and more data holders will be willing to share their data considering the increased benefits reaped. Therefore PO3, the high intensity option, will achieve a EUR 300 million benefit. This benefit is so much larger, than compared to PO2, because data altruism is still very new and calls for many privacy concerns. Mandatory certification could, likely will, increase the trust of data holders to share data and more shared data, which has a market value although not being used commercially considering it is for research purpose, will lead to higher benefits. In the case of healthcare for example, more shared data could save lives or be the solution to a pandemic.

The key argument is that while the costs are initially high, the recurring costs are very low, and the benefits of the volume of data shred strongly outweigh the costs.

2.3.2.5.2.3 Findings of the Cost-Benefit Analysis

Policy Option 3 entails significant benefits including an acceleration of data altruism mechanism in the European Union due to the mandatory certification mechanisms. This will especially establish an increased trust into data altruism mechanism which will increase the amount of data shared by data holders. This will then lead to the substantial benefit of EUR 300 million, which translates to a Cost-Benefits Ratio of 6.3, as referenced in Annex I .

2.3.2.5.3 Coherence of the option

The option is coherent with the Digital Single Market strategy and aim to build the data economy, however obliging Member States to create a national authorisation authority could create additional costs for them and thereby society. In addition, it could potentially increase the burden of regulation on Member States and businesses, thereby slowing down the process of data altruism scheme implementation in Europe.

2.3.2.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 31 - Summary of impacts for Establishing a certification scheme for data altruism mechanisms

Economic impacts	(Depending on the shared data) • Increased GDP due to improved policy making • Fairer access to data altruism schemes • Increased transparency and security of data altruism schemes • Enable all organisations and companies to participate in data altruism, not just the firms that have sufficient funds to navigate the ICT and legislative labyrinth
Social impacts	(Depending on the shared data) • Improved social policy due to improved policy making based on data (insights) • Empower data holders
Environmental impacts	(Depending on the shared data) • Improved environmental policy making due to better data insights

Fundamental rights impacts	(Depending on the shared data) • Improved policies concerning fundamental rights due to better/more data insights • Empower data holders and increase citizens ability to make their data available for the public good

2.3.3 Establishing a European structure for governance aspects of data sharing

Under this domain, we started by identifying and selecting data points related to the indicators about the costs and benefits of deciding and implementing data standards that effectively contribute to foster data sharing within and across sectors in Europe. These indicators - linked to the causes of the cost occurring due to not sharing data - were categorised and quantified to the extent possible.

Because of the nature of the policy option, the costs and benefits are very indirectly related to the policy measure. An expert group or even the "hard" regulatory option of a legal body has very limited influence over the development and adoption of data standards, as ultimately the decision on how the standard is designed and whether to use it lies with companies.

As the problem analysis (reproduced in the table below) shows, the logical links are clear.

1. Data sharing is opening up huge opportunities for efficiency gains. In the manufacturing sector alone, data sharing accounts for 80% of the potential efficiency gains, equal to 1.4 trillion Euros by 2027.
2. There is consensus that the main barriers to data sharing are trust and interoperability. Companies do not want to lose the control over the data and are wary of potential risks from misuse. The costs for establishing ad hoc agreements is too high.[170]
3. These barrier is precisely the target of initiatives such as IDSA, IHAN and iShare. They provide scalable standardised tool to implement data sharing on a large scale while ensuring the control of the data holder over the data. These tools include technological protocols, process templates and legal agreements. However, these initiatives are at an early stage and being applied still at limited scale – there is no robust evidence about their traction and impact.
4. The four policy options do not envisage the establishment of a standard or standardisation activities, but a set of accompanying activities to support the development and adoption of data standards.

Table 32 - overview of the problem analysis

Ongoing initiatives	Causes	Problem	Effects
Standardisation and coordination initiatives	Lack of data and metadata standards , data schemes within sectors Lack of technical interoperability across sector	Lack of data sharing within/across sector	Lower productivity and innovation

Because policy option 1, 2 and 3 are similar in nature and vary only by the level of formality of the group, the impact will be similar across option. The main difference lies in:

[170] The Everis study on data sharing places technical interoperability as the most mentioned obstacle, by 73% of companies. Legal uncertainty about data ownership is the second, with 54%, and control over usage the third with 42%. The Deloitte studies reports costs of normalizing data, lack of standard protocols, cumbersome legal procedures, involuntary disclosure of commercial secrets as the main barriers. The WEF "Share to gain" report identifies standars, trust and legal arrangement as the key enablers.

- Increased costs for the more institutionalized options
- Increased benefits from a greater possibility to foster the adoption of standards.

The benefits framework is reproduced below.

Table 33 - overview of the cost and benefits framework

		Types of stakeholders (economic impact)					
Types of Impact		**Data holders**	**Data re-user same sector**	**Data re-user other sectors**	**Data intermediaries**	**Society**	**Environment**
Direct Benefits	**Costs Savings**	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	Easier reuse of data, lower cost of data processing and management	
	Efficiency gains	Greater access to data	Greater access to data	Greater access to data			
	Time savings	n.a					
	Resource / energy savings	n.a					
	Revenues/ user charges	n.a.					
Indirect Benefits	**New products and services**	Faster development of new products and services Anticipate disruption by players from different sectors	Faster development of new products and services Anticipate disruption by players from different sectors	New markets	Greater demand for data intermediaries	New products and services	
	Digital single market	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data	Greater access to data from other countries Reduced friction in reusing data		

	Other non-monetisable benefits					New discoveries for health	Environmental efficiency and new products
	Health, Safety & Security improvement						
Direct Costs	**CAPEX**			✓	Cost of setting up governance bodies	n.a	n.a
	OPEX	Cost of participating to standardization activities	Cost of participating to standardization activities	Cost of participating to standardization activities		n.a	n.a
	R&D Costs	n.a.	n.a.	n.a.	n.a.	n.a	n.a
	Implementation	n.a.	n.a.	n.a.	n.a.	n.a	n.a
	Training	n.a.	n.a.	n.a.	n.a.	v	n.a
	Compliance costs	Cost of complying to standards	Cost of complying to standards	Cost of complying to standards			
	Administrative burden	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Indirect Costs	**Compliance costs**	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Digital single market	Greater competition within and across sectors and countries	Greater competition within and across sectors and countries				

2.3.3.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 34 – Overview of stakeholders affected by policy options for Establishing a European structure for governance aspects of data sharing

Who?	**How?**
Data holders	Companies would benefit from the possibility to share data and reap the opportunities of data driven innovation

Intermediary initiatives	European initiatives, data spaces and data standardisation initiatives would benefit from increased awareness and standardisation
Data (re)users	Same as data holders, as data sharing happen between companies in a peer to peer fashion. But with the addition of data companies who would benefit from access to data

2.3.3.2 Policy option 0 – no action at EU level

1.3.1.1 Effectiveness in achieving the policy objectives

1.3.1.1.1 Achievement of specific objectives

The current status is slowly progressing towards the achievement of the objectives. Traditional standardization efforts are ongoing within sectors. Several initiatives (illustrated in the problem analysis) are addressing the full architecture of data sharing, such as iShare, IHAN, IDSA. These initiatives are mostly national but are trying to grow at European level. They encounter different rates of success in achieving the buy in of business not only in the standard definition (where they managed to achieve momentum, for instance IDSA involves 100 leading EU companies) but also in the adoption of the standards.

This will also result in limited access to data, because of the additional cost of agreement on data standards and most importantly on the legal and trust framework, which have to be developed ad hoc in the absence of such protocols and standards.

According to interviewed experts, the present rate of growth in developing and adopting data schemes and standards is slow and unlikely to achieve the promised 1.4 trillion euros benefits by 2027.

1.3.1.1.2 Achievement of general objectives

Because of the slow progress towards the digital single market for data. There is a clear risk that the slow progress towards data sharing could reduce the capacity of European industry to guarantee the sovereignty over their data and standards, opening the way for competitors from third countries to enter the European market and potentially achieve a gatekeeper role for industrial data – just as platforms did for personal data.

In other words, the policy option 0 allows for a slow progress that might expose the European economy to strong risks.

1.3.1.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

1.3.1.2.1 Costs of the option

Traditional business would spend limited time in developing standards as well as in adopting them, hence face no additional costs.

Similarly, there would be no additional costs for technological companies, intermediaries as well as for the European Commission

1.3.1.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

There would be no benefits in terms of reducing costs for data sharing and reuse. For traditional business, the current situation entails very high costs for data sharing and reusing, because of the lack of standards. Taking the analogy of scientific data, the application of FAIR principles would reduce the time spent integrating data by 30%.

Similarly, trust would continue to be managed on a bilateral level, hindering large scale initiatives such as the data spaces.

As a result, data would be insufficiently shared. There would be only a moderate increase in data access and reuse. This would limit the capacity for productivity gains in traditional sectors that are described in the problem analysis. For instance, the gains seen in the iShare project would only slowly be apparent in other sectors and countries. Existing standardization initiatives will scale very slowly.

Intermediaries will then grow slowly, based on individual negotiations.

In the short term, however, it would protect business from increased competition by new entrants as well as competitors.

On a similar note, data companies would have less opportunities to develop new products and services because of limited data sharing.

The specific lack of standard on metadata for machine learning would not improve the risk for bias and incorrect decisions, as well as exposing to human rights violation, as shown by recent examples. It would potentially lead to a limited adoption of machine learning.

1.3.1.2.3 Findings of the Cost-Benefit Analysis

Option zero entails, by definition, no additional costs for any stakeholder. The option would maintain the current level of adoption for data schemes for data sharing between companies, which is marginal, thereby leading to no efficiency increases.

1.3.1.3 Coherence of the option

As there would be no intervention, coherence with existing legislation can largely be confirmed. However, there is a case for arguing that the issues identified as part of the problem assessment would hinder the achievement of a Single Market for Data. In particular, there seems to be a missing link with the creation of data spaces included in the data strategy.

2.3.3.3 Policy option 1 – Informal expert group

1.3.1.4 Effectiveness in achieving the policy objectives

1.3.1.4.1 Achievement of specific objectives

The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

It is clear that the creation of an informal expert group will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives. The informal nature of the group raises questions on its capacity to achieve an impact, as it represents the weakest of the possible options. It will however allow for a more varied set of expertise, which is appropriate in view of the fast-evolving nature of data sharing.

The agile nature of the expert group is also more complementary with the large ecosystem of existing instruments and institution for fostering standard creation and adoption.

1.3.1.4.2 Achievement of general objectives

On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated.

Moreover, the expert group nature might allow for closer collaboration with industry than a legal body, thereby helping with the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard.

1.3.1.5 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 1.

1.3.1.5.1 Costs of the option

Cost for business:

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1 phone call per week (average 3 h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US- EU- Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the companies.

Cost for intermediaries:

- Cost of paying for the expert group. Expert groups are typically only reimbursed for expenses, but only in exceptional cases for work. The costs are therefore limited to around 200 Euros per day per person – so a group of 10 experts would cost about 24,000 Euros per year based on the engagement above (four three-days meetings a year)..
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 35 – Overview of costs for Establishing a European structure for governance aspects of data sharing | PO 1

Overview of costs – PO 1

		Data holders		**Intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure	*Direct costs*	-	200,000/year for running the group	-	24,000/year for running the group	-	As data holders

for governance aspects of data sharing	*Indirect costs*	-	-	-	-	-	-

1.3.1.5.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.
- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500.000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700,000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing data schemes is below 100, we can estimate conservatively that under this option 700 companies would be involved by 2025.
- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.
- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 36 – Overview of benefits – Establishing a European structure for governance aspects of data sharing | PO 1

Type of action	Description	Amount	Stakeholders
	Direct benefits		
European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

1.3.1.5.3 Findings of the Cost-Benefit Analysis

Option 1 would lead according to our estimations validated by experts, to a small increase in adoption of such schemas, estimated in 700, or 0,1% of the current number of "data users" in the EU data market study (700.000).

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 700 companies would benefit from an operational efficiency of 15% of an average OPEX of 50M EUR in 5 years, as estimated by interviewed experts. This leads to an average 1050 M EUR benefit per year, or 5250 M EUR in five years (NPV 4.668 M EUR).

As regard the costs, we estimate no additional costs for the European Commission since the participation to the experts group is voluntary, only a reimbursement of expenses for about 24.000 EUR per year. Costs by participating companies would be for a total of 200.000 EUR per year. This is based on an estimate of four three-day meetings a year by ten company representatives.

The final results is a NPV of 4668 Million EUR and a BCR of 200.362 M EUR. However, the BCR is not the most reliable indicator because of the very small scale of expenses.

1.3.1.6 Coherence of the option

The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

2.3.3.4 Policy option 2 – Formal expert group

1.3.1.7 Effectiveness in achieving the policy objectives

1.3.1.7.1 Achievement of specific objectives

The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

It is clear that the creation of a formal expert group will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives. The formal nature of the group is likely to provide some kind of authority, while not requiring the investment that is devoted to a legal body, and allowing for a more varied set of expertise, which is appropriate in view of the fast-evolving nature of data sharing.

The agile nature of the expert group is also more complementary with the large ecosystem of existing instruments and institution for fostering standard creation and adoption.

1.3.1.7.2 Achievement of general objectives

On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated.

Moreover, the expert group nature might allow for closer collaboration with industry than a legal body, thereby helping with the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard.

1.3.1.8 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

1.3.1.8.1 Costs of the option

Cost for business

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1 phone call per week (average 3 h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US- EU- Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the funder of the expert group, usually the European Commission, under the budget below.

Cost for intermediaries:

- Cost of paying for the expert group or scientific committee. Formal expert groups and scientific committees include a daily fee for work in addition to expenses. The costs are higher than for informal expert groups. For instance, the current Scientific Committees cost on average 280.000 euros per year (four three-day meetings a year).[171] This average includes travel costs and miscellaneous costs.
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 37 – Overview of costs for Establishing a European structure for governance aspects of data sharing | PO 2

Overview of costs – PO 2

		Data holders		**Intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure for governance aspects of data sharing	*Direct costs*	-		-	280,000/year for running the group	-	As data holders
	Indirect costs	-	-	-	-	-	-

1.3.1.8.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.

[171] European Commission, 2016. Report on the activity of the scientific committee's term 2013-2016.

- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500,000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700.000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing data schemes is below 100, we can estimate conservatively that under this option 800 companies would be involved by 2025.
- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from reuse of scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.
- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 38 – Overview of benefits – Establishing a European structure for governance aspects of data sharing | PO 2

Type of action	Description	Amount	Stakeholders
	Direct benefits		
European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs* *15% overall efficiency gain*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**

	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

1.3.1.8.3 Findings of the Cost-Benefit Analysis

Option 2 would lead according to our estimations, validated by experts, to a slightly greater increase in adoption of such schemas, estimated in 800, against the 700 of option 1. Interviewed experts were adamant that we cannot expect a major difference in terms of adoption between an informal and formal expert group, which cannot have any way to directly act (not to mention enforce) companies' choice for adoption of a specific standard. In other words, the general effect of such a soft policy measure as an expert group on adoption by companies is limited, as it is heavily mediated by other factors, and the difference between a formal and informal expert group is expected to be minor.

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 800 companies would benefit from an operational efficiency of 15% of an average OPEX of 45M EUR in 5 years, assuming that the companies joining are typically a bit smaller than in option 1. This leads to an average 1.200,0 M EUR benefit per year, or 6000 M EUR in five years (NPV 4. 5335,6 M EUR).

As regard the costs, we estimate some additional costs for the European Commission with respect to option 2, for the reimbursement of travel expenses and a fee for participation, for a total of 280.000 (based on experience with other formal expert groups). This is based on an estimate of four three-day meetings a year by ten company representatives.

The final results is a NPV of 5.335,3 Million EUR and a BCR of 19.627,3 M EUR. However, the BCR is not the most reliable indicator because of the very small scale of expenses.

1.3.1.9 Coherence of the option

The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

2.3.3.5 Policy option 3 – Legal body

1.3.1.10 Effectiveness in achieving the policy objectives

1.3.1.10.1 Achievement of specific objectives

The policy option is suitable to the achievement of the specific objectives, notably:

- Ensuring trust, by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing
- Facilitating interoperability, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors

However, it is clear that the creation of a legal body will only provide an indirect impact on the achievement of the objectives, as it will merely facilitate the activities of the existing standardization initiatives.

There are already well-developed instruments and institution for fostering standard creation and adoption.

1.3.1.10.2 Achievement of general objectives

On a similar tone, the present option is certainly in line with the objective to increase data sharing and create a European Single Market for data, although the relation is indirect and heavily mediated. Moreover, it does not offer particular advantages with regard to the variable that will determine the ultimate success: guaranteeing the participation of companies in the definition and in the adoption of the standard. Indeed, a legal body could even prove counterproductive in a context with multiple actors and initiatives in place, and where leadership sit squarely with business.

1.3.1.11 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

1.3.1.11.1 Costs of the option

Cost for business

- Standardization is costly, in particular to ensure wide adoption. Full interoperability based on the FAIR principles requires 5 to 10% of a research project budget. So there will be additional costs as in the case of all standards. However, these additional costs will be ultimately part of the decision process that is controlled by industry as the policy option does not entail any mandatory standards.
- Another important cost category refers to participating in the expert groups activities, which is expected to be additional to other effort in the field of standardization. One interviewee summarizes the effort as the involvement of 1 to 7 people in a company team, engagement in 1

phone call per week (average 3 h), 1 meeting each 3 months that spans from 3 to one full week. This entails high travel expenses via continental flight (overseas because combines US- EU- Asia travels) and 1-week hotel every 3 months in average. The IDSA experience suggest the need for 20% of a full time equivalent per company. These costs will have to be covered by the legal body as part of the budget below.

Cost for intermediaries

- Cost of setting up and running the legal body which are higher than the expert groups. Based on comparable initiatives, such as the European Data Protection Board, we estimate a yearly budget of 3.5 million euros.
- Cost of standard documentation and education. Costs of guidelines, toolkits, tutorials, webinars etc. It is discretionary to the team that lead the standardization effort and how much they want to invest on the engagement – but there is a direct correlation between the width of the support measures and the level of uptake. These costs are highly variable and are difficult to estimate.

Table 39 – Overview of costs for Establishing a European structure for governance aspects of data sharing | PO 3

Overview of costs – PO 3							
		Data holders		**Intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
European structure for governance aspects of data sharing	*Direct costs*	-		-	3.5 million euros/year	-	As data holders
	Indirect costs	-	-	-	-	-	-

1.3.1.11.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For traditional business:

- Increased clarity on the existing standard and standardization initiatives, leading to a more informed choice by business over the standards to adopt.
- Increased adoption of standards, leading to reduction in costs for acquiring, integrating and processing data. Lack of technical standards for data and metadata is considered a major barrier to data sharing in traditional sectors.
- Adoption of standards provides easier access to data. Estimates from individual case studies show that adoption of standards for data sharing results in increased adoption. For instance, in the case of sport activities, the openactive.io standard led to 200,000 new activities being posed by 29 organisations, resulting in 150 to 500,000 new activities carried out every month.
- This will lead to accelerating the progress towards achieving the benefits from data sharing – enable the achievement of the high growth scenario put forward by the EU data market study, or to grasp the opportunities of the Internet of Things outlined in the Deloitte study. IDSA estimate an efficiency gain of 15% by the adoption of the reference architecture. Considering a potential addressable market of 700.000 companies (the data users identified by the EU data market study), and considering that the current number of companies involved in implementing

data schemes is below 100, we can estimate conservatively that under this option 900 companies would be involved by 2025.

- For the scientific domain alone, FAIR data introduces efficiency measures by at least 10 billion euros, by reducing the effort for data reuse and avoiding duplication of data collection. When it comes to innovation, the expected gains from scientific data are quantified in 16 billion euros annually.

For intermediaries:

- Increased capacity to scale up standardization initiatives.
- Increased market and demand for standardization from traditional business.
- Reduced costs for the initiatives. The typical annual budget in the initial phase for such initiatives is around 5 million euros (for AboutML and iShare). Replication would allow for economies of scale.

For society:

- Lower costs for consumer and greater competition
- New products and services, particularly important for health.
- Reduction of bias in algorithms

For environment

- Reduction in carbon footprint due to increased efficiency

Table 40 – Overview of benefits – Establishing a European structure for governance aspects of data sharing | PO 3

Type of action	Description	Amount	Stakeholders
	Direct benefits		
European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		

	New products and services Anticipate disruption by players from different sectors	*1,4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**
	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

Type of action	**Description**	**Amount**	**Stakeholders**
	Direct benefits		
European structure for governance aspects of data sharing	Costs Savings and efficiency gains - Easier reuse of data, lower cost of data processing and management	*30% reduction in data processing costs* *15% overall efficiency gain*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries, the Society**
	Greater access to data for faster development of new products and services Anticipate disruption by players from different sectors	*10 bn euros/year cost of not having FAIR research data*	**Data holders in traditional industries**
	Possibility to access new markets	*Not possible to predict*	**Data re-users in same sector**
	Greater access to data sources for analytics and machine learning as well as development of new products and services	*16bn euros/year from innovative reuse of FAIR data*	**Researchers, Data holders in traditional industries, Tech companies**
	Indirect benefits		
	New products and services Anticipate disruption by players from different sectors	*1.4 trillion euros by 2027 for manufacturing*	**Data holders, Data re-users in same sector, the Society**
	Greater demand for data intermediaries	*Not possible to predict*	**Data intermediaries**
	Digital single market – greater access to data from other countries Reduced friction in reusing data	*Not possible to predict*	**Data holders, Data re-users in same sector, Data re-users in other sectors, Data Intermediaries**

	Other non-monetisable benefits: new discoveries for health, environmental efficiency and new products	*Not possible to predict*	**Society**

1.3.1.11.3 Findings of the Cost-Benefit Analysis

Option 3 would lead according to our estimations validated by experts, to a slightly greater increase in adoption of such schemas, estimated in 900, against the 700 of option 1 and 800 for option 2. Interviewed experts were adamant that we cannot expect a major difference in terms of adoption between an expert group and a legal body, as the initiative should remain mostly with industry and government should not be in a position of picking winners among standards. The legal body cannot have any way to directly act (not to mention enforce) companies' choice for adoption of a specific standard. In other words, the general effect of such a soft policy measure as a legal body on adoption by companies is limited, as it is heavily mediated by other factors, and the difference between an expert group and a legal body is expected to be minor.

Because of the nature of this measure, which entails the peer to peer sharing of data between companies, there is no distinction between data holders and reusers.

These 900 companies would benefit from an operational efficiency of 15% of an average OPEX of 40M EUR in 5 years, assuming that the companies joining are typically a bit smaller than in option 1 and 2. This leads to an average 1350 M EUR benefit per year, or 6750 M EUR in five years (NPV 6.002,5 M EUR).

As regard the costs, we estimate additional costs for the European Commission with respect to option 2, related to fixed costs for the organization for a total of 3,5 M EUR (based on analogy with other legal bodies such as the European Data Protection Board. This includes also the budget for travelling and reimbursement of speakers' fees.

The final results is a NPV of 5.999,1 Million EUR and a BCR of 1.766,5 M EUR.

1.3.1.12 Coherence of the option

The option is fully coherent with the Digital Single Market strategy and will be highly complementary with the creation of data spaces.

2.3.3.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 41 - Summary of impacts for Establishing a European structure for governance aspects of data sharing

Economic impacts	• costs for business ○ staff for participating to standardisation meeting • benefits for business ○ efficiency from data sharing • costs for intermediaries : ○ staff for participating • benefits for intermediaries : ○ increased demand and awareness of their solutions • costs for European Commission ○ Cost of running expert group
Social impacts	• Increased social innovation such as new drug discoveries
Environmental impacts	• Increased energy efficiency through sharing of consumption data

Fundamental rights impacts	• Reduced risk for algorithmic bias and discrimination through dataset metadata

2.3.4 Establishing a certification framework for data intermediaries

2.3.4.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 42 – Overview of stakeholders affected by policy options for Establishing a certification framework for data intermediaries

Who?	How?
Data holders	Data holders will not be directly affected in terms of costs, but will benefit mainly from data monetisation via the data sharing through certified intermediaries, cost and time savings through digitisation of the transactions, increased control over their data and increased revenue generated from the network growth the increased volume of data sharing.
Data intermediaries	Certification will help data intermediaries to scale up and grow in terms of revenue, resources, client base and volume of data transactions. However, they will be the main stakeholders affected directly from the certification cost.
Data (re)users	Data re-users will benefit mainly from cost and time savings through digitisation of the transactions, easier access to data and the creation of mechanisms to assess the quality of data intermediaries' services. They might also face some indirect certification costs due to potentially increased charges of the certified data intermediaries' services.
Society	Societal benefit will be twofold. On the one hand society will benefit as the potential of the European data market will be unlocked through certification. On the other hand, data flows of intermediaries serving societal purposes (i.e. health, research) will be increased.

2.3.4.2 Policy option 0: Baseline scenario- No horizontal action at EU level

1.3.1.13 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of the baseline scenario in achieving the specific and general policy objectives.

1.3.1.13.1 Achievement of specific objectives

The absence of horizontal action at EU level would not allow to create **trust** in common European data spaces, neither to build **common data spaces** that is making more data usable where data holders could agree to it through technical, legal and organisational support. Finally, current data **interoperability** issues across sectors would persist.

Since the data intermediaries would remain uncertified, there would remain a lack of trust and mechanisms for data holders and users to assess their "neutrality", while at the same time there would be no differentiation between neutral and non-neutral data intermediaries. This will lead to further interoperability issues and regulatory fragmentation in the internal market. Ultimately, there will be less data available for reuse across sectors. This will prevent the EU from reaping the full benefits of horizontal data sharing which account for 20% of all the benefits of data sharing in general[172].

In particular, according to the summary report of the open public consultation on the European strategy for data, almost 80% of the 512 respondents have encountered difficulties in using data from other companies, related to technical aspects (data interoperability and transfer mechanisms),

[172] Realising the economic potential of machine-generated, non-personal data in the EU, Deloitte Report for Vodafone Group, July 2018

denied data access, and prohibitive prices or other conditions considered unfair or prohibitive. A very large share of respondents (87.7%) supported the idea that the EU should make major investments in technologies and infrastructures that enhance data access and use, while giving individuals as well as public and private organisations full control over the data they generate. Around the same proportion of respondents considered that the development of common European data spaces should be supported by the EU in strategic industrial sectors and domains of public interest. [173] This idea cannot be supported by the policy option 0 and the absence of action at EU level.

1.3.1.13.2 Achievement of general objectives

The absence of horizontal action at EU level would not contribute to setting **the foundations of a Single Market for Data, neither to strengthen the EU data economy.** Since the data intermediaries would remain unregulated, such businesses would not be empowered with respect to the data use they generate to create value for the society.

1.3.1.14 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

1.3.1.14.1 Costs of the option

The majority of stakeholders interviewed expressed opinions against the baseline scenario and the lack of action at EU level, due to the costs arising from the lack of trust within the market, which does not allow the potential of data to be unlocked and the data intermediaries to grow. Furthermore, data holder and data users would also have to bear the cost of more expensive transactions, while losing time savings and efficiency gains from transactions that would be otherwise facilitated by the certified intermediaries.

1.3.1.14.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The absence of action at EU level could create cost and time savings for data intermediaries who were against certification, as they would not need to bear the cost of obtaining and maintaining the certification, neither to have a competitive disadvantage compared to certified data intermediaries.

1.3.1.14.3 Findings of the Cost-Benefit Analysis for Policy Option 0

Policy Option O entails no significant costs and benefits for the stakeholders, since no action is taken. The assessment of both costs and benefits under the baseline scenario is linked to the current absence of a certification framework for data intermediaries in the European market.

1.3.1.15 Coherence of the option

The absence of action at EU level would not change the status quo, therefore coherence with the existing EU policy and legal framework is ensured. However, incoherence issues might arise from the fact that the provisions of the European Data Strategy for (i) data flow within the EU and across sectors, for the benefit of all (ii) and the rules for access and the use of data are fair, practical and clear would not be strengthened or promoted by the maintenance of the baseline scenario.

[173] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-european-strategy-data

2.3.4.3 Policy option 1: Coordination at EU level (industry driven self-regulatory certification framework)

1.3.1.16 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the specific and general policy objectives.

1.3.1.16.1 Achievement of specific objectives

The establishment of a self-regulatory certification framework could contribute into the three specific policy objectives however not in a significant way. Stakeholders from both B2B and C2B interviewed expressed concerns that policy option no. 1 might have little added value compared to Policy option 0. There is already a self-regulatory effort being currently conducted in the frame of MyData Community for personal data operators.

Concerns were also raised that big industry players would have a stronger role and could potentially influence the outcome of the discussions taking place in the frame of the stakeholder forum.

1.3.1.16.2 Achievement of general objectives

In line with arguments presented in the above mentioned section, within the limits of its effectiveness this policy option could promote and contribute into setting the foundations of a single market for data as well as strengthen the EU data economy as a broader impact, although in a limited fashion.

1.3.1.17 Efficiency: Costs and benefits of the option

Policy option 1 was not examined for a detailed cost-benefit analysis. During stakeholder interviews, several **data intermediaries in both B2B and C2B markets expressed concerns that big industry players would have a stronger role compared to SMEs** and **could potentially influence in their favour the outcome of the discussions** taking place within the stakeholder forum. The absence of a regulatory measure adopted at EU level would not help ensuring a **fair and well-representative selection of certification criteria/requirements** for all the various types of data intermediaries active in the European market. Furthermore, **there is already a completely industry-driven, self-regulatory certification (self-description) process in place**, initiated in 2020 within the MyData Community for personal data spaces. For these reasons, it was estimated that this policy option would have **limited added-value and impacts** compared to the baseline scenario, and therefore the option was not further considered for detailed cost and benefits analysis. Indicative categories of costs and benefits are provided below.

1.3.1.17.1 Costs of the option

Data intermediaries would have in this case to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria decided during the stakeholder forum discussions, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. Given that this is a self-regulatory approach, the requirements are not expected to be hard. This affects the effectiveness and efficiency of the option, but also limits the related costs. In particular the expected costs by the stakeholders interviewed, linked to this policy option, are 15K to 20K EUR one-off costs for obtaining the certification for the first time (including 10K EUR internal preparatory costs and 5K-10K EUR external certification costs) and 10K to 20K EUR/year recurrent costs for renewing it (including 5K-10K EUR internal preparatory costs and up to 5K EUR external certification costs).

In this policy option data intermediaries might have to face also costs of **setting up the scheme** and potentially funding a private **certification agency**, depending on how this framework is being designed. Time and resource costs are also involved for the participation to the stakeholder forum discussions.

In terms of indirect impacts, the **competition on the market** is expected to be between 25%-1% decreased in both C2B and B2B data sharing markets, within the first year after the first operators will obtain the certification.

Data re-users might be affected as well from the certification from **indirect transaction costs**, as the certified intermediaries might increase the user charges to cover the certification cost.

1.3.1.17.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Depending on its effectiveness described in the previous section, policy option 1 could provide benefits in all the stakeholders of the value chain, due to the **trust between the actors** that would be brought in the ecosystem.

This trust could lead into further **efficiency gains and time savings, increase in the client base and data transactions and therefore increase of revenues,** allowing data intermediaries to scale up but also other stakeholders in the value chain. However, if the effectiveness rate of this option remains low, then the benefits would be lower as well, compared to the full certification potential under other policy options. Given that this is a self regulatory approach, the requirements are not expected to be hard. This affects the effectiveness and efficiency of the option, but also limits its associated benefits. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 20%-25% business development time acceleration
- Growth in terms of revenues and employee number:
    - between 25-35% growth, within the 1st year after obtaining the certification;
    - between 25-35% growth, from 2 to 5 years' timeframe;
    - up to 25% growth, beyond 5 years' timeframe
- Growth in terms of client base (including both number of clients and number use cases)
    - up to 25% increase within the 1st year after obtaining the certification;
    - 25% increase from 2 to 5 years' timeframe;
    - 25%-35% increase beyond 5 years' timeframe.

Data holders would have the opportunity to monetise more from data sharing while more individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while at the same time they would have a kind of mechanisms for assessing the quality of services provided by data intermediaries.

In terms of indirect benefits, while certification is expected to limit market competition in a short-term plan, the opposite impact is expected in a longer-term plan. In particular, the interviewed stakeholders expect a competition increase in both C2B and B2B data sharing markets of 1%-25% from 2-5 years’ timeframe after obtaining the certification, and a further 1%-25% increase in a beyond 5 year’s timeframe.

1.3.1.17.3 Findings of the Cost-Benefit Analysis for Policy Option 1

Policy Option 1 entails relatively low levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market after certification. These benefits are also linked to relatively low levels of recurrent and one-off costs for data intermediaries in order to obtain and maintain the certification. However, the expected benefits significantly outweigh the costs. Details on the costs and benefits for policy option 1 are presented in Annex I.

1.3.1.18 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

2.3.4.4 Policy option 2: Regulatory intervention with low intensity (voluntary certification framework)

1.3.1.19 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the specific and general policy objectives.

1.3.1.19.1 Achievement of specific objectives

This policy option could significantly contribute to the three specific objectives, particularly, in creating trust in common European data spaces, building common data spaces as well as ensuring data interoperability across sectors, through the certification framework. The majority of stakeholders interviewed agreed to this policy option as the most effective one, for the current status of the market, given also that the certification criteria would be defined by a legal instrument. Therefore, if a big number of industry players decides to proceed to the certification process, the trust between the stakeholders involved in the market would be increased significantly, allowing the data intermediary market to flourish and bringing various economic and societal benefits.

1.3.1.19.2 Achievement of general objectives

Following the effectiveness in achieving the specific objectives described in the above section, this policy option would also further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy, since the European data market overall will be significantly boosted through certification, increasing the volume of data flows.

1.3.1.20 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

1.3.1.20.1 Costs of the option

Similarly, to the policy option 1, data intermediaries would have in this case as well to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria defined in the legal instrument, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. Small industry players of early stage growth pointed out the importance of these costs to remain in an affordable level (not exceeding the 20K as a point of reference), and explained that the costs might also involve the need for additional resources (e.g such as lawyers or developers to ensure compliance) depending on how exactly the framework is being designed. If the certification cost cannot be minimised to a reasonable cost for SMEs then it should be subsidised. However, opinions on this vary depending on the growth stage of a company. In particular, a well-established industry player in the B2B market, raised the argument that the certification cost should not be kept low while the criteria should rather be strict, in order to ensure a proper differentiation between the various players in the market. The interviewed stakeholders estimate that certification costs under this policy option would be of 20K – 50K EUR one-off costs for obtaining the certification for a first time (including 10K-25K EUR internal preparatory costs and 10K-25K EUR external certification costs) and of 20K-35K EUR/year recurrent costs for renewing it (including 10-25K EUR/year internal preparatory costs and 10K EUR/year external certification costs).

In terms of indirect impacts, the **competition in the market** might be slightly affected in this policy option as well, once these first intermediaries will obtain the certification, with an expected decrease of 25% in the B2B market. However, in the long term new competitors will be attracted in the market, due to the acceleration of the market and the speed of adoption. Some stakeholders pointed out that **innovation** might also be decreased in the market depending on how strict the criteria will be. On the contrary, other stakeholders face certification as an opportunity for innovation increase, based on the fact that the certification will be significantly beneficial for the market, providing the opportunity for the development of new products or services. A stakeholder representing the personal data market, expressed the opinion that the way the certification requirements will be written will determine whether the certification will limit innovation or not. If the requirements are only about what needs to be proven but not about the way this will be implemented, then it should not limit innovation. Another stakeholder in growth stage in the B2B market highlighted the importance of the need for the framework to represent and benefit the market as a whole, and not only specific big-industry players, as this would lead to both reduced innovation and reduced competition.

Data re-users might be affected as well from the certification from **indirect transaction and implementation costs**, as the certified intermediaries might increase the user charges to cover the certification cost. However not all stakeholders agree to this argument.

Table 43 – Overview of costs for Establishing a certification framework for data intermediaries | PO 2[174]

Overview of costs – PO 2							
		Data holders		**Data intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Certification framework for data intermediaries	*Direct costs*	-	-	**20K – 50K EUR** •10K-25K EUR internal preparatory costs •10K-25K EUR external certification costs	**20K-35K EUR/year** •10-25K EUR/year internal preparatory costs •10K EUR/year external certification costs	-	-
	Indirect costs	-	-	-	Around 25% decreased market competition	-	Non-quantifiable costs due to lack of data

1.3.1.20.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

The majority of the interviewed stakeholders expressed views in favour of this policy option as the preferred one, seeing a broad number of benefits. These mainly include the **trust** between the actors that would be brought in the ecosystem leading into **further efficiency gains, time savings, increase of the client base** and data transactions and therefore **increase of revenues**, allowing

[174] The table presents the estimated amounts of costs by the interviewed stakeholders

data intermediaries to scale up but also other stakeholders in the value chain. An aspect linked to the increased trust, where many stakeholders focused on, is the acceleration of the market, speed of adoption and time savings in business development cycle of data intermediaries. Furthermore, certification under this policy option is expected to guarantee uniform and transparent security levels, interoperability, safety and quality of intermediaries' services. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 25%-50% business development time acceleration
- Growth in terms of revenues and employee number:
  - between 35%-40% increase, within the 1st year after obtaining the certification;
  - between 40%-50% increase, from 2 to 5 years' timeframe;
  - between 40%-50% increase, beyond 5 years' timeframe
- Growth in terms of client base (number of clients)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 40%-50% increase beyond 5 years' timeframe.
- Growth in terms of client base (number use cases)
  - between 35%-40% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 35%-40% increase beyond 5 years' timeframe.

According to a stakeholder of growth stage in the C2B value chain, an increase of approximately 20% in client base and revenues expected after certification, corresponds to additional 30M-50M EUR revenue increase per year.

With regards to the indirect benefits under this policy option, the interviewed data intermediaries expect also increased innovation-related benefits once the market is boosted through the certification, explaining market acceleration will bring new use cases, which wouldn't be viable before certification. Furthermore, increased trust in the market could also lead to increase of funding, as investors will find it safer to invest in certified companies, A stakeholder representing a data union (C2B value chain) focused on the legitimacy-related benefits, to provided under this policy option which is expected to lead to additional efficiency and trust between the company and its members/clients as well as to strengthen the role of newly established C2B data intermediaries, such as data unions or data cooperatives by setting the scene behind their mission.. As a consequence, the market competition would be boosted for these newly established data intermediaries being in growth phase, empowering them towards the monopoly or oligopoly of big industrial players. Increased competition benefits are expected to be met however in both C2B and B2B value chains under this policy option, especially in a long-term timeframe. In particular, the interviewed stakeholders, expect:

- between 1%-25% increased competition in the B2B market from 2 to 5 years' timeframe;
- between 1%-25% increased competition in the B2B market beyond 5 years' timeframe;
- between 1%-25% increased competition in the C2B market, within the 1st year after obtaining the certification;
- between 1%-25% increased competition in the C2B market, from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the C2B market, beyond 5 years' timeframe.

Data holders would have the opportunity to monetise more from data sharing, while they will have easier access to a bigger network of data re-users. At the same time more individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while mechanisms for assessing the quality of services provided by data intermediaries will be created for them through certification.

Table 44 – Overview of benefits – Establishing a certification framework for data intermediaries | PO 2[175]

Type of action	Description	Amount	Stakeholders
Certification framework for data intermediaries	**Direct benefits**		
	Time savings and business development cycle acceleration; speed of adoption of the market	25%-50% business development time acceleration after certification	**Data Intermediaries**
	Company growth in terms of revenue and number of employees	• Between 35%-40% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe • 20% increase corresponds to 30M-50M EUR/year revenue increase for a growth-stage intermediary	**Data Intermediaries**
	Company growth in terms of client base (number of clients)	• between 25%-50% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Company growth in terms of client base (number of use cases)	• between 35%-40% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 35%-40% increase beyond 5 years' timeframe	**Data Intermediaries**
	Data monetisation via data sharing through certified platforms	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
	Cost and time savings, including through digitization of interaction with data (re)users by certified intermediaries and quicker access to data suppliers (in the case of certified intermediaries who provide data pooling services) as well as efficiency gains when sharing supply	*Not quantifiable due to lack of data*	**Data holders/ Data providers, Data (re)users**

[175] The table presents the estimated amounts of benefits by the interviewed stakeholders

chain data through a data intermediary		
Growing network and additional revenue through easier access to data (re)users (in the case of certified intermediaries who provide data pooling services) and larger client base (especially in the case of intermediaries providing data pooling services).	*Not quantifiable due to lack of data*	**Data holders/ Data providers, (Certified) data intermediaries, Data (re)users**
Data users obtain personal data with legal clarity on usability (consented data). Creation of mechanism for intermediaries' clients to assess the quality of intermediaries' operations (i.e. in terms of compliance with legislation) Additional revenue and increased productivity generated through the increased volume of re-used data (unlocked and facilitated through the certified intermediaries)	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Increased market competition – B2B Market	• between 1%-25% increase from 2 to 5 years' timeframe, • between 1%-25% increase beyond 5 years' timeframe;	**Data Intermediaries**
Increased market competition – C2B market	• between 1%-25% increase within the 1st year after obtaining the certification, • between 1%-25% increase from 2 to 5 years' timeframe, • between 25%-50% increase beyond 5 years' timeframe	**Data Intermediaries**
Additional revenue potentially generated through the "re-use" of the data unlocked and facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
Enhancing of trust between the main actors involved	*Not quantifiable due to lack of data*	**(Certified) Data Intermediaries, Data holders/ Data providers, Data (re)users**
Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	
Economic value of data (both personal and industrial) will be maximised and additional revenue potentially generated through the re-use of the data facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data (re)users, Data holders/ Data providers**

1.3.1.20.3 Findings of the Cost-Benefit Analysis for Policy Option 2

Policy Option 2 entails significantly high levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market, after certification. Recurrent and one-off costs are expected as well for data intermediaries in order to obtain and maintain the certification, however, the expected benefits significantly outweigh the expected costs. This policy option presents a great cost-benefit relation, with benefits exceeding more than 10 times the costs. Details on the costs and benefits for policy option 2 are available in Annex I.

1.3.1.21 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

2.3.4.5 Policy option 3: Regulatory intervention with high intensity (compulsory certification framework)

1.3.1.22 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the specific and general policy objectives.

1.3.1.22.1 Achievement of specific objectives

This policy option is expected to contribute to some of the specific objectives, namely creating trust in common European data spaces as well as contributing to data interoperability across sectors. However, concerns were raised regarding its effectiveness to build common data spaces, due to the fact that a compulsory certification process with hard neutrality requirements is likely to prevent small industry players from getting into the market due to the potentially prohibitive certification cost.

1.3.1.22.2 Achievement of general objectives

In the same line of argumentation, there are doubts whether the overall impact of a compulsory certification framework would be positive by boosting the market, since it could create significant burdens for new players to get into the market. It is therefore doubtful whether this policy option could further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy.

1.3.1.23 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

2.3.4.5.1 Costs of the option

The types of costs under this policy option are similar to those of the previous policy options. However, the hard neutrality requirements of a compulsory certification framework are expected to increase the level of such costs. In terms of direct costs, data intermediaries would have in this case as well to bear the cost of **obtaining and maintaining the certification**. Depending on the requirements and the criteria defined in the legal instrument, some intermediaries would also have to face implementation costs to ensure compliance with the requirements. In particular, the interviewed stakeholders estimate that certification costs under this policy option would be of 35K – 75K EUR one-off costs for obtaining the certification for a first time (including 25K-50K EUR internal preparatory costs and 10K-25K EUR external certification costs) and of 20K-50K EUR/year recurrent costs for renewing it (including 10K-25K EUR/year internal preparatory costs and 10-25K EUR/year external certification costs).

In terms of indirect costs, concerns are raised by data intermediaries in relation to the innovation in the market that might be limited, due to hard neutrality requirements defined by such a compulsory framework. Competition in the market might also be reduced in a short-term timeframe, as some companies might fail to enter and remain in the market due to the high certification cost while others in early stage might fail to afford the high certification cost. In particular, within the first year timeframe after the establishment of the certification framework, the interviewed stakeholders expect a decrease of competition between 50% - 25% in B2B market and of 25%-1% in the C2B market. Data re-users are also likely to be affected by **indirect transaction and implementation costs**, as some of the certified intermediaries might increase the user charges to balance the certification cost.

Finally, stakeholders expressed concerns that a compulsory certification framework would be prohibitive for data intermediaries which are not necessarily legal entities, but the intermediation is provided through the form of a project or product. Under this policy option, this type of data intermediation initiatives would have to be locked out of the market.

Table 45 – Overview of costs for Establishing a certification framework for data intermediaries | PO 3[176]

Overview of costs – PO 3

		Data holders		**Data intermediaries**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Certification framework for data intermediaries	*Direct costs*	-	-	35K – 75K EUR •25K-50K EUR internal preparatory costs •10-25K EUR external certification costs	20K-50K EUR/year •10-25K EUR/year internal preparatory costs •10-25K EUR/year external certification costs	-	-
	Indirect costs	-	-	-	• Between 50% and 25% decrease within the 1st year after obtaining the certification in B2B market; • between 25%-1% decrease within the 1st year after obtaining the certification in C2B market	-	Non-quantifiable costs due to lack of data

[176] The table presents the estimated amounts of costs by the interviewed stakeholders

1.3.1.23.1 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

Even though the majority of stakeholders involved was not in favour of this policy option, due to the additional burdens it might create in a yet not mature market, this option still presents a broad number of benefits similar to the policy options 1 and 2, even at higher levels in some cases. These are interwoven to the increased trust between the actors that would be brought in the ecosystem through compulsory certification. It would lead to further time and cost savings for data holders and re-users as well as efficiency gains, increased client base and volumes of data transactions. This would increase revenues for data intermediaries, allowing them to scale up. Additionally, it will guarantee uniform and transparent security levels, interoperability as well as safety and quality of data intermediaries' services. The role of newly established data intermediaries in the market such as data unions or data cooperatives would still be empowered, within their ecosystem, through this legal framework which would provide legitimacy to their work. Additional benefits are expected for data intermediaries who already respect certain neutrality requirements, as they would have a competitive advantage towards the other players, without having to bear a high level of costs for the certification. In particular, the benefits expected by the interviewed stakeholders under this policy option include:

- 45%-50% business development time acceleration
- Growth in terms of revenues and employee number:
  - between 25%-35% increase, within the 1st year after obtaining the certification;
  - between 40%-50% increase, from 2 to 5 years' timeframe;
  - between 40%-50% increase, beyond 5 years' timeframe
- Growth in terms of client base (number of clients)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - approximately 50% increase from 2 to 5 years' timeframe;
  - approximately 50% increase beyond 5 years' timeframe.
- Growth in terms of client base (number use cases)
  - between 25%-50% increase within the 1st year after obtaining the certification;
  - between 40%-50% increase from 2 to 5 years' timeframe;
  - between 40%-50% increase beyond 5 years' timeframe.

With regards to the indirect benefits expected under this policy option, increased competition benefits are expected to be met in both C2B and B2B value chains, especially in a long-term timeframe. In particular, the interviewed stakeholders expect:

- between 1%-25% increased competition in the B2B market from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the B2B market beyond 5 years' timeframe;
- between 1%-25% increased competition in the C2B market, from 2 to 5 years' timeframe;
- between 25%-50% increased competition in the C2B market, beyond 5 years' timeframe.

Data holders would still have the opportunity to monetise more from data sharing, while they will have easier access to a bigger network of data re-users. At the same time increased number of individuals will be attracted to share their personal data through the certified platforms. Data holders as well as data re-users would also have cost and time savings through digitization of interaction with data re-users by certified intermediaries, while mechanisms for assessing the quality of services provided by data intermediaries will be created for them through certification.

Several interviewed stakeholders, mainly in the C2B market, expressed the view that, even though the European data market is not mature at this stage for the establishment of a compulsory certification framework, in a long-term timeframe this policy option might become the preferred one. This is also based on the idea that compulsory certification will initially decrease the number of suitable data intermediaries, however it will significantly increase confidence in the market and fair market competition in the long-term, with same rules applying to everyone.

Table 46 – Overview of benefits – Establishing a certification framework for data intermediaries | PO 3[177]

Type of action	Description	Amount	Stakeholders
Certification framework for data intermediaries	**Direct benefits**		
	Time savings and business development cycle acceleration; speed of adoption of the market	45%-50% business development time acceleration *expected after certification*	**Data Intermediaries**
	Company growth in terms of revenue and number of employees	• between 25%-35% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between to 40%-50% increase beyond 5 years' timeframe • 20% increase corresponds to 30M-50M EUR/year revenue increase for a growth-stage intermediary	**Data Intermediaries**
	Company growth in terms of client base (number of clients)	• between 25%-50% increase within the 1st year after obtaining the certification; • 50% increase from 2 to 5 years' timeframe; • 50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Company growth in terms of client base (number of use cases)	• between 25%-50% increase within the 1st year after obtaining the certification; • between 40%-50% increase from 2 to 5 years' timeframe; • between 40%-50% increase beyond 5 years' timeframe	**Data Intermediaries**
	Data monetisation via data sharing through certified platforms	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
	Cost and time savings, including through digitization of interaction with data (re)users by certified intermediaries and quicker access to data suppliers (in the case of certified intermediaries who provide data pooling services) as well as efficiency gains when sharing supply chain data through a data intermediary	*Not quantifiable due to lack of data*	**Data holders/ Data providers, Data (re)users**
	Growing network and additional revenue through easier access to data (re)users (in the case of certified intermediaries who provide data pooling services)	*Not quantifiable due to lack of data*	**Data holders/ Data providers, (Certified) data**

[177] The table presents the estimated amounts of benefits by the interviewed stakeholders

	and larger client base (especially in the case of intermediaries providing data pooling services).		**intermediaries, Data (re)users**
	Data users obtain personal data with legal clarity on usability (consented data). Creation of mechanism for intermediaries' clients to assess the quality of intermediaries' operations (i.e. in terms of compliance with legislation) Additional revenue and increased productivity generated through the increased volume of re-used data (unlocked and facilitated through the certified intermediaries)	*Not quantifiable due to lack of data*	**Data (re)users**
	Indirect benefits		
	Increased market competition – B2B Market	• between 1%-25% increased competition from 2 to 5 years' timeframe, • between 25%-50% increased competition beyond 5 years' timeframe;	**Data Intermediaries**
	Increased market competition – C2B market	• between 1%-25% increase from 2 to 5 years' timeframe, • between 25%-50% beyond 5 years' timeframe	**Data Intermediaries**
	Additional revenue potentially generated through the "re-use" of the data unlocked and facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data holders/ Data providers**
	Enhancing of trust between the main actors involved	*Not quantifiable due to lack of data*	**(Certified) Data Intermediaries, Data holders/ Data providers, Data (re)users**
	Effect on Gross Domestic Product (GDP)	*Not quantifiable due to lack of data*	
	Economic value of data (both personal and industrial) will be maximised and additional revenue potentially generated through the re-use of the data facilitated through the certified intermediaries	*Not quantifiable due to lack of data*	**Data (re)users, Data holders/ Data providers**

1.3.1.23.2 Findings of the Cost-Benefit Analysis for Policy Option 3

Policy Option 3 entails significantly high levels of benefits including business development time acceleration, client base and revenues increases, market competition, as a result of the increased trust between the stakeholders in the market, after certification. Increased recurrent and one-off costs are expected as well for data intermediaries in order to obtain and maintain the certification, however, the expected benefits significantly outweigh the expected costs. This option presents a great cost-benefit relation as well, as the benefits exceed more than 10 times the costs. Details on the costs and benefits for policy option 2 are presented in Annex I.

1.3.1.24 Coherence of the option

This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.

2.3.4.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 47 - Summary of impacts for Establishing a certification framework for data intermediaries

Economic impacts	• Data intermediaries would be affected from the certification cost. However, certification would enable them to grow in terms of revenue, employees and client base. Competition and innovation in the data intermediaries market might also be affected. • Data holders and re-users will mainly benefit from efficiency gains including cost and time savings through digitisation of the transactions and facilitation of data sharing via the certified intermediaries.
Social impacts	• Societal benefit will be twofold. On the one hand society will benefit as the potential of the European data market will be unlocked through certification. On the other hand, data flows of intermediaries serving societal purposes (i.e. health, research) will be increased
Environmental impacts	• N/A
Fundamental rights impacts	• Protection of privacy and personal data will be promoted through the certification of data intermediaries, especially of personal data spaces.

2.4 Comparison of the policy options

The aim of this section is to compare of the policy options in order to identify the preferred policy option for each of the domains.

The following MCA has been performed in line with the European Commission's *Better Regulation Guidelines*[178] and its toolbox[179], most importantly tool 63[180]. The assessment builds on the prior analysis of each individual option.

It has been concluded in the previous section that for none of the area under investigation the baseline will be able to achieve the desired results and resolved identify problems. The assessment concludes that a policy intervention is needed. It remains to be seen the type (regulatory vs. non regulatory) and the intensity (low vs. high) of intervention. The MCA will assess which of the three policy options under each area is the most adequate:

- Measures facilitating secondary use of sensitive data held by the public sector
  - PO 1: Guidelines
  - PO 2: One-Stop-Shop
  - PO 3: Single Data authorisation Body
- Establishing a certification/authorisation scheme for data altruism mechanisms

[178] http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm
[179] http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm
[180] https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-63_en_0.pdf

    - PO 1: Coordination at EU level
    - PO 2: Voluntary certification scheme
    - PO 3: Mandatory authorisation
- Establishing a European structure for certain governance aspects of data sharing
    - PO 1: Informal expert group
    - PO 2: Formal expert group
    - PO 3: Legal body
- Establishing a certification framework for data intermediaries
    - PO 1: Industry driven certification framework
    - PO 2: Voluntary certification framework
    - PO 3: Compulsory certification framework

The MCA was carried out in the following three distinct steps:

- *Step 1:* Establish indicators or assessment criteria against which the policy options are assessed and compared. This includes establishing the performance of a policy option (i.e. the magnitude of its impact), the weight of the criteria in relation to each other, as well as the direction of the impact (negative/positive). The indicators are established in an analytical grid;
- *Step 2:* Build an outranking matrix in which the scores for all policy options and criteria are provided in order to summarise how the policy options compare with each other in relation to established criteria; and
- *Step 3:* Prepare a permutation matrix that enables the selection of a final ranking of all the possible policy options against each other for each domain. This means that it is possible not only to select a preferred policy option but also a ranking of all other options against each other.

2.4.1 Assessment criteria and indicators

The following assessment criteria were agreed with the European Commission for the assessment of the impacts of the options. A weight has been defined for each criterion. The direction of the change desired are all positive. The proportionality assessment criteria is considered as an exclusion criteria, and is therefore not included in the MCA.

Table 48 – Weight, direction and performance value allocated to the assessment criteria

Assessment criterion	Weight	Direction	Performance value
Effectiveness	0.3	1	Qualitative scale +/-4
Efficiency	0.3	1	Benefit/Cost-ratio (BCR)
Coherence	0.25	1	Qualitative scale +/-2
Legal and political feasibility	0.15	1	Qualitative scale +/-2
Proportionality	This exclusion criteria will not be assessed as part of the MCA	N/A	N/A

Based on the results of the Cost-Benefit analysis and the qualitative assessment of each individual options, we have drafted an **input grid** for each domain in which the scores for all policy options are collected and compared in relation to each criterion towards each other.

Table 49 – Input Matrix

Input matrix			Measures facilitating secondary use of sensitive data held by the public sector					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	3	0,9
Efficiency (BCR)	0,3	1	0	0	6	1,8	2,4	0,72
Coherence of the policy options	0,25	1	0,5	0,125	1	0,25	-1	-0,25
Legal and Political feasibility	0,15	1	1	0,15	0,5	0,075	-1	-0,15
			Establishing a certification/authorisation scheme for data altruism mechanisms					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1,5	0,45	2	0,6	2	0,6
Efficiency	0,3	1	0	0	2,7	0,81	6,3	1,89
Coherence of the policy options	0,25	1	1	0,25	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	1	0,15
			Establishing a European structure for governance aspects of data sharing					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	2	0,6
Efficiency	0,3	1	200.362,19	60108,657	19627,32	5888,196	1766,46	529,938
Coherence of the policy options	0,25	1	-1	-0,25	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	-1	-0,15
			Establishing a certification framework for data intermediaries					
			PO 1 - Non regulatory		PO 2 - Low intensity		PO 3 - High intensity	
Critera	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness of the policy options in reaching the specific and general policy objectives	0,3	1	1	0,3	2	0,6	1	0,3
Efficiency	0,3	1	5,21	1,563	2,68	0,804	2,68	0,804
Coherence of the policy options	0,25	1	0,5	0,125	1	0,25	1	0,25
Legal and Political feasibility	0,15	1	1	0,15	1	0,15	1	0,15

2.4.1.1 Measures facilitating secondary use of sensitive data held by the public sector

The analysis above shows that of all policy options, the single data authorisation body (Policy Option 3) is the option likely to achieve best the policy objectives, as it is also the most ambitious. This policy option combines the establishment of a one-stop shop and of secure data processing environments (Policy Option 2) with additional tasks designed to facilitate secondary re-use of data the use of which is subject to the rights of others. As a consequence, it brings more benefits, but also additional costs.

However, the ambition of policy option 3 to enable the re-use of data the use of which is subject to the rights of others for commercial purposes is incompatible with the national legislation of some Member States. As noted above, a single secure data processing environment combining data the use of which is subject to the rights of others from different holders may not be permitted by some national laws. The introduction of a single data authorisation body can also damage public trust. In one Member States, such establishment has been stranded because of trust-related considerations.

Policy option 2 – establishing a one-stop shop and data processing environments is both feasible and proportionate: the tasks linked with data discovery and with advisory services performed by the one-stop shops appear feasible and proportionate in relation to the policy objectives (although the provision of advice specific to datasets should remain the responsibility of data holders), while limiting sensitive data re-use to research and to commercial purposes serving a public interest also appears proportionate (and appears to correspond to the legal status quo in a large number of Member States). Likewise, allowing Member States to design more than one data processing environment as needed, appears proportionate in view of the policy objectives, and is likely to be feasible/compatible with existing national laws. As pointed out by participants to the workshop held on 8 July 2020, the very different nature of datasets in scope (i.e. in health, and other sectors) militate for several data processing environments coexisting. In addition, policy-option 2 is the most efficient, with a Benefit-Cost ratio higher than policy option 3.

When asked their view, **participants to the 8 July workshop expressed a strong preference for PO2** over the other policy options, and a slight preference for PO3 over PO1. The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 50 – Summary comparison table between low/high intensity policy options for Measures facilitating secondary use of sensitive data held by the public sector

	Regulatory intervention with low intensity (PO2)	**Regulatory intervention with high intensity (PO3)**
Efficiency	This option would bring costs likely ranging in millions of euros for each Member State. However, it would also result in very significant savings that are likely to significantly outweigh the costs.	This option would create higher costs than PO2 (linked to processing data access applications), but would also yield higher benefits for both data holders and data reusers. Opening the re-use of sensitive data to commercial purposes would likely produce innovation and growth, but comes with risks linked to trust – risks which could undermine these benefits. Overall, the Benefit-Cost ratio of PO3 is lower than that of PO2.
Effectiveness	Contributes to achieving the specific and general objectives.	Contributes more than PO2 to achieving the specific and general objectives, as it also includes centralisation of data access applications.
Coherence	Coherent with EU law and with national laws.[181]	Coherent with EU law. Enabling re-use of data for strictly commercial purposes would be incompatible with national laws of several Member States. A single data processing environment would be incompatible with national law of some Member States.[182]
Legal/political feasibility	This option appears to be feasible.	There are doubts as to the political feasibility of this option in some countries (e.g. where public trust is government is lower).
Proportionality	This option appears proportionate.	A single data processing environment would be disproportionate, particularly in view of the very different types of datasets concerned.

As regards guidelines/recommendations (policy option 1), it remains unclear whether these would result in a larger number of Member States setting up structures to enhance the reuse of data the use of which is subject to the rights of others compared to a no intervention scenario (i.e. whether many Member States would do so *as a result of* these recommendations).

2.4.1.2 Establishing a certification/authorisation scheme for data altruism mechanisms

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

[181] The national laws of stakeholders interviewed.
[182] The national laws of stakeholders interviewed.

Table 51 – Summary comparison table between low/high intensity policy options for Establishing a certification scheme for data altruism mechanisms

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	This option would bring substantial costs to Member States, however also increasing benefits.	This option would bring substantial costs to Member States, however also increasing benefits.
Effectiveness	This option would increase transparency for and decrease security concerns of citizens.	This option would increase transparency for and decrease security concerns for citizens.
Coherence	Coherent with EU law as there is currently no EU law on data altruism.	Coherent with EU law as there is currently no EU law on data altruism.
Legal/political feasibility	This option appears to be feasible.	This option appears to be feasible.
Proportionality	This option appears to be proportionate.	This option appears to be proportionate.

The preliminary analysis of the different policy options shows there is great potential for data altruism mechanisms in the European Union considering that the COVID-19 crisis catapulted data altruism in the limelight, it also uncovered the national differences amongst Member States on core discussion points such as data privacy and the discussion of costs and benefits of such mechanisms. This was highlighted by the different approaches Member States took to implement data altruism mechanisms, if at all.

Of the four policy options, the baseline scenario is the least effective and efficient to achieve the general objectives. It would likely exclude SMEs and organisations from data altruism because only large cooperation's would have the necessary resources to handle legal fragmentation across member states and a European data economy including data altruism would be difficult to achieve. This would be an economic and societal loss.

Policy option 1 would be a positive step towards achieving the specific and general objectives, however considering the fast-paced developments in this field, it could also have a very limited effectiveness since cooperation or expert groups and voluntary soft regulation can be lengthy process.

Policy option 2 and 3 appear to be the most effective and efficient options to achieve the general and specific outcomes. Whereas option 2 and 3 would both cost member states, option 3 is less favourable because the costs would be significantly higher and could create additional regulatory burden for member states. Nevertheless, in the light of the COVID-19 crisis, the benefits of data altruism schemes have been highlighted in the public and political debate and the Commission could use this momentum to highlight the expected benefits and how this could outweigh costs.

2.4.1.3 Establishing a European structure for governance aspects of data sharing

The analysis of the different options allows to draw a preliminary comparison of the different options, based on the current available evidence.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 52 – Summary comparison table between low/high intensity policy options for Establishing a European structure for governance aspects of data sharing

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	The costs of a formal expert group are limited, while the benefit are similar to PO3: an expert group will have a limited positive effect over the development and adoption of standardisation initiatives	The cost of a legal body are typically higher than a formal expert group (5 to 10 times higher), while the benefits are similarly positive.
Effectiveness	The overall effect on data sharing are moderately positive – there is a very mediated relation between an expert group and the adoption of data sharing standards	Equivalent to PO2. It is unlikely that a legal body status would visibly make a difference when it comes to business adoption of the standards
Coherence	Strong coherence with digital strategy, notably with the creation of data spaces.	Equivalent to PO2
Legal/political feasibility	High. Positive reception by industry interviewees, who are reluctant to strong EU role in standardisation but calls for soft support and stakeholders engagement	Moderate. Industry as well as standardisation initiatives are suspicious towards government led efforts in standardisation. Also, the creation of yet another legal entity would have to be strongly justified internally.
Proportionality	High. The intervention is very limited and proportionate to the stakes at hand	High. The intervention is stronger and proportionate to the stakes at hand

First and foremost, the specific and general objectives are very demanding and ambitious. They concern the internal decision of how companies behave when it comes to data standardization and sharing. In addition, the options have to take into account the multiplicity of actors already carrying out activities in this area.

As such, it can be expected that no single option can deliver the specific and general objective, as their achievement is based on how companies react to the policy option. It will take time and a multiplicity of intervention to achieve them.

Option 0 appears as the worse option, not just because it will be less efficient and effective, but it has possible negative effects because of the fast-evolving nature of the economic context. Many other countries worldwide are active in data standardization, and there is the risk that Europe follows the lead of others, with all the negative consequences in terms of competitiveness. In other words, there is the risk that after personal data, also industrial data become controlled by non-European players.

Furthermore, option 0 is inconsistent with the initiatives in place on data spaces. The absence of activity on data standardization will weaken the effectiveness of the data spaces.

Option 1, 2 and 3 have limited differences – basically around an informal, formal expert group or a legal body. An informal expert group is unlikely to have an effect at all, while the legal body appears overly heavy from a bureaucratic point of view, considering the wealth of existing bodies working on standardization. In other words, option 1 has limited costs, but also very little impact. Option 3 on the other hand seem to achieve a comparable impact to option 2 in terms of standardisation, but with increased costs related to the creation of a new legal entity or the expansion of an existing one and lower political feasibility.

2.4.1.4 Establishing a certification framework for data intermediaries

This section provides intermediary conclusions with regard to the policy options for this domain. Whereas this measurement unit cannot be added, a qualitative summary assessment is provided in the Average row of the table.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 53 – Summary comparison table between low/high intensity policy options for Establishing a certification framework for data intermediaries

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	PO 2 presents the best balance between costs and benefits for the stakeholders and appears to be the preferred policy option by the interviewed stakeholders.	PO 3 presents a broad number of benefits for the stakeholders, at similar levels to the benefits of PO 2. However, it is doubtful whether the European market is mature enough for the establishment of a compulsory certification framework. There are concerns that the latest might impose unnecessary burdens to data intermediaries, leading in the end to opposite results from the desired ones.
Effectiveness	PO 2 could significantly contribute to the three specific objectives, particularly, in creating trust in common European data spaces, building common data spaces as well as ensuring data interoperability across sectors, through the certification framework. The majority of stakeholders interviewed agreed to this policy option as the most effective one given that the certification criteria would be defined by a legal instrument. Therefore, if a big number of industry players decides to proceed to the certification process, the trust between the stakeholders involved in the market would be increased significantly, allowing the data intermediary market to flourish and bringing various economic and societal benefits. Following the effectiveness in achieving the specific objectives described in the above section, this policy option would also further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy, since the European data market overall will be significantly boosted through certification, increasing the volume of data flows.	PO3 is expected to contribute to some of the specific objectives, namely creating trust in common European data spaces as well as contributing to data interoperability across sectors. However, concerns were raised regarding its effectiveness to build common data spaces, due to the fact that a compulsory certification process with hard neutrality requirements is likely to prevent small industry players from getting into the market due to the potentially prohibitive certification cost. In the same line of argumentation, there are doubts whether the overall impact of a compulsory certification framework would be positive by boosting the market, since it could create significant burdens for new players to get into the market. It is therefore under question whether this policy option could further contribute to set the foundations of a Single Market for Data, and furthermore, strengthen the EU data economy.
Coherence	This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.	This policy option could promote the objectives of the single market for data under the European Data Strategy and therefore remains coherent with the current EU legal and policy framework.
Legal/political feasibility	This policy option appears to be legally and politically feasible to be adopted and implemented at European level,	This policy option appears to be legally and politically feasible to be adopted and implemented at European level,

Proportionality	This policy option is proportionate as its intensity matches the identified problem and objectives of this study	This policy option is not proportionate as its intensity is deemed too strong for the identified problem and objectives of this study.

The majority of stakeholders involved expect benefits and costs in all three policy options while they are not in favour of policy option 0 meaning no action is taken at EU level.

The preferred policy option seems to be Policy Option 2 establishing a voluntary certification framework through a legal instrument by the European Commissions. Even though the level of costs and benefits remain common and similar for all the three policy options, stakeholders expressed concerns that policy option 1 might be too weak compared to the other two, in which case it would place it close to the effectiveness levels of policy option 0. On the other side, according to the stakeholders, it is doubtful whether the European market is mature enough for the establishment of a compulsory certification framework. There are concerns that the latest might impose unnecessary burdens to data intermediaries, leading in the end to opposite results from the desired ones.

2.4.2 Comparison of the policy options

In relation to Step 2, the following table provides an **outranking matrix** in which all the weights indicated in the table under step 1 are totalled for the criteria in relation to which a policy option is favoured over another policy option (abbreviated e.g. as "PO1/PO2") as indicated by the weighted performance of each criterion.

This means that the outranking matrix provides an overview of the overall scores of the policy options compared to each other (i.e. the differences between them).

Table 54 – Outranking matrix

Outranking matrix			
Measures facilitating secondary use of sensitive data held by the public sector	PO 1	PO 2	PO 3
PO 1	**0**	**0,15**	**0,4**
PO 2	**0,85**	**0**	**0,7**
PO 3	**0,6**	**0,3**	**0**
Establishing a certification/authorisation scheme for data altruism mechanisms	PO 1	PO 2	PO 3
PO 1	**0**	**0**	**0**
PO 2	**0,6**	**0**	**0**
PO 3	**0,6**	**0,3**	**0**
Establishing a European structure for governance aspects of data sharing	PO 1	PO 2	PO 3
PO 1	**0**	**0,3**	**0,45**
PO 2	**0,55**	**0**	**0,45**
PO 3	**0,55**	**0**	**0**
Establishing a certification framework for data intermediaries	PO 1	PO 2	PO 3
PO 1	**0**	**0,3**	**0,3**
PO 2	**0,55**	**0**	**0,55**
PO 3	**0,25**	**0**	**0**

Naturally, the grey combinations received a score of 0 as it does not make sense to compare these. In essence, the table shows that the impacts of the policy options outrank those of the baseline scenario and that policy options with a higher score outrank those with a lower score.

The differences between the overall rankings of each policy option between each other as presented above are derived from the sum of the individual scores per policy option and assessment criterion in the analytical grid.

The table below present the six different combination of policy options for the four areas under investigation.

Table 55 – Policy ranking permutation

Policy ranking permutation	Policy parings	Coefficients of policy pairings	Final score
Measures facilitating secondary use of sensitive data held by the public sector			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,15 + 0,4 + 0,7	1,25
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,4 + 0,3 + 0,15	0,85
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,85 + 0,4+ 0,7*	*1,95*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,7 + 0,6 + 0,3	1,6
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,6 + 0,15 + 0,3	1,05
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,6 + 0,85	1,75
Establishing a certification scheme for data altruism mechanisms			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0 + 0 + 0	0
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0 + 0,3 + 0	0,3
PO2/PO1/PO3	PO2/PO1 + PO1/PO3 + PO2/PO3	*0,6 + 0 + 0*	*0,6*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0 + 0,6 + 0,3	0,9
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,6 + 0 + 0,3	0,9
PO3/PO2/PO1	*PO3/PO2 + PO3/PO1 + PO2/PO1*	*0,3 + 0,6 + 0,6*	*1,5*
Establishing a European structure for governance aspects of data sharing			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,3 + 0,45 + 0,45	1,2
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,45 + 0 + 0,3	0,75
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,55 + 0,45 + 0,45*	*1,45*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,45 + 0,55 + 0	1
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,55 + 0,3 + 0	0,85
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0 + 0,55 + 0,55	1,1
Establishing a certification framework for data intermediaries			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,3 + 0,3 + 0,55	1,15
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,3 + 0 + 0,3	0,6
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,55 + 0,3 + 0,55*	*1,4*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,55 + 0,25 + 0	0,8
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,25 + 0,3 + 0	0,55
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0 + 0,25 + 0,55	0,8

This means the following:

- For Measures facilitating secondary use of sensitive data held by the public sector, **policy option PO2 – One-stop-shop** is the preferred option as it providers the most combination of effectiveness, efficiency and coherence.
- For Establishing a certification scheme for data altruism mechanisms, **policy option PO3 – mandatory authorization scheme** for data altruism is the preferred option;
- For Establishing a European structure for governance aspects of data sharing, **policy option PO2 – formal expert group** is the preferred option;
- For Establishing a certification framework for data intermediaries**, policy option PO2 – Voluntary certification framework** for intermediaries is the preferred option.

It must be noted that although PO1s for the four areas, rank rather well in comparison to the other policy options, the actual amounts of costs and benefits expected by a non-regulatory intervention are extremely limited. Additionally, it is assumed that the schemes foreseen under the second and fourth domains would rely on either the one-stop shop mechanisms set up by Member States (PO2) or a private conformity assessment body, accredited by the European Data Innovation Board (PO3) to issue the respective certifications and authorizations to data intermediaries.

2.5 Assessment of macro-economic impacts

This section consists in comparing the expected macro-economic impacts of a low intensity, high intensity regulatory intervention and the preferred association of policy options on the overall economy and society compared to the baseline scenario.

2.5.1 Definition of policy packages

The assessment of impacts on the overall data economy and society can only be performed at an aggregated level, by creating policy packages composed of one policy option per domain (area). Based on the multi-criteria analysis performed, the fact that the PO1 (non-regulatory options) were deemed to create low impacts in terms of costs and benefits, and taking into account the interdependences between the policy options (reliance on PO2 or PO3 under Measures facilitating secondary use of sensitive data held by the public sector, to support the certification mechanisms linked to data altruism and data intermediaries), four policy option packages where identified:

- **Policy Package 0 – Baseline**: the baseline scenario consists in applying no policy changes to the four areas for which problems could be identified: sensitive data held by the public sector, data altruism schemes, certification of data intermediaries, governance and standards. The EU economy will not be able to reap the benefits of data sharing.
- **Policy Package 1** – Low intensity regulatory options: this package includes the creation of a one-stop shop to foster the sharing of (sensitive) data whose use is subject to the rights of others held by the public sector. A voluntary certification scheme would be established by EU Member States for data altruism mechanism and organisations offering such schemes. Data intermediaries will also be able to obtain a certification to demonstrate their neutrality and absence of conflict of interest (e.g. absence of competition with data users) on a voluntary basis. Finally, the European Data innovation Board would take the form of a formal expert group created by the European Commission, including Member States representatives and industry representatives.
- **Policy Package 2** – High intensity regulatory options: Under this package, Member States will be required to set up a Single Data Authorisation body in charge of providing the authorisation to enable the further use of data that is subject to the rights of others contained held by the public sector. This entity will also be in charge of delivering the compulsory authorisation required from organisations offering data altruism schemes, as well as mandatory certification scheme for data intermediaries. Under this package, the European Data Innovation Board would consist of an independent European body with legal personality, supported by a secretariat.
- **Policy Package 3** – Preferred policy options: this package is similar to Policy Package 1, with the exception that a compulsory authorisation mechanism is set-up for organisations offering data altruism schemes.

Table 56 – Policy Packages composition

		Measures facilitating secondary use of sensitive data held by the public sector	Establishing a certification scheme for data altruism mechanisms	Establishing a European structure for governance aspects of data sharing	Establishing a certification framework for data intermediaries
Policy Package 1	Low regulatory options	PO2	PO2	PO2	PO2
Policy Package 2	High regulatory options	PO3	PO3	PO3	PO3
Policy Package 3	Preferred options	PO2	PO3	PO2	PO2

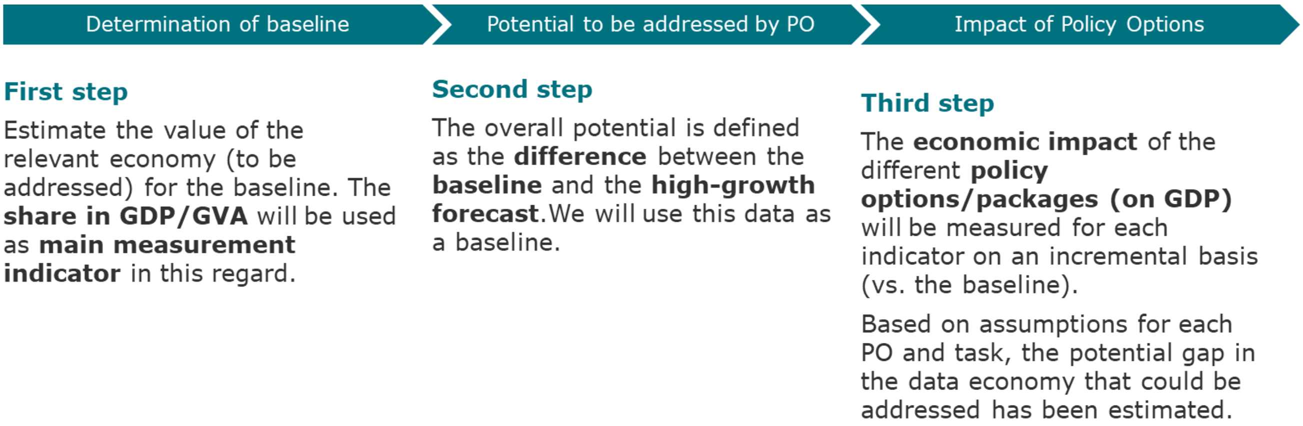
2.5.2 Methodological approach

This section provides a brief explanation about the methodological approach for the macroeconomic analysis.

For the analysis of the economic impact both a top-down and a bottom-up analysis is conducted. The top-down approach is the primary method of analysis, whereas the bottom-up approach serves to validate the results. However, both approaches can be used to estimate a range of results and to calculate averages.

The top-down approach will be based on the estimation of a broader baseline for the size of the relevant data economy in a first step. In a second step, the potential is calculated by estimating a high growth scenario and calculating the difference to the baseline. The core of the top-down analysis is the estimation of the (positive) impact of each policy option per domain in terms of contributing to reach the high-growth scenario. In this regard, ratios are estimated. The top-down approach is summarized in the figure below and described in more detail in section 4.1.4.

Figure 3 – Approach to the top-down analysis



Determination of baseline

First step

Estimate the value of the relevant economy (to be addressed) for the baseline. The **share in GDP/GVA** will be used as **main measurement indicator** in this regard.

Potential to be addressed by PO

Second step

The overall potential is defined as the **difference** between the **baseline** and the **high-growth forecast**.We will use this data as a baseline.

Impact of Policy Options

Third step

The **economic impact** of the different **policy options/packages (on GDP)** will be measured for each indicator on an incremental basis (vs. the baseline).

Based on assumptions for each PO and task, the potential gap in the data economy that could be addressed has been estimated.

The bottom-up approach, on the other hand, is based on the micro-analysis of estimated impacts conducted for each domain. Within the CBA, certain benefits and costs are assessed. As far as possible, the impact on GDP is estimated based on the CBA results and/or case studies. The results and estimations of the micro-analyses are extrapolated and scaled in this regard. The bottom-up approach is described in more detail in Annex I.

2.5.2.1 Calculation of the baseline

The baseline has been calculated based on the forecast of the European Data Market Monitoring Tool.[183] The monitoring tool provides three forecasts at the 2025 Horizon: a baseline scenario, a challenge scenario and a high growth scenario.

Our study aims at understanding to what extent the introduction of a regulatory intervention will aim at reaching the expected levels of the high growth forecast from the European Data Market Monitoring Tool.[184]

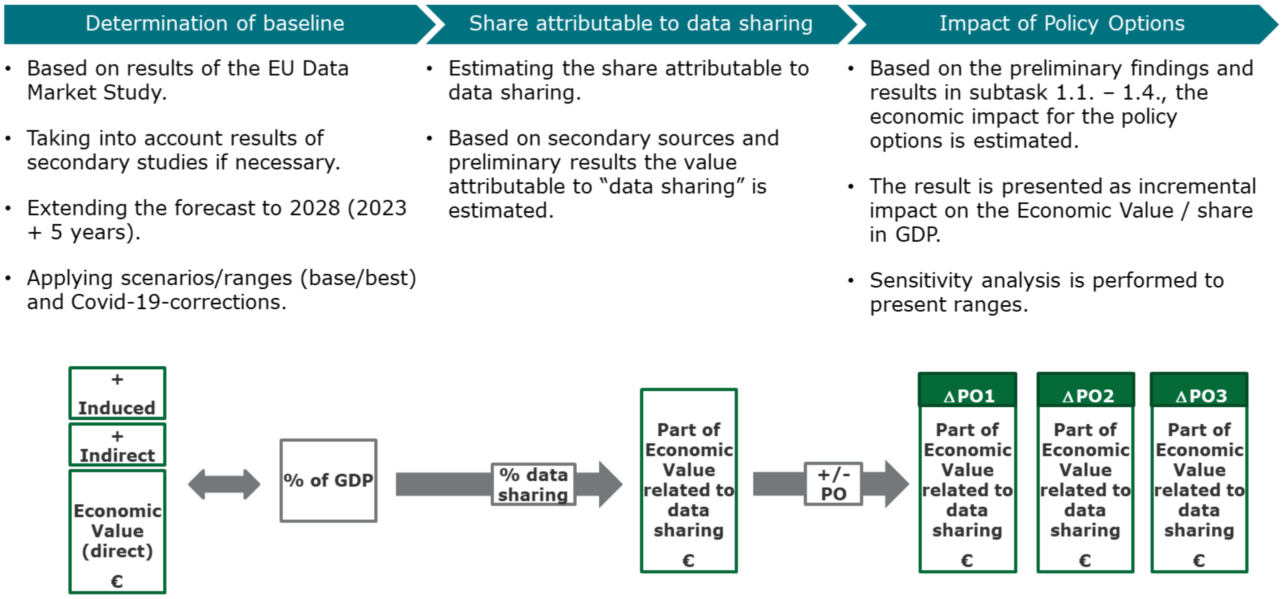
In the year 2020, the outbreak of Covid-19 massively affected the European economy. Expected figures have been corrected to take into account the impact of this crisis. The European Data Market Monitoring Tool already provides a Covid19 correction of the 2025 forecast. We have made further annual adjustments according to The Economist intelligence Unit data forecast of GDP, which includes Covid19 corrections for 2020.

The impacts are calculated until 2025 on the basis of the value of the data economy as projected by the EU Data Monitoring Tool, which is the basis (or baseline) of our analysis. The EU Data Monitoring Tool forecast projects a growth of the data economy of approx. 8% p.a.. This forecast for the growth of the EU data economy, however, ends in 2025. In order to calculate impacts beyond 2025 we have taken a conservative approach and calculated the impacts on the basis of the GDP growth rate forecast of the OECD (1.5%-1.6% p.a.). For this reason the impacts are based on a much lower per annum growth rate for the period 2026-2028.

2.5.2.2 Top-Down analysis

In order to obtain the economic impact of data sharing, in relation to its contribution to GDP, a top-down analysis has been performed.

Figure 4 – Approach to the top-down analysis



Determination of baseline

- Based on results of the EU Data Market Study.
- Taking into account results of secondary studies if necessary.
- Extending the forecast to 2028 (2023 + 5 years).
- Applying scenarios/ranges (base/best) and Covid-19-corrections.

Share attributable to data sharing

- Estimating the share attributable to data sharing.
- Based on secondary sources and preliminary results the value attributable to "data sharing" is estimated.

Impact of Policy Options

- Based on the preliminary findings and results in subtask 1.1. – 1.4., the economic impact for the policy options is estimated.
- The result is presented as incremental impact on the Economic Value / share in GDP.
- Sensitivity analysis is performed to present ranges.

+ Induced
+ Indirect
Economic Value (direct) €
% of GDP
% data sharing
Part of Economic Value related to data sharing €
+/- PO
ΔPO1
Part of Economic Value related to data sharing €
ΔPO2
Part of Economic Value related to data sharing €
ΔPO3
Part of Economic Value related to data sharing €

[183] Data landscape, The European Data Market Monitoring Tool see: http://datalandscape.eu/european-data-market-monitoring-tool-2018

[184] According to the European Data Market Study, "The High Growth scenario is characterised by a high level of data innovation, low data power concentration, an open and transparent data governance model with high data sharing, and a wide distribution of the benefits of data innovation in the society".

The data from the European Data Monitoring Tool provides a baseline for the economic value of the data economy and relates it to GDP. We have used this data to calculate the baseline. Adjustments with regard to Covid-19 outbreak macroeconomic impact have been included. The overall potential is defined as the difference between the baseline and the high-growth forecast.

In order to define the relevant part of the value of the data economy that the policy intervention foreseen as part of this study can address, two conservative assumptions were made:

- The full potential of data sharing represent 80% of the total data economy;[185]
- 50% of barriers to data sharing are linked to interoperability and trust related issues, which are the specific problems the policy options under scrutiny in this study aim to address:
- The open public consultation on the European Strategy for Data revealed that almost 80% of the participants of the public consultation have encountered difficulties in using data from other companies. These obstacles mainly relate to technical aspects (data interoperability and transfer mechanisms), denied data access, and prohibitive prices or other conditions considered unfair or prohibitive;[186]
- A study from the World Economic Forum also estimates that the trust and technical issues are the most pregnant barriers to data sharing.[187]
- At the same time, there is a general consensus (91% of the participants of the open public consultation) that standardisation is necessary to improve interoperability and ultimately data re-use across sectors.[188]

Based on these assumptions, the potential gap in the data economy that could be addressed if these problems were resolved has been estimated.

For each area, further assumptions have been made to understand in more details the magnitude of impact of the specific policy options on this potential gap. The experts' assumptions are based on the findings of the research, interviews carried out and the literature studies.[189] The baseline and the potential gap to be addressed are presented in the table below. The growth rates of the reference period 2026-2028 are based on the OECD long-term GDP forecast highlighted in dark green.

[185] Deloitte study for Vodafone group, Realising the economic potential of machine-generated, nonpersonal data in the EU, see: https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

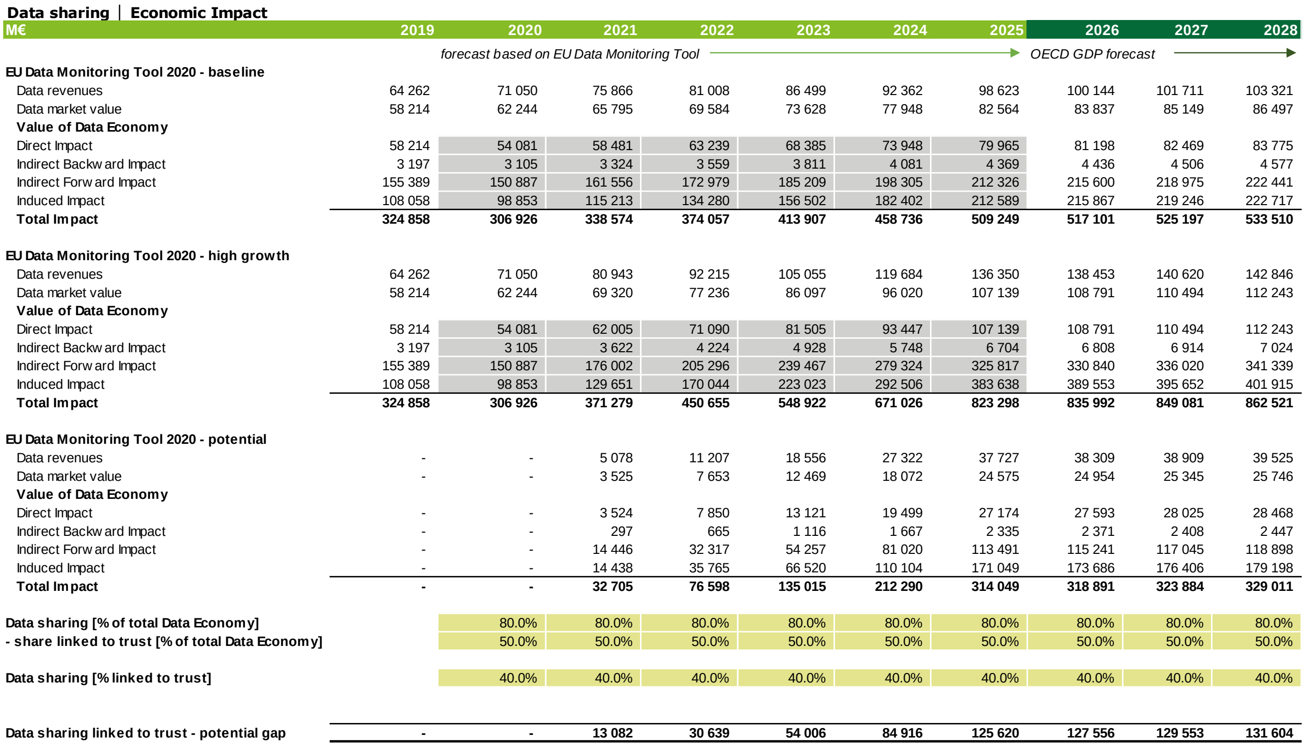
[186] European Commission, 2020, Open public Consultation on the European Strategy for Data. Summary Report on the open public consultation on the European Strategy for Data.

[187] WEF, Share to Gain: Unlocking Data Value in Manufacturing, see : https://www.weforum.org/whitepapers/share-to-gain-unlocking-data-value-in-manufacturing

[188] European Commission, 2020, Open public consultation on the European Strategy for Data.

[189] Accordingly, the assumptions to a certain extent reflect the results of the bottom-up approach.

Figure 5 - Baseline estimates



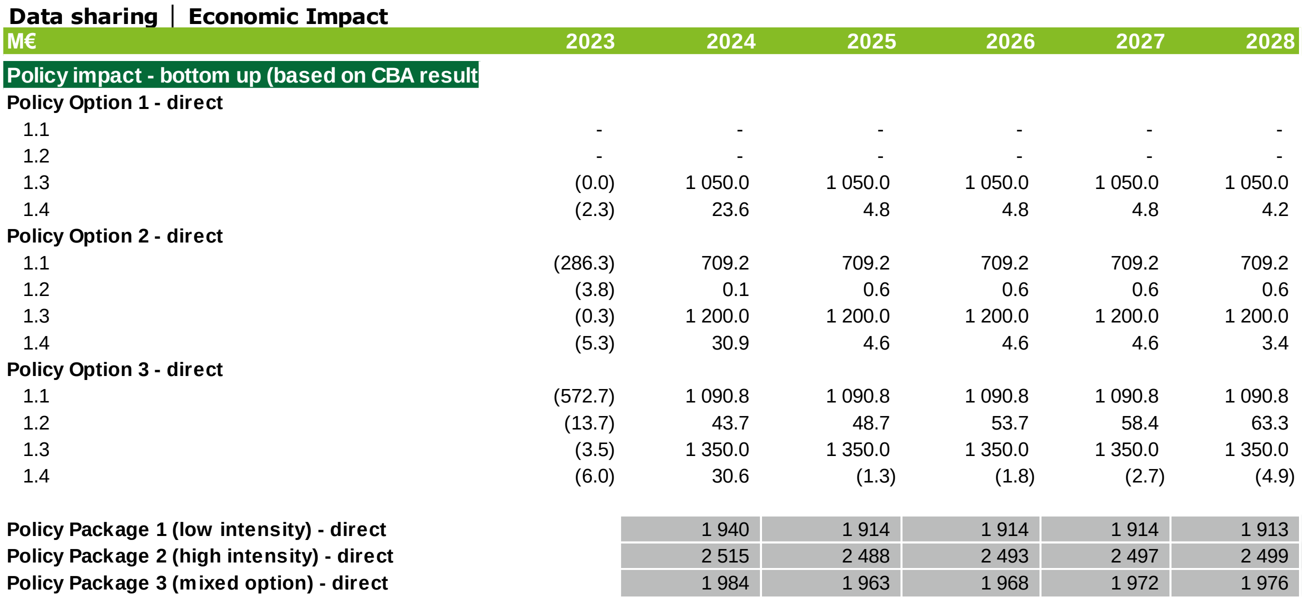
Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
			forecast based on EU Data Monitoring Tool →					*OECD GDP forecast* →		
EU Data Monitoring Tool 2020 - baseline										
Data revenues	64 262	71 050	75 866	81 008	86 499	92 362	98 623	100 144	101 711	103 321
Data market value	58 214	62 244	65 795	69 584	73 628	77 948	82 564	83 837	85 149	86 497
Value of Data Economy										
Direct Impact	58 214	54 081	58 481	63 239	68 385	73 948	79 965	81 198	82 469	83 775
Indirect Backward Impact	3 197	3 105	3 324	3 559	3 811	4 081	4 369	4 436	4 506	4 577
Indirect Forward Impact	155 389	150 887	161 556	172 979	185 209	198 305	212 326	215 600	218 975	222 441
Induced Impact	108 058	98 853	115 213	134 280	156 502	182 402	212 589	215 867	219 246	222 717
Total Impact	**324 858**	**306 926**	**338 574**	**374 057**	**413 907**	**458 736**	**509 249**	**517 101**	**525 197**	**533 510**
EU Data Monitoring Tool 2020 - high growth										
Data revenues	64 262	71 050	80 943	92 215	105 055	119 684	136 350	138 453	140 620	142 846
Data market value	58 214	62 244	69 320	77 236	86 097	96 020	107 139	108 791	110 494	112 243
Value of Data Economy										
Direct Impact	58 214	54 081	62 005	71 090	81 505	93 447	107 139	108 791	110 494	112 243
Indirect Backward Impact	3 197	3 105	3 622	4 224	4 928	5 748	6 704	6 808	6 914	7 024
Indirect Forward Impact	155 389	150 887	176 002	205 296	239 467	279 324	325 817	330 840	336 020	341 339
Induced Impact	108 058	98 853	129 651	170 044	223 023	292 506	383 638	389 553	395 652	401 915
Total Impact	**324 858**	**306 926**	**371 279**	**450 655**	**548 922**	**671 026**	**823 298**	**835 992**	**849 081**	**862 521**
EU Data Monitoring Tool 2020 - potential										
Data revenues	-	-	5 078	11 207	18 556	27 322	37 727	38 309	38 909	39 525
Data market value	-	-	3 525	7 653	12 469	18 072	24 575	24 954	25 345	25 746
Value of Data Economy										
Direct Impact	-	-	3 524	7 850	13 121	19 499	27 174	27 593	28 025	28 468
Indirect Backward Impact	-	-	297	665	1 116	1 667	2 335	2 371	2 408	2 447
Indirect Forward Impact	-	-	14 446	32 317	54 257	81 020	113 491	115 241	117 045	118 898
Induced Impact	-	-	14 438	35 765	66 520	110 104	171 049	173 686	176 406	179 198
Total Impact	**-**	**-**	**32 705**	**76 598**	**135 015**	**212 290**	**314 049**	**318 891**	**323 884**	**329 011**
Data sharing [% of total Data Economy]		80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%
- share linked to trust [% of total Data Economy]		50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%	50.0%
Data sharing [% linked to trust]		40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%
Data sharing linked to trust - potential gap	**-**	**-**	**13 082**	**30 639**	**54 006**	**84 916**	**125 620**	**127 556**	**129 553**	**131 604**

2.5.2.3 Bottom-up analysis

A Bottom-up validation of these estimates has been performed based on the results of the Cost-Benefit analysis for each domain.

Figure 6 – Results of the CBA per policy options and Policy Packages

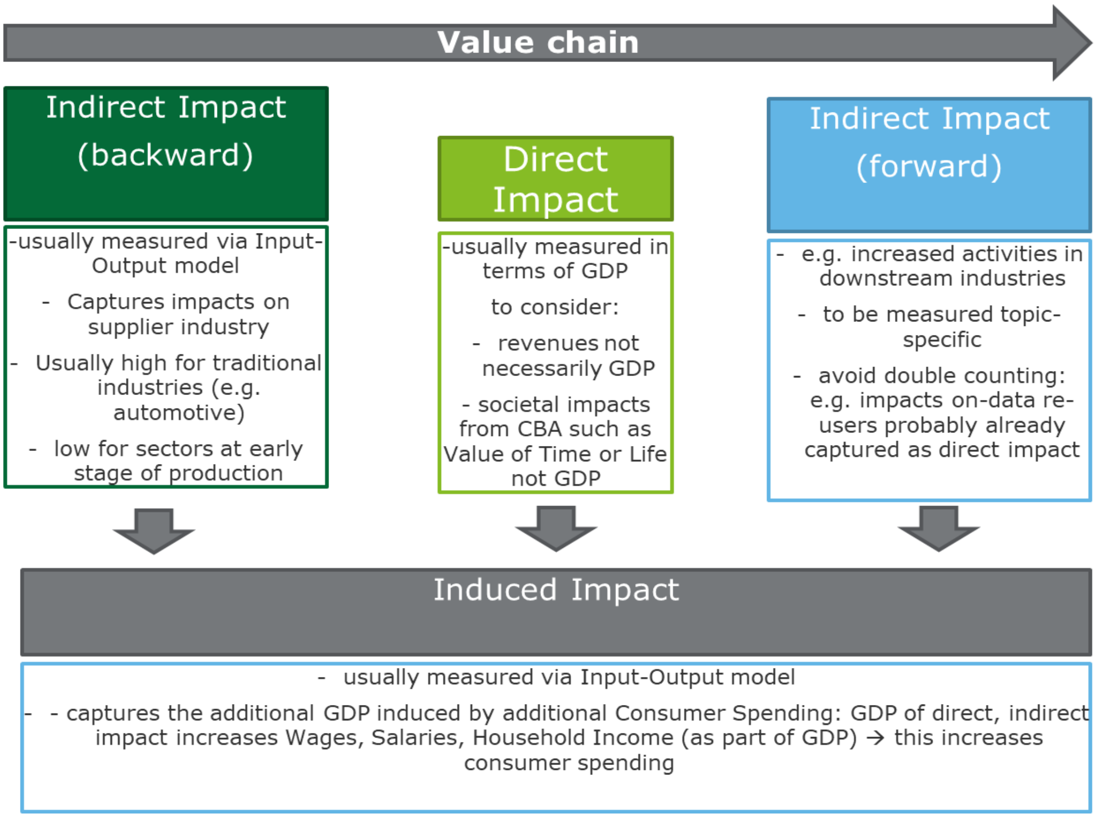


Data sharing | Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - bottom up (based on CBA result						
Policy Option 1 - direct						
1.1	-	-	-	-	-	-
1.2	-	-	-	-	-	-
1.3	(0.0)	1 050.0	1 050.0	1 050.0	1 050.0	1 050.0
1.4	(2.3)	23.6	4.8	4.8	4.8	4.2
Policy Option 2 - direct						
1.1	(286.3)	709.2	709.2	709.2	709.2	709.2
1.2	(3.8)	0.1	0.6	0.6	0.6	0.6
1.3	(0.3)	1 200.0	1 200.0	1 200.0	1 200.0	1 200.0
1.4	(5.3)	30.9	4.6	4.6	4.6	3.4
Policy Option 3 - direct						
1.1	(572.7)	1 090.8	1 090.8	1 090.8	1 090.8	1 090.8
1.2	(13.7)	43.7	48.7	53.7	58.4	63.3
1.3	(3.5)	1 350.0	1 350.0	1 350.0	1 350.0	1 350.0
1.4	(6.0)	30.6	(1.3)	(1.8)	(2.7)	(4.9)
Policy Package 1 (low intensity) - direct		1 940	1 914	1 914	1 914	1 913
Policy Package 2 (high intensity) - direct		2 515	2 488	2 493	2 497	2 499
Policy Package 3 (mixed option) - direct		1 984	1 963	1 968	1 972	1 976

In order to fully reflect on the reality of the impact, the indirect impacts have been added to the estimates based on the CBA results. A coefficient of 2.6 has been used, in line with the results of the European Data Monitoring Tool.[190]

Figure 7 – Overview of direct, indirect and induced impacts.


Value chain
Indirect Impact (backward)
-usually measured via Input-Output model
- Captures impacts on supplier industry
- Usually high for traditional industries (e.g. automotive)
- low for sectors at early stage of production
Direct Impact
-usually measured in terms of GDP
to consider:
- revenues not necessarily GDP
- societal impacts from CBA such as Value of Time or Life not GDP
Indirect Impact (forward)
- e.g. increased activities in downstream industries
- to be measured topic-specific
- avoid double counting: e.g. impacts on-data re-users probably already captured as direct impact
Induced Impact
- usually measured via Input-Output model
- - captures the additional GDP induced by additional Consumer Spending: GDP of direct, indirect impact increases Wages, Salaries, Household Income (as part of GDP) → this increases consumer spending

2.5.3 Macroeconomic impacts of the policy packages

A **Single Market for Data with common data spaces** has the potential to bring **immense benefits** to the economy, notably through increased innovation and opportunities, as well as to society and the environment through new insights informing and improving decision-making.

As noted in section 2.2.1, the **objective of this intervention** is to set **the foundations** of a Single Market for Data by **enabling a range of actors** to make data available for access and secondary use, by ensuring greater interoperability across sectors, and by fostering trust. Therefore, this intervention is a **necessary first step in the process of creating these common data spaces**. The full range of benefits incurred by the latter heavily rely on **other actors seizing the opportunities** offered by these building blocks to complete these data spaces. Without this intervention, however, these actors would have no incentives *ceteris paribus* to change their behaviour.

For instance, the low-intensity **regulatory intervention under Measures facilitating secondary use of sensitive data held by the public sector** – requiring Member States to set up one-stop

[190] The European Data Monitoring implicitly includes several types of multipliers, including indirect and induced impacts, which estimate impacts on the supplier industries and the overall economy generated through additional income and consumption (both could be classically estimated using e.g. Input-Output models), as well as indirect forward impacts, which estimate the effects downstream in the economy. To stay conservative, the later one have been considered here based on the European Data Monitoring Tool, since those impacts are expected to be of major interest. The European Data Monitoring Tool in this regard estimates coefficients between 2.6 in the baseline as a lower bound and 3.0 in the high growth scenario as an upper bound.

shops to facilitate the secondary use of sensitive data held by the public sector – contributes to fostering trust and to increasing the amount of data available for reuse. By this study's estimates, this low-intensity intervention would bring annual benefits of EUR 725 million for the EU27. Yet, it is by itself **insufficient to reap the full benefits of increased secondary use of such data**. It requires **subsequent action, particularly by public sector data holders** that will need to make their data available via this one-stop shop, **as well as by reusers** who will need to familiarise themselves and make use of this new service. This is a **long-term process**, in part because building trust is a lengthy process, and due to potential path dependence within public sector data holders or research organisations.

As regards with Establishing a certification scheme for data altruism mechanisms, the **regulatory intervention with high intensity** – requiring Member States to establish a compulsory authorisation of data altruism mechanisms, administered by the one-stop shop established under Measures facilitating secondary use of sensitive data held by the public sector – could lead to the most beneficial outcome for Member States. It would increase the **trust** of data holders in certified data altruism mechanisms leading to an increase of shared data and thereby available data for data reusers. However, the **data altruism certification scheme** is merely the **first step** to this goal. **Member States** who have done so would **need to set up data altruism schemes**. Data reusers would still have to continue working on reaping all benefits such as analysing the data and utilising the data for e.g. new policy initiatives for the public good. In addition, data reusers will continuously have to work on **building trust** with data holders/subjects for these to share data for altruistic purposes. Only when data reusers can adequately present the benefits of data altruism to encourage data holders to share data, will this succeed in the long-term.

Likewise, as part of **Establishing a European structure for governance aspects of data sharing**, the creation of a **formal expert group** with the low-intensity regulatory intervention would **reinforce trust** by facilitating the definition and adoption of standard data sharing schemes and reference architectures that include easy legal and trust arrangements for data sharing. It would also **facilitate interoperability**, by promoting the activities for the definition of data and metadata standards, and principles for interoperability between sectors. Yet, this will **merely facilitate the activities of the existing standardisation initiatives**, and is therefore only a **first step in achieving the objectives of this intervention**. Realising the full benefits of interoperability requires these existing initiatives to flourish, and private sector actors to increase data sharing and reuse.

The low-intensity regulatory intervention under **Establishing a certification framework for data intermediaries**, creating a **voluntary labelling/certification framework for data intermediaries** administered by the one-stop shop established under Measures facilitating secondary use of sensitive data held by the public sector is also expected to significantly **increase trust** between the stakeholders in the European data market. As most of these novel data intermediaries have recently made their appearance in the market, certification is expected to provide legitimacy to their operations, functionalities offered and business models, while it would also provide mechanisms for data holders and data reusers to assess the quality and neutrality of data intermediaries' services. However, since certification will be voluntary under this framework, the **positive impacts** of this regulatory intervention **depend also on the number of data intermediaries who will decide to proceed** to the certification and comply with the certification requirements.

This study, including the **cost-benefit analysis and the macroeconomic analysis focuses solely on the direct and indirect impacts of this initial first step** taken by the Commission. It does not assess the overall benefits that the EU's economy and societies would reap following the development of data spaces by other actors. Other studies estimate that in manufacturing alone, data sharing of IoT data is expected to generate - if fully implemented – 1.3 trillion euros in increased productivity by 2027. The policy options of this intervention are a necessary first step to encourage increased data sharing in the EU.

The Impact Assessment support study took as the baseline the total economic value of the data economy for the EU27 of 306.93 billion EUR in 2020 (2.7% of the GDP).[191] These numbers take into account a correction linked to Covid-19 impact on the overall EU economy.

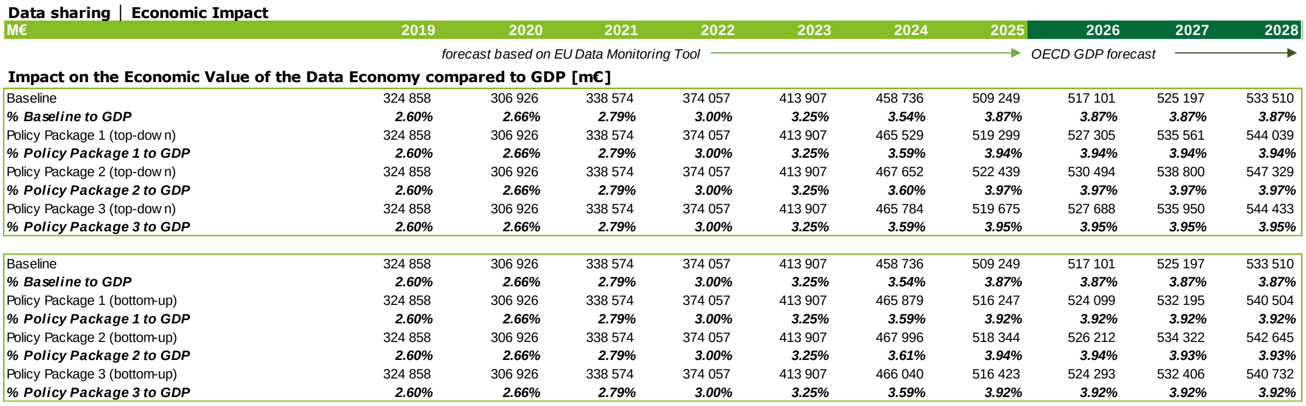
The baseline scenario foresees an autonomous growth to 533.51 billion EUR (+74%) in 2028.

In 2028, the value of the data economy could increase from 533.51 billion EUR to between 540.5 billion EUR and 544.04 billion EUR if the lower intensity regulatory intervention was introduced (from 3.87% to between 3.92% and 3.94% of the GDP).

In 2028, the value of the data economy could increase to between 542.65 million EUR to 547.33 million EUR if the high intensity regulatory intervention was introduced (from 3.87% to between 3.93% and 3.97% of the GDP);

In 2028, the value of the data economy could increase from 540.73 billion EUR to 544.43 billion EUR if the mixed regulatory intervention was introduced (from 3.87% to between 3.92% and 3.95% of the GDP).

Figure 8 – Results of the top-down and bottom-up macroeconomic impact calculations



Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
			forecast based on EU Data Monitoring Tool →					*OECD GDP forecast* →		
Impact on the Economic Value of the Data Economy compared to GDP [m€]										
Baseline	324 858	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.54%***	***3.87%***	***3.87%***	***3.87%***	***3.87%***
Policy Package 1 (top-dow n)	324 858	306 926	338 574	374 057	413 907	465 529	519 299	527 305	535 561	544 039
% Policy Package 1 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.94%***	***3.94%***	***3.94%***	***3.94%***
Policy Package 2 (top-dow n)	324 858	306 926	338 574	374 057	413 907	467 652	522 439	530 494	538 800	547 329
% Policy Package 2 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.60%***	***3.97%***	***3.97%***	***3.97%***	***3.97%***
Policy Package 3 (top-dow n)	324 858	306 926	338 574	374 057	413 907	465 784	519 675	527 688	535 950	544 433
% Policy Package 3 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.95%***	***3.95%***	***3.95%***	***3.95%***
Baseline	324 858	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.54%***	***3.87%***	***3.87%***	***3.87%***	***3.87%***
Policy Package 1 (bottom-up)	324 858	306 926	338 574	374 057	413 907	465 879	516 247	524 099	532 195	540 504
% Policy Package 1 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.92%***	***3.92%***	***3.92%***	***3.92%***
Policy Package 2 (bottom-up)	324 858	306 926	338 574	374 057	413 907	467 996	518 344	526 212	534 322	542 645
% Policy Package 2 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.61%***	***3.94%***	***3.94%***	***3.93%***	***3.93%***
Policy Package 3 (bottom-up)	324 858	306 926	338 574	374 057	413 907	466 040	516 423	524 293	532 406	540 732
% Policy Package 3 to GDP	***2.60%***	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.92%***	***3.92%***	***3.92%***	***3.92%***

[191] It must be noted that the European Data Market Monitoring Tool uses the "Value of the Data Market" as a proxy for the direct economic value. The Value of the Market is calculated based on revenues of data companies, excluding exports and including imports. It should at least be noted, that imports usually do not contribute directly to GDP, those will affect foreign GDP (whereas exports contribute to domestic GDP).

Figure 9 – Impact of the Economic Value of the Data economy compared to GDP (top-down calculation)

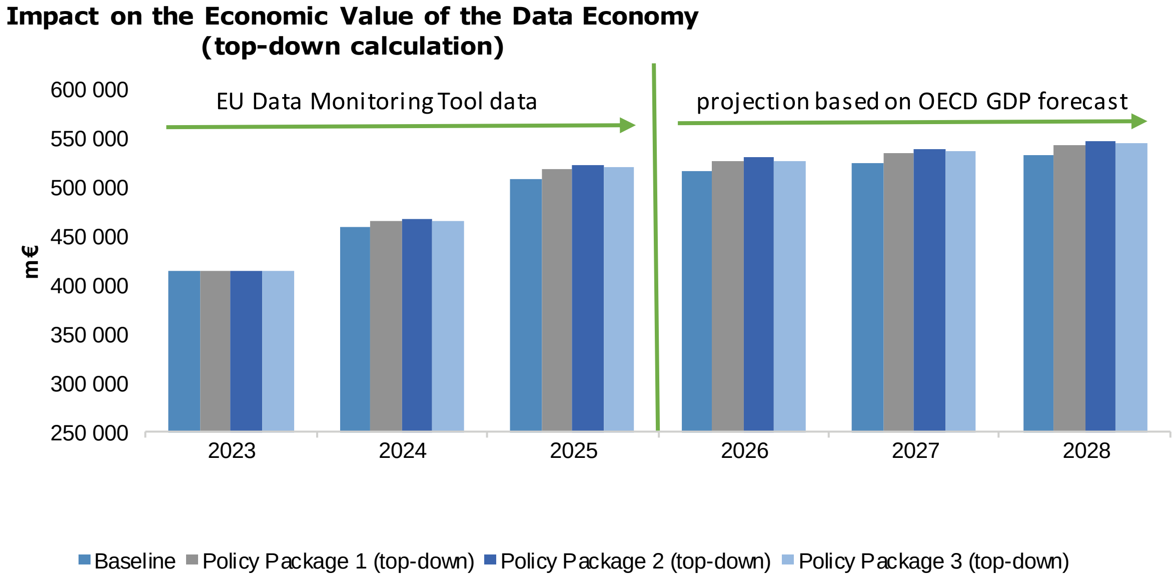
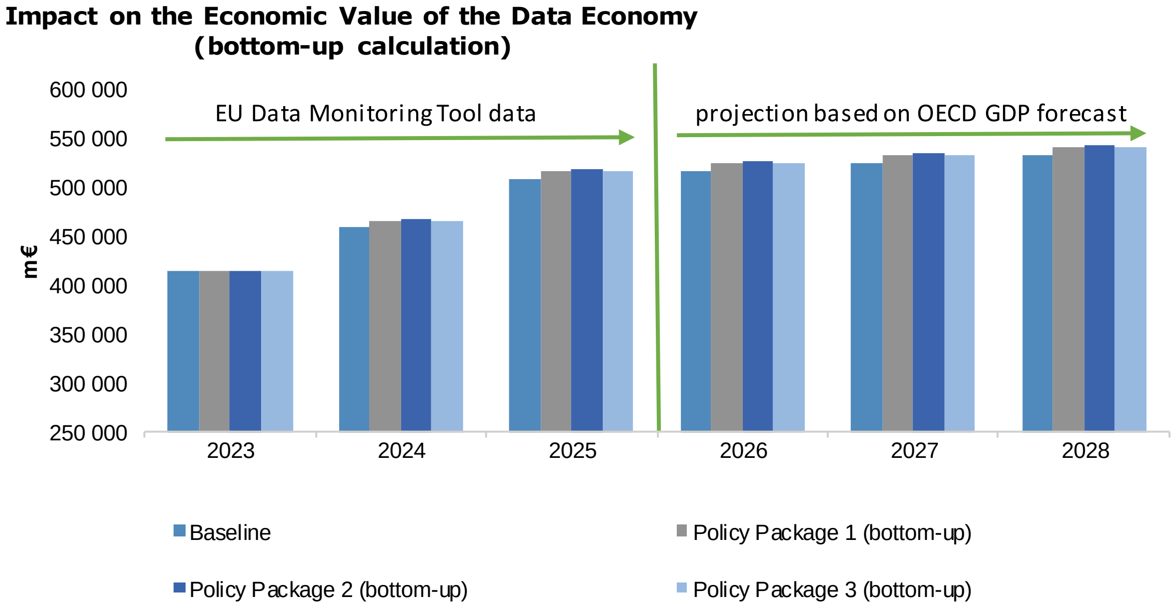


Figure 10 – Impact of the Economic Value of the Data economy compared to GDP (bottom-up calculation)



The impacts are calculated until 2025 on the basis of the value of the data economy as projected by the EU Data Monitoring Tool, which is the basis (or baseline) of our analysis. The EU Data Monitoring Tool forecast projects a growth of the data economy of approx. 8% p.a.. This forecast for the growth of the EU data economy, however, ends in 2025. In order to calculate impacts beyond 2025 we have taken a conservative approach and calculated the impacts on the basis of the GDP growth rate forecast of the OECD (1.5%-1.6% p.a.). For this reason the impacts are based on a much lower per annum growth rate for the period 2026-2028. Due to a lack of specific growth rates for the data industry, the overall OECD GDP long-term forecast was chosen as a conservative alternative. Even though it could be expected, that growth rates for the data industry might exceed the general GDP growth, for the calculation of the impact in this analysis, the incremental impact of each policy option/ policy package compared to the baseline is considered to be of most relevance, rather than the growth rate of the baseline itself.

The results regarding the economic value have been compared to the overall GDP for the EU27. For 2019, a ratio of 2.60 % compared to GDP has been estimated. This ratio was estimated to increase to 3.87% in 2028 in the baseline scenario. With regard to policy package 1 (low intensity intervention), the ratio was estimated to increase to between 3.94% (top-down calculation) and 3.92% (bottom-up calculation). For policy package 2 (high intensity intervention), an increase to 3.97% (top-down) and 3.93% (bottom-up) has been estimated. In the mixed intensity intervention (policy package 3), an increase to 3.95% (top-down) and 3.92% (bottom-up) was forecasted respectively. However, with regard to the ratios of the economic value compared to GDP, as presented in Figure 31, it should be noted, that the baseline for the years 2026-2028 has been forecasted based on growth rates of the long-term GDP forecast of the OECD.

In 2028, the total impact of the lower intensity regulatory intervention is estimated between billion EUR 6.993 and billion EUR 10.528 (bottom-up vs. top-down estimation).

In 2028, the total impact of the high intensity regulatory intervention is estimated between billion EUR 9.135 and billion EUR 13.818.

In 2028, the total impact of the preferred policy option package is estimated between billion EUR 7.222 and billion EUR 10.923.

Figure 11 - Total impact of Policy Packages (top-down calculation)

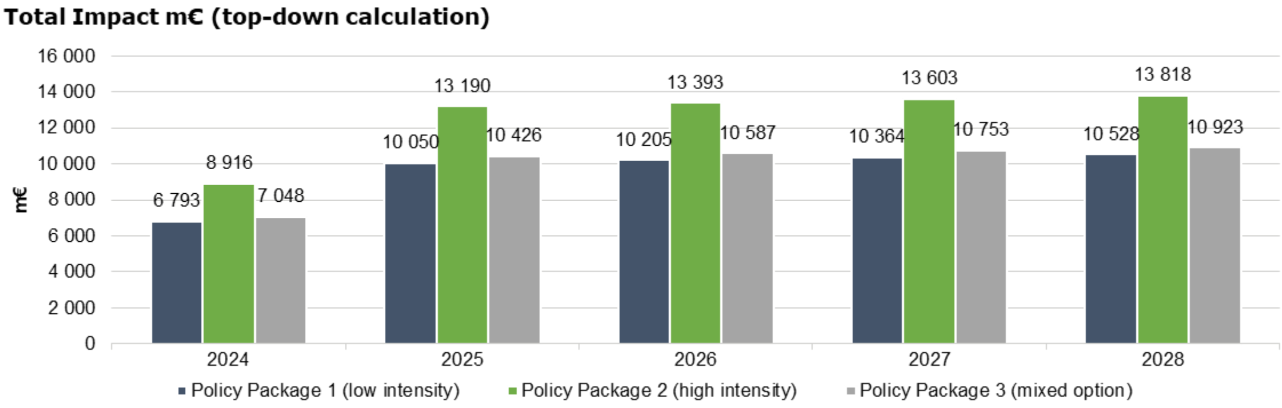
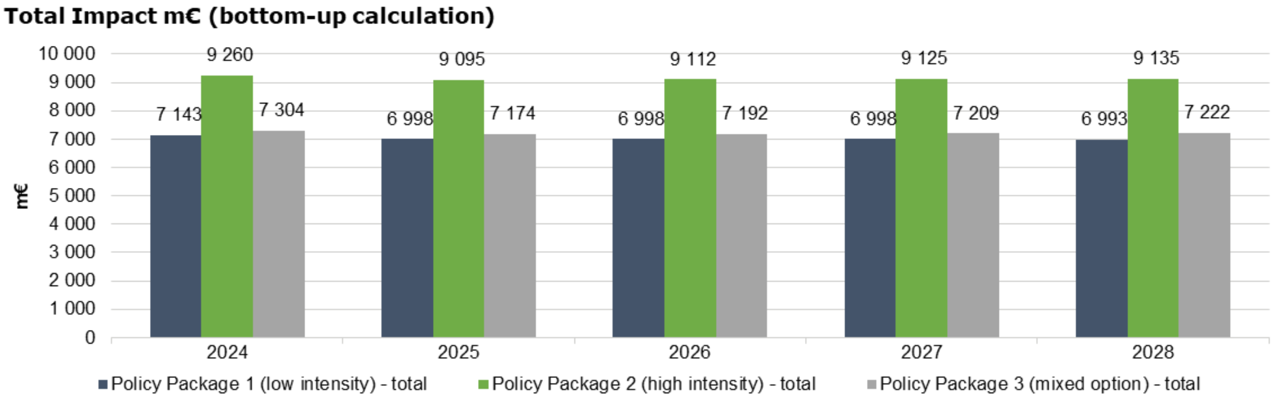


Figure 12 - Total impact of Policy Packages (bottom-up calculation)



The following figures provide an overview of the share of contribution of the policy options to each Policy Option Package. The qualitative analysis and assumptions that were made for the top-down calculation takes into account total impacts according to the European Data Monitoring Tool. The weight of each policy option was assessed in detail to understand the magnitude of its contribution to the total effect. In the case of the bottom-up calculation, a multiplier was applied to the overall package in order to integrate the most relevant indirect impacts. As a consequence, the results of the CBA, on which the bottom-up approach is based, is equally amplified across all policy options. Taken this into account, we consider that in this case, the share of contribution provided by the top-down calculation is the most relevant.

For all Policy Packages, it is the setting up of a voluntary certification scheme for data intermediaries that realises the most benefits followed by the creation of a European Data Innovation Board aiming at improving coordination in the domain of data interoperability, standards and governance. This order of share of contribution is logical, as these policy options create cross-sectorial effects. The authorization scheme for data altruism and increased sharing of sensitive data held by the public sector are also expected to yield impacts, but at a lower scale considered the smaller scope of the domain at stake compared to the overall economy.

Figure 13 – Economic impact Package 1 by Policy Option

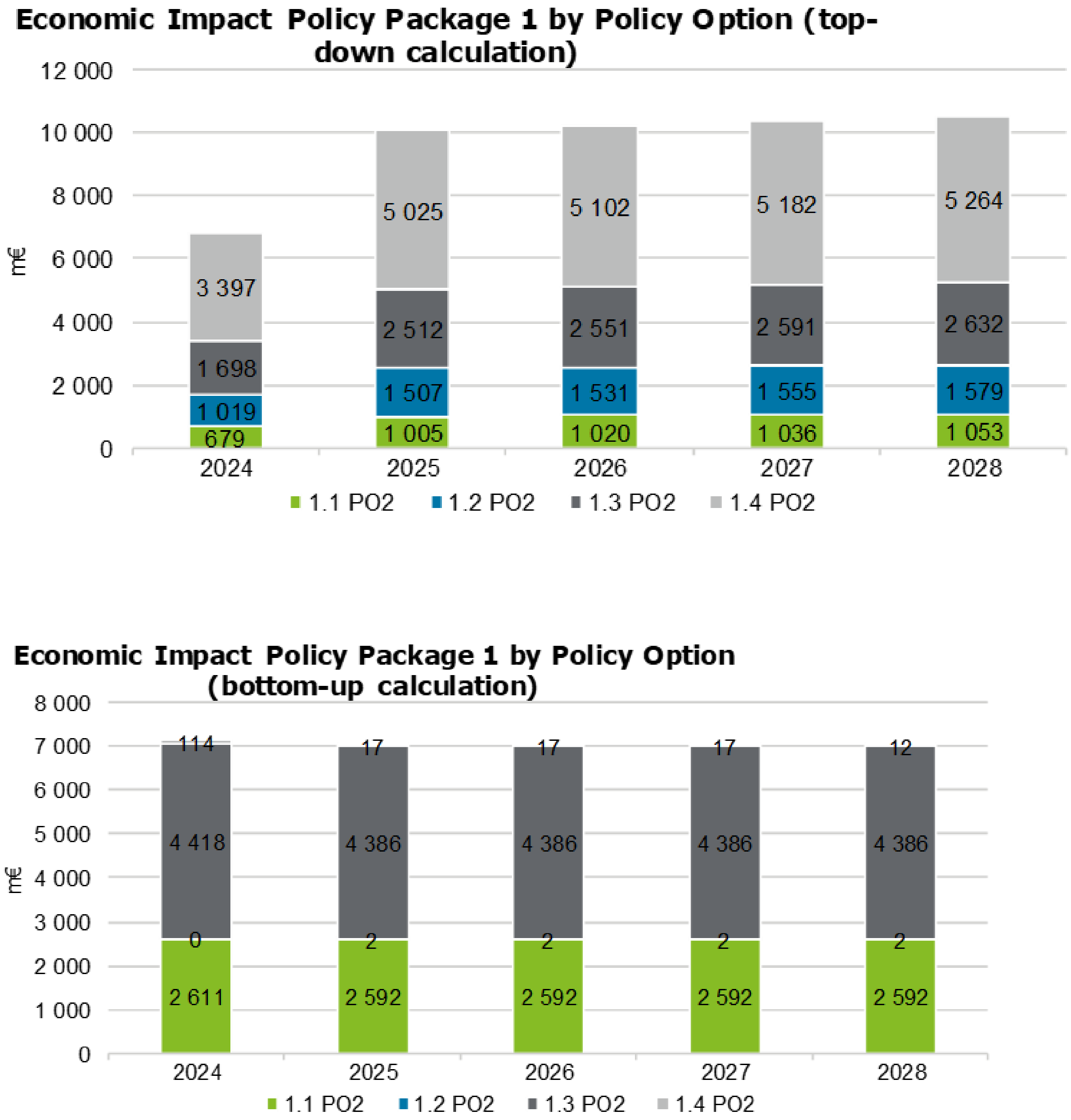
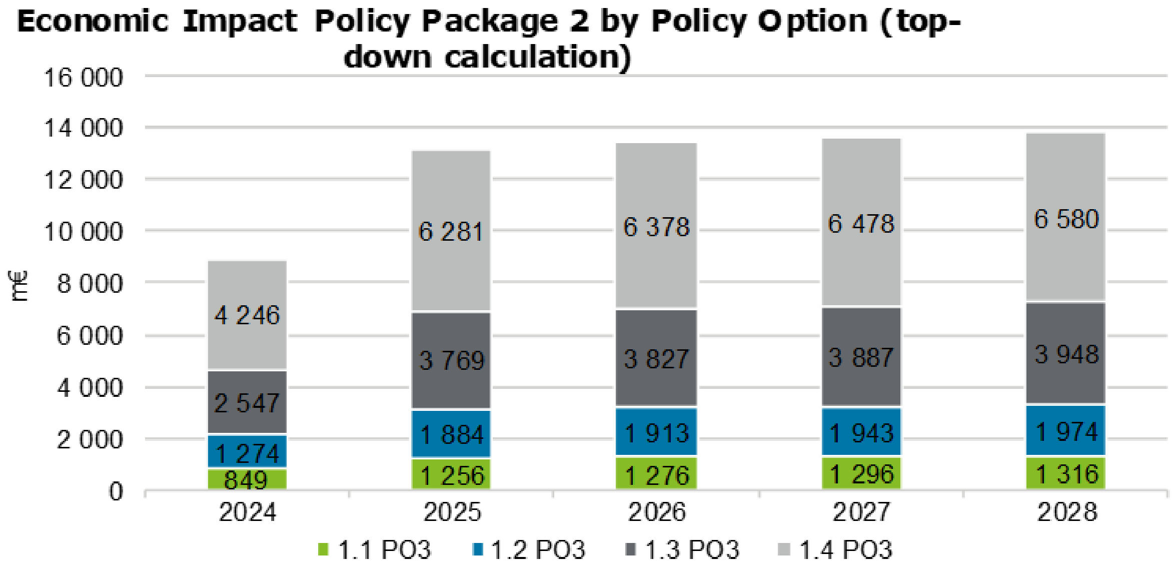


Figure 14 - Economic impact Package 2 by Policy Option



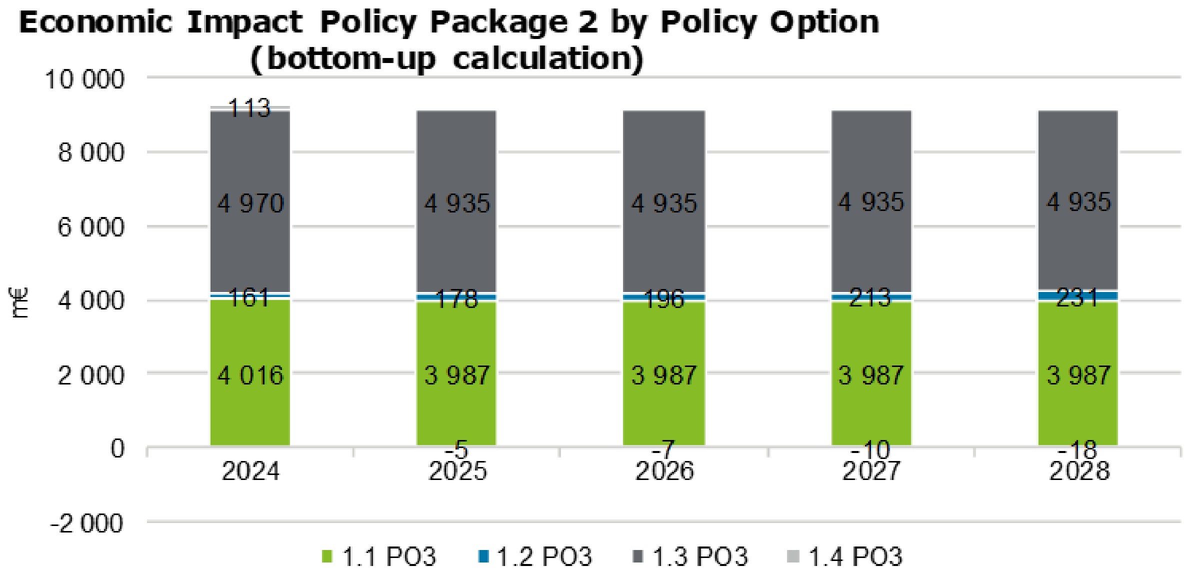
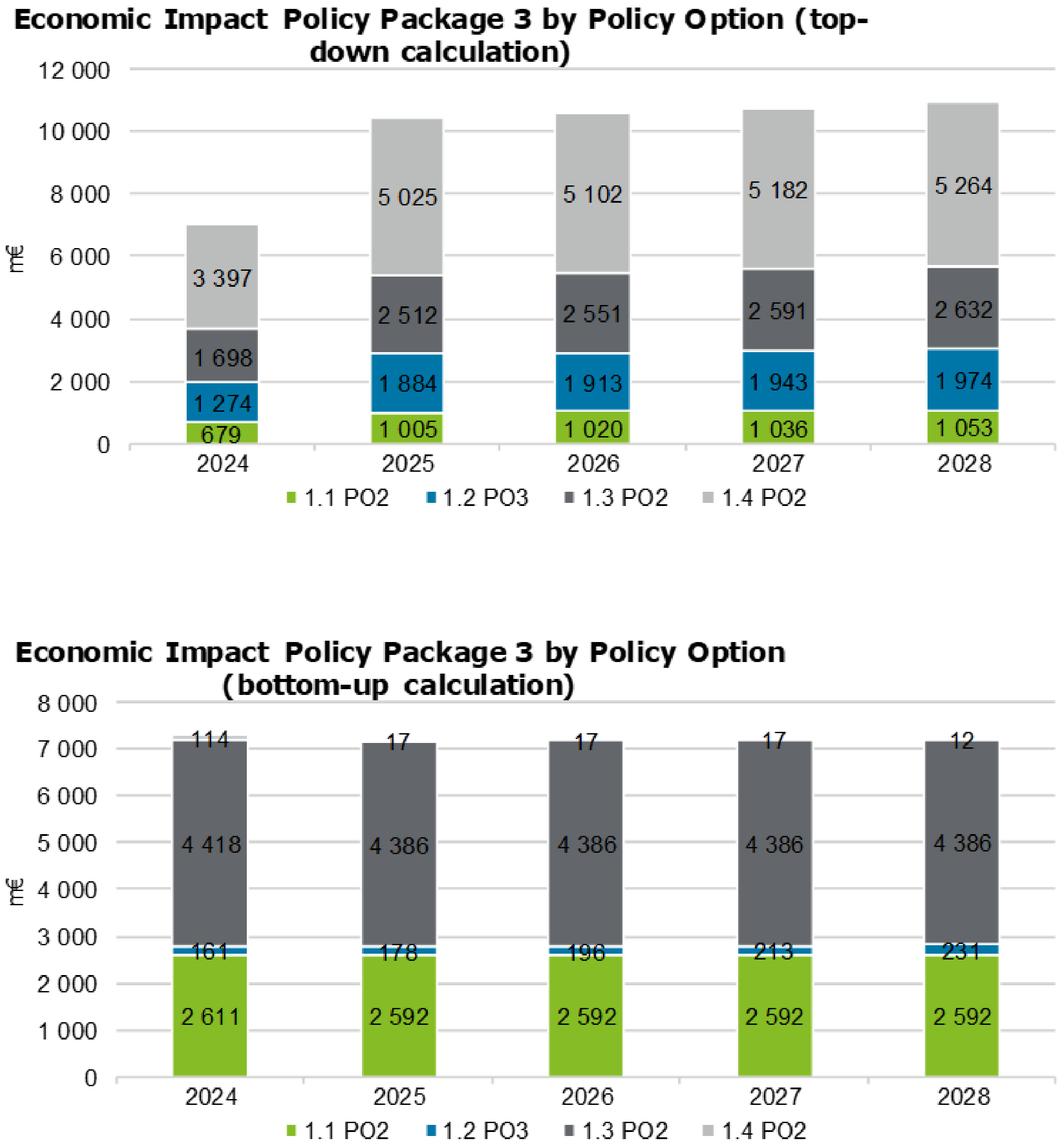


Figure 15 – Economic impact Package 3 by Policy Option



Our assessment is that all Policy Packages (low/high intensity regulatory intervention and preferred policy option packages) are creating positive impacts at the macroeconomic level, by boosting the value of the total economic value of the data economy for the EU27 from a projected autonomous growth from million EUR 306 930 in 2020 (2.7% of the GDP) to million EUR 533 510 (3.87% of the GDP) to between million EUR 540 504 and million EUR 547 329 in 2028 (3.92% to 3.97% of the GDP).

The Policy Package 2 (High intensity regulatory intervention) creates the highest impact on the total economic value of the data economy. This result is logical, as a more stringent intervention will affect more stakeholders (compared to voluntary approaches) and possibly more costs (that contribute to the European GDP at the macroeconomic level).

Policy Package 3 (Preferred Policy option package) creates important impacts on the total economic value of the data economy: more than the baseline and low intensity regulatory option but less than the high intensity intervention. It has been assessed that the combination of PO2 for Measures facilitating secondary use of sensitive data held by the public sector, PO3 for Establishing a certification scheme for data altruism mechanisms, PO2 for Establishing a European structure for governance aspects of data sharing and PO2 for Establishing a certification framework for data intermediaries hits the best score when all assessment criteria (effectiveness, efficiency, coherence, legal/political feasibility and proportionality) are taken into account.

2.5.3.1 Additional indicators

Based on the macroeconomic impacts we have estimated the impact of the policy options and policy packages on the following economic and socio-economic indicators:

- Employment (total number of additional persons employed, direct and indirect)
- Number of additional companies created statistically
- Additional governmental revenues (total gross as % of GDP incl. SSC, taxes, subsidies, governmental revenues etc.)
- Additional investment activity

To estimate the impact on these indicators, coefficients in terms of GDP-ratios have been used based on official data provided by Eurostat. With regard to numbers of person's employment, the number of additional companies and additional investment activities the GDP-ratios of the ICT-sector have been applied. Governmental revenues were calculated based on the data on tax revenue and its relationship to gross domestic product (GDP) for the EU27 in general.

2.5.3.1.1 Employment

The first indicator, employment, indicates the total number of additional persons employed (directly and indirectly) in the case the respective Policy Package will be implemented. To calculate the total number of additionally employed people, the coefficient of employment as per mEUR gross value added (GVA) was determined. This coefficient was determined to be a weighted coefficient of the EU27 per mEUR GDP/GVA in the ICT services sector. Proceeding these calculations, a constant coefficient of 10.6 for the years 2024-2028 was applied.[192] The employment coefficient indicates the per-ratio increase in employment (number of persons employed) throughout the economy which result from an increase in GDP/GVA. On average (combining the bottom-up and the top-down approaches), for the low intensity Policy Package, an additional number of employed persons of 93 045 in 2028 is expected, for the high intensity Policy Package, an incremental of 121 890 persons and for the mixed Policy Package, an additional number of 96 357 persons is expected.

The following two figures provide a detailed overview of the employment impact incremental for the three Policy Packages, based on the top-down and a bottom-up calculation of the GDP impact.

[192] The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the employment ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

Figure 16 – Employment impact incremental (top-down) in 2024-2028 for the different Policy Packages

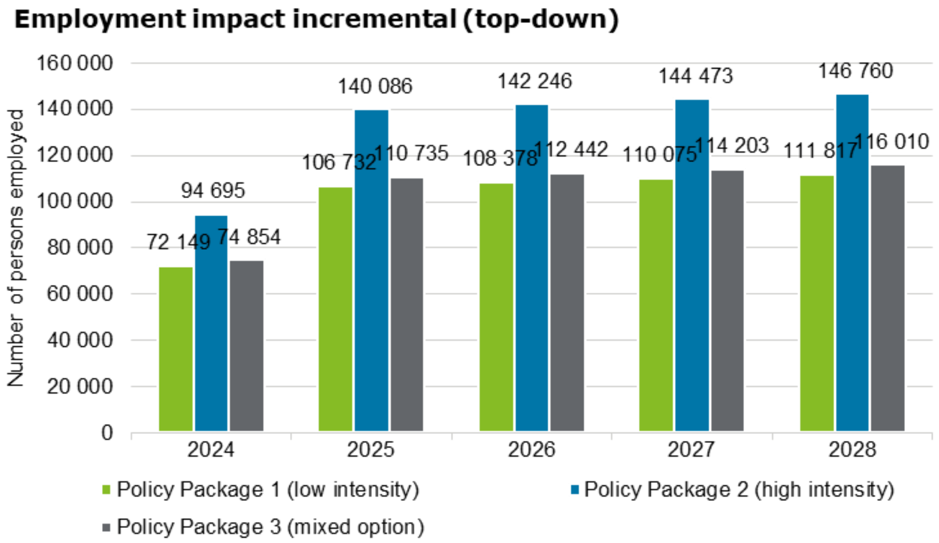
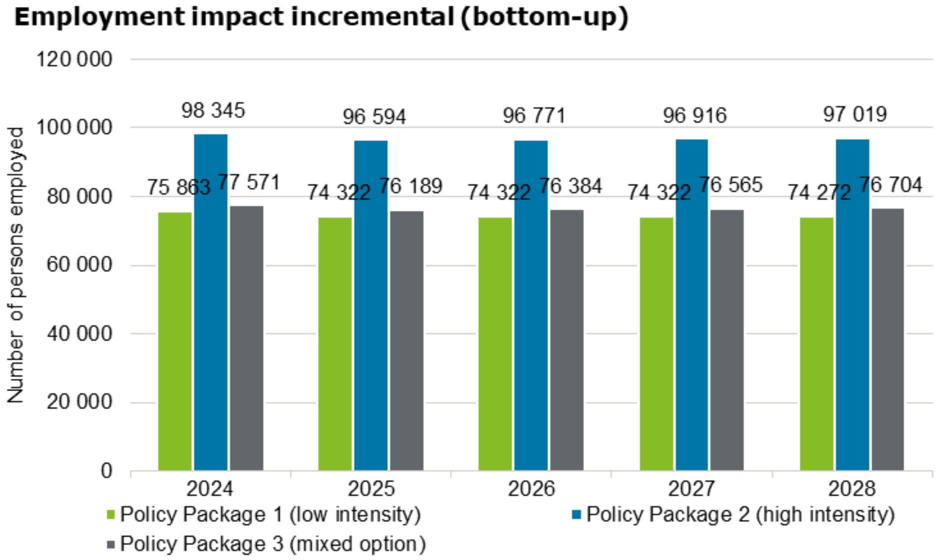


Figure 17: Employment impact incremental (bottom-up) in 2024-2028 for the different Policy Packages



2.5.3.1.2 Number of additional enterprises

The second indicator to be included in the next stage of the impact assessment is the number of additional enterprises, which would be created statistically. This variable was calculated with a weighted coefficient for the EU27 ICT-service sector, representing the ratio of enterprises per GVA. For the weighted coefficient it was assumed that an average firm in the EU27 ICT-sector has 6 employees, respectively a statistical ratio of ca. 2 enterprises per 1 million € GVA.[193] Averaging between the top-down and the bottom-up approach, in 2028 18.585 new firms will exist with the low intensity Policy Package, 24 363 additional firms with the high intensity Policy Package and 19 260 with the mixed Policy Package.

[193] The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

However, the results should be interpreted as a proxy and represent a statistical value. It should also be noted, that a part of the impact refers to indirect impacts, which are linked to downstream activities in other industries. The economic impact can also occur in existing companies in the form of expanding their activities.

Figure 18 - Number of incremental enterprises from 2024-2028 (top-down approach)

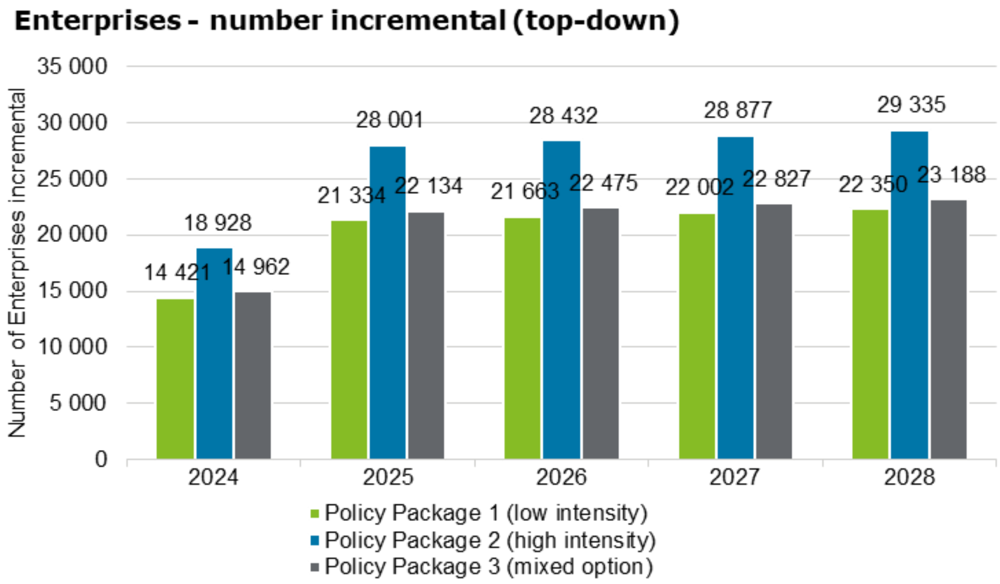
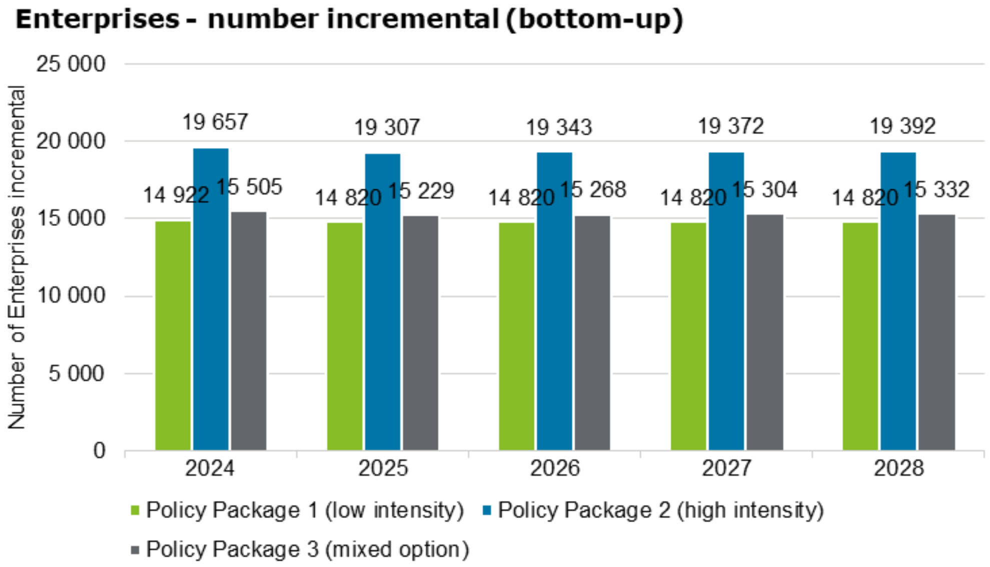


Figure 19 - Number of incremental enterprises from 2024-2028 (bottom-up approach)



2.5.3.1.3 Governmental revenues

The third indicator to be included is the governmental revenues. According to the definition of Eurostat[194], the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Total taxes are composed of taxes on production and imports, current taxes on income and wealth and capital taxes. The net social

[194] Eurostat 2020, Statistics Explained, Glossary: government revenue and expenditure. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government_revenue_and_expenditure

contribution is composed of actual social contributions by employers and households and the imputed social contributions, households' social contribution supplements and social insurance scheme service charges. Other current revenues consist of the categories property income earned, other subsidies on production received and current transfers. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained. Following the calculations of Eurostat, this coefficient has the value of 46% of GDP for the EU27. It should be noted, that part of this is related to governmental output, including market output, output for own final use and payments for non-market output, which could be linked to increased economic activity, but does not represent governmental inflows from taxes, social security payments or similar revenues.

For the governmental revenues, the top-down and the bottom-up calculations for the years 2024-2028 were averaged. This yields to an average governmental revenue of 4 030 mEUR in 2028 with the Policy Package 1. The highest governmental revenue on average can be obtained with an implementation of Policy Package 2, yielding to 5 279 mEUR in 2028. The mixed option, Policy Package 3, yields to an average amount of governmental revenue in 2028 of 4 173 mEUR.

Figure 20 – Governmental revenue from 2024-2028 (top-down approach)

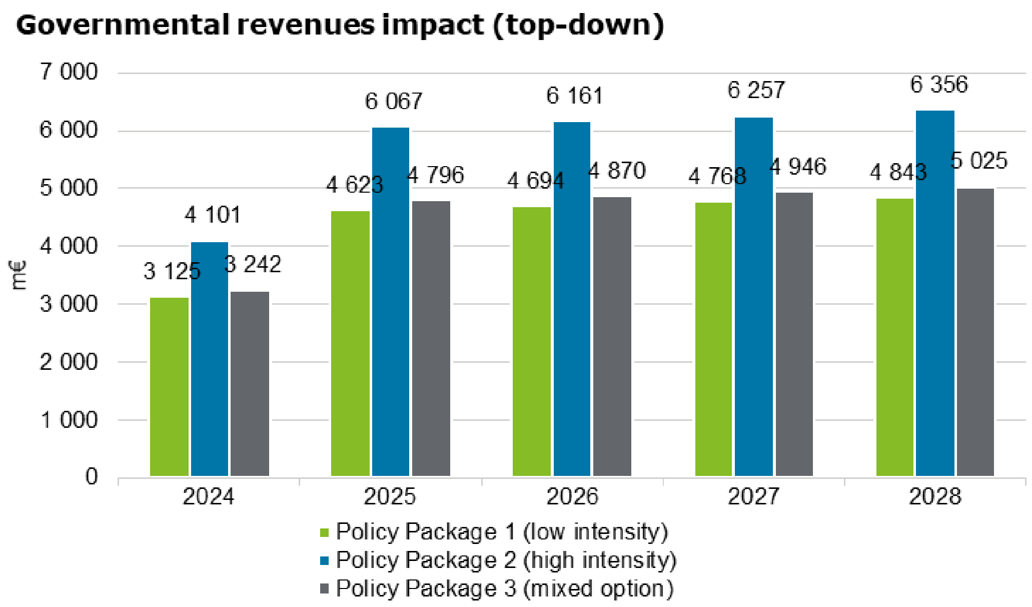
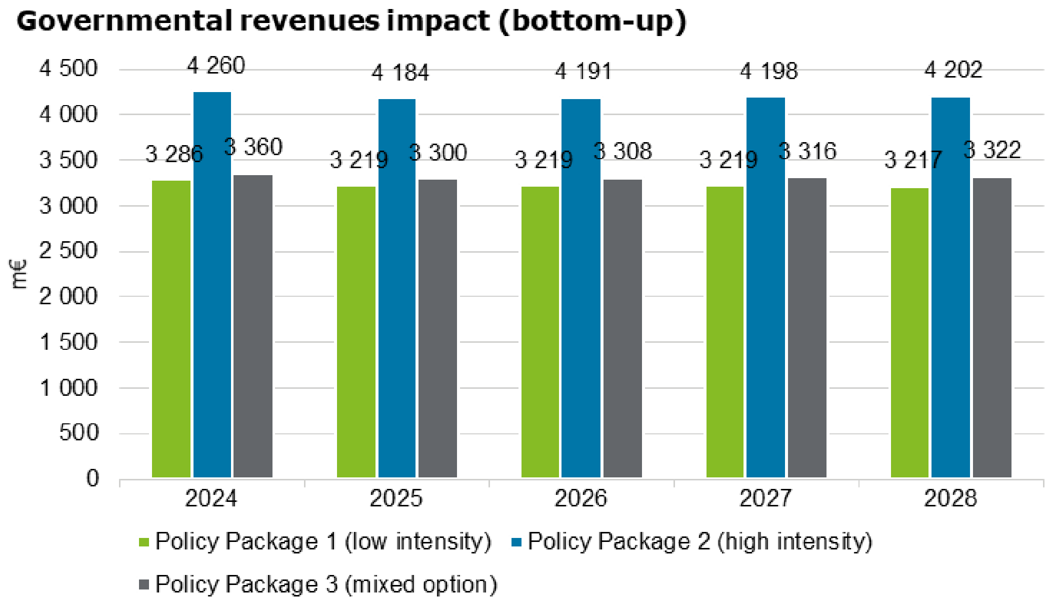


Figure 21 – Governmental revenue from 2024-2028 (bottom-up approach)



It must be noted, however, that this total governmental revenue includes – as defined in the European System of Accounts 2010 – also the market output, output for own final use and payments for non-market production. As this definition is a rather broad concept and as the macroeconomic effect of the introduction of the Policy Packages depends on a lot yet unknown factors, market output, output for own final use and payments for non-market production cannot be predicted as precisely as the other variables of governmental revenues. Excluding the categories mentioned, the adjusted governmental revenues would lower to approximately 38% of GDP according to OECD estimates.[195]

2.5.3.1.4 Investment activity

As a fourth indicator to be added we suggest to include investment activity. The investment rate is defined as the investment per value added at factor costs and is indicated as a percentage of the GDP of the EU27. The investment rate which was obtained by Eurostat[196] is at 14.4% of the GDP of the EU27 ICT-sector.

On average, an investment of 1 264 mEUR in 2028 for the EU27 can be expected with the Policy Package 1. The high intensity Policy Package 2 yields on average an investment of 1 657 mEUR in 2028, whereas with the mixed Policy Package 3 an investment of 1 310 mEUR in 2028 is expected.

[195] OECD, 2020, Comparative Statistics: Governmental Revenue. https://stats.oecd.org/Index.aspx?DataSetCode=REV
[196] Eurostat, 2020, Investment share of GDP. See: https://ec.europa.eu/eurostat/web/products-datasets/product?code=sdg_08_11

Figure 22 - Investment activities (top-down) for 2024-2028

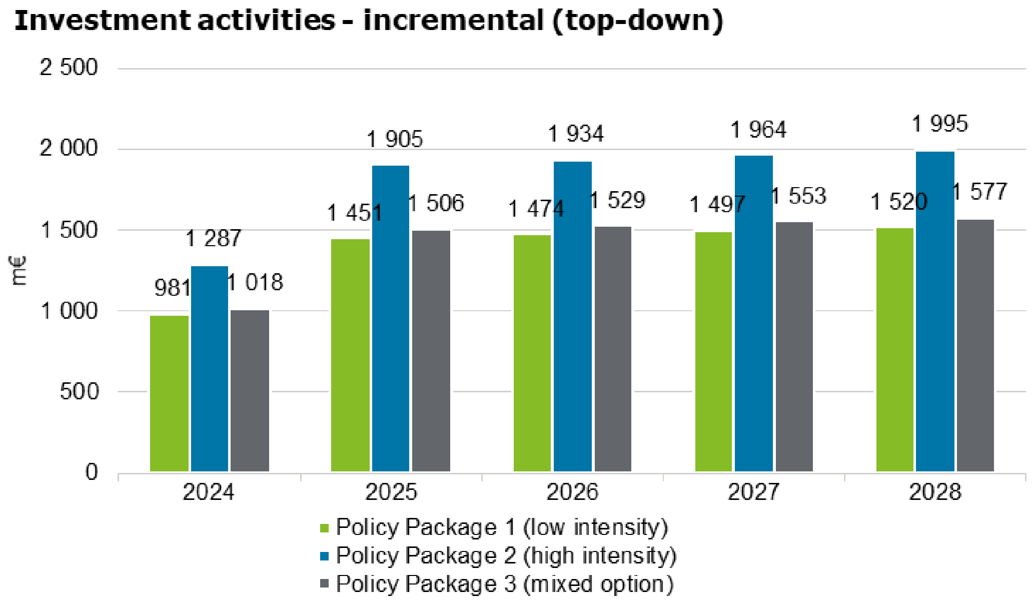
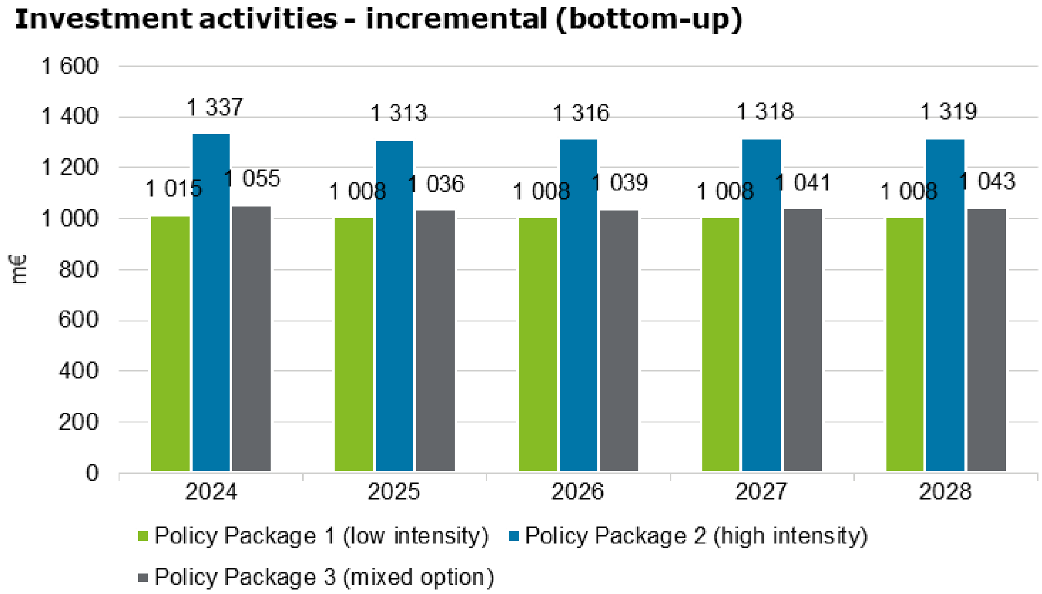


Figure 23 - Investment activities (bottom-up) for 2024-2028



2.5.3.2 Impacts linked to specific elements of preferred policy package

2.5.3.2.1 Impacts linked to the additional sharing of sensitive data by the public sector

When looking at **health data** specifically, the **potential economic benefits of increased data re-use may be very large**. A 2019 Ernst & Young report estimates that the **UK's 55 million patient records may have a value of '*several billion pounds to a commercial organisation*'**. Through increased efficiency, enhanced patient outcomes and wider economic benefits (e.g. big data, AI and personalised medicine), the curated **NHS dataset could deliver benefits worth as much as GBP 9.6 billion** (~EUR 10.7 Billion).[197] Likewise, a 2013 McKinsey study estimated that **increased re-use of health data** by both the public and private sectors would lead to **12% to 17% reduction in healthcare spending** in the United States, representing between **USD 300 billion and USD 450 billion in savings** (~EUR 266 to 399 billion).[198]

Currently, the extent to which the preferred Policy Package that includes establishing one-stop shops and secure data processing environments would contribute to achieving these economic benefits is unknown, since many **other factors have an impact** on these. For instance, the NHS' single medical market resulting in a large pool of unified data contributes to the potential value of the patient records datasets.

Nevertheless, the establishment of one-stop shops and secure data processing environments, by facilitating the reuse of health-related sensitive data and by improving interoperability between datasets, would make it easier for smaller research organisations, as well as for foreign researchers, to reuse sensitive data. It could also facilitate research re-using sensitive data from more than one Member State – and thus reusing larger datasets leading to potentially **better research outcomes and new insights**. This would contribute to **unlocking part of the value identified above**.

In addition to financial benefits, one-stop shops facilitating the reuse of sensitive health data may result in studies with a potential to **improve the lives of EU citizens**. For instance, the CASD in France facilitated the reuse by a private company of datasets related to home hospitalisation, follow-up/readaptation care, and medicine, surgery, obstetrics and odontology. This reuse resulted in the publication of several publications, including one on post-stroke spasticity and BoNT treatment in French hospitals, with a potential to improve the lives of post-stroke patients by reducing the occurrence of spasticity following strokes.[199]

Other studies resulting from reuse of data via the CASD have concerned wage inequality in different types of private companies,[200] impacts of territorial policies in France,[201] or productivity gains arising from agglomeration economies in Greater Paris.[202] While specific effects from these studies cannot

[197] EY, How we can place a value on health care data. See: https://www.ey.com/en_gl/life-sciences/how-we-can-place-a-value-on-health-care-data

[198] McKinsey & Company, The big-data revolution in US health care: Accelerating value and innovation. See: https://www.mckinsey.com/~/media/mckinsey/industries/healthcare%20systems%20and%20services/our%20insights/the%20big%20data%20revolution%20in%20us%20health%20care/the_big_data_revolution_in_health care.pdf

[199] Value in Health Journal, Patient Care Pathway for Post-Stroke Spasticity and Bont Management in French Hospitals Through the Prism of PMSI Data. See: https://www.valueinhealthjournal.com/article/S1098-3015(17)31733-3/fulltext

[200] CASD, Qualité de l'emploi dans les coopératives de travailleurs. See: https://www.casd.eu/project/qualite-de-lemploi-dans-les-cooperatives-de-travailleurs/

[201] CASD, Effets des dispositifs ZUS, ZRU, ZFU. See: https://www.casd.eu/project/effets-des-dispositifs-zus-zru-zfu/

[202] CASD, Impact des économies d'agglomération sur la productivité. See: https://www.casd.eu/project/impact-des-economies-dagglomeration-sur-la-productivite/

be precisely estimated, they are nonetheless likely to have an **indirect societal impact** by improving awareness of the issues they tackle, and by informing policy-making.

2.5.3.2.2 Impacts linked to the establishment of a certification/authorisation scheme for data altruism mechanisms

To date, the expected impact linked to the establishment of certification/authorisation scheme for data altruism mechanisms is expected to be largest for healthcare related mechanisms. This is because data holders, citizens and companies, appear to be the most willing to share data for the public good when the direct impact, better healthcare or mitigation of a health crisis, can be achieved by sharing data. Again, while this argument has been repeated multiple times throughout this study, the COVID-19 pandemic accelerated the impact data altruism can have on a society.

To achieve this, PO3 is expected to be the most effective and impactful because data holder trust is the most important to achieve a high volume of data altruism to achieve a positive outcome for society and on the economy.

Considering that the European Union currently has 446 million inhabitants the future potential of data altruism, at least considering this group of data holders, is very large and a large resource for the public good. The most important factor however is that Member States coordinate their efforts on data altruism and that citizen trust and awareness of the benefits of data altruism are continuously increased. Scientists and governments, such as the German government, are already in favour of data altruism and the benefits it can have for greater society, now data holders need to be educated and empowered, while sufficient mechanisms are created to enable data altruism.

2.5.3.2.3 Impacts linked to further governance of data and data standards

The economic benefits of greater adoption of data and metadata standards and schemes by companies are very significant: just for manufacturing, 1,4 trillion Euros in benefits of data sharing are estimated by 2027, and the few available studies convene that the main obstacles lie in lack of standards and trusted legal models for data sharing – hence we attribute a conservative estimate of 50% of the gains to the solution of these barriers.[203] There are many initiatives already today in place, but have reached limited traction so far, and there is certainly a need for increased European activity to foster the development and adoption of such initiatives. It is estimated that effective adoption of such instruments could lead to a reduction of 15% in operational expenditure.

Obviously, most of these benefits depend on decision by companies to adopt and comply with such standard, which lies entirely upon their business decision. Any European intervention, while necessary, will have only a very indirect effect on the ultimate adoption of such instruments by companies.

On the other hand, the economic benefits are only one side of the coin. Ultimately, the massive economic benefits deriving from increased efficiency will also be reflected in environmental benefits, namely through increased energy efficiency. These benefits are massive in size: the industrial sector consumes about 54% of the world's total energy according to the International Energy Outlook 2016.

[203] As previously illustrated, the Everis study on data sharing places technical interoperability as the most mentioned obstacle, by 73% of companies. Legal uncertainty about data ownership is the second, with 54%, and control over usage the third with 42%. The Deloitte Vodafone study reports costs of normalizing data, lack of standard protocols, cumbersome legal procedures, involuntary disclosure of commercial secrets as the main barriers. The WEF "Share to gain" report identifies standars, trust and legal arrangement as the key enablers.

Existing cases show a 15% energy reduction thanks to improved IoT based controls.[204] Last but not least, data sharing schemes are crucial for the efficiency of the energy sector itself.[205]

On the other hand, because the main application domain of such instruments is industry, it is difficult to quantify any form of societal impact. Of course, data sharing in domains such as pharmaceutical is already seen as key to develop new drugs, ultimately leading to gains in health,[206] but these benefits are very indirectly related to the measures under discussion.

2.5.3.2.4 Impacts linked to projects that would have benefited from the certification of intermediaries

Currently, several data intermediaries in Europe have launched initiatives that encourage and facilitate both B2B and C2B data sharing, aiming to tackle the COVID19 global health crisis and restore the economy faster through data sharing. Non-exhaustive examples of such companies include Digi.me and CitizenMe in the United Kingdom, MIDATA in Switzerland, Dawex in France, de Volksbank in the Netherlands, Polypoly in Germany and many others. In particular, a "COVID19 Hub" has been created in the Digi.me application which counts approximately 700,000 users over time across 140 countries, enabling personal data sharing in order to flatten the COVID curve faster and help restore economy for business and citizens.[207] CitizenMe has launched a research project enabling people to share information in order to tackle COVID19. CitizenMe platform has 250,000 users worldwide who use the app to share data, information and answers to questions anonymously. The results are shared openly with institutions, health organisations, researchers, journalists, charities, and the general public.[208] MIDATA's Corona Science project aims to make available to the public, as quickly as possible, a collection of anonymized/aggregated health and symptom data (stored in the MIDATA platform) in a semantic standard defined with eHealth Suisse as Open Data.[209] Finally, Polypoly's GDPR compliant Corona Protector for corporates is helping them to manage the crisis, without harming the privacy of the employees, while also enabling the trade unions to monitor the data behavior of the employer.

Given that there are currently approximately 150 data intermediaries in the European Market, with thousands of users each of them, it is estimated that the increased trust between the market stakeholders, after certification of data intermediaries, and the resulting increase in the volume of data sharing could lead to the acceleration of the time needed for resolving a global health crisis and restore the economy of approximately 25% under PO1 (industry driven self-certification scheme); 40% under PO2 (voluntary certification scheme) and 45% under PO3 (compulsory certification scheme), as significantly more stakeholders would be eager to share and use data, through data sharing platforms. This assumption is made, based on the on expected benefits by the interviewed stakeholders to arise after certification (including business development time acceleration, client base and revenue increase).

[204] See Us Energy Information Administration, 2016, International Energy Outlook 2016 and https://www.emersontopquartile.com/z-featureditems/featured-2/industrial-internet-of-things-empowering-big-time-energy-savings

[205] Douwe Lycklama et al, Data sharing: a new source for the Energy Transition in Smart Energy International 5-2019

[206] Mugdha Khaladkar and others, 'Uncovering Novel Repositioning Opportunities Using the Open Targets Platform', Drug Discovery Today, 22.12 (2017), 1800–1807 <https://doi.org/10.1016/j.drudis.2017.09.007>.

[207] https://digi.me/covid19/

[208] https://covid19.citizenme.com/public/wp/ and https://www.citizenme.com/public/wp/covid19/covid19/

[209] https://coronascience.ch/en/

2.6 Conclusion

The chapter focused in particular on four key issues which were outlined in the Data Governance Act, namely:

- The question of access and **reuse of sensitive public sector data** which are currently not disclosed by public sector bodies and not covered by the Public Sector Information (PSI)/Open data directive[210] (e.g. health data, statistical microdata, company ownership data, microdata from public transport systems and others).[211]
- The possibility of **establishing "data altruism" schemes** in Europe, defined as means of making data available (whether anonymised or non-anonymised) without expecting anything (not even services) in exchange.
- The question of facilitating data sharing through the establishment of **metadata standards** across or within sectors and including both technical and legal standards.
- The relevance of building a **certification framework** for European data intermediaries or data marketplaces which help data demand and supply to match through independent platforms.

For each of these key aspects, the study explored the state of play in Europe and determines the impact of a number of possible policy options.

Concerning the geographical scope, the study focused on the **27 European Union Member States** but it also covered case studies, examples and literature coming from third countries when relevant (i.e. experiences of B2G data sharing). Furthermore, for specific domains (i.e. domain on data altruism) the data collection and analysis activities focused on a sample of Member States. From a stakeholder perspective, the study focuses on the relevant stakeholders in the data value chain for each of the topics in scope, meaning on data holders, data intermediaries and data re-users.

This study collected data from a range of sources, including desk research, stakeholder interviews, workshops and case studies. The data collection was hampered by the fact that the public and private sector are still relatively new to navigating the data economy and could only share insights regarding costs and benefits to a very limited extent. While this study was able to collect qualitative feedback from the public and private sector on the different policy interventions discussed for each domain, it was more difficult to quantify their costs and benefits, e.g. because case numbers are still small or the data sharing practices are just emerging and stakeholders themselves do not yet know their scale and/or costs of making data available. In addition, the stakeholders consulted do not yet have a final and consolidated perception on for example the potential benefits they could draw from increased data use and availability in their respective domain, besides speculative thoughts. This report should be considered as a first attempt at examining this topic and gathering the existing data on these subjects. This analysis is therefore based on the limited data available and provides a preliminary (mainly qualitative) overview of the costs and benefits for the different topics under scrutiny. The conclusions reached are based on independent judgement and specific to this study.

The assessment of the policy options for each domain enabled the study to formulate several policy packages combining one policy option per domain. These policy packages are:

- **Policy Package 0 – Baseline**: the baseline scenario consists in applying no policy changes to the four areas for which problems could be identified: sensitive data held by the public sector,

[210] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024
[211] In agreement with the Commission, this study focuses on the former two.

data altruism schemes, certification of data intermediaries, governance and standards. The EU economy will not be able to reap the benefits of data sharing.

- **Policy Package 1** – Low intensity regulatory options: this package includes the creation of a one-stop shop to foster the sharing of (sensitive) data whose use is subject to the rights of others held by the public sector. A voluntary certification scheme would be established by EU Member States for data altruism mechanism and organisations offering such schemes. Data intermediaries will also be able to obtain a certification to demonstrate their neutrality and absence of conflict of interest (e.g. absence of competition with data users) on a voluntary basis. Finally, the European Data innovation Board would take the form of a formal expert group created by the European Commission, including Member States representatives and industry representatives.
- **Policy Package 2** – High intensity regulatory options: Under this package, Member States will be required to set up a Single Data Authorisation body in charge of providing the authorisation to enable the further use of data that is subject to the rights of others contained held by the public sector. This entity will also be in charge of delivering the compulsory authorisation required from organisations offering data altruism schemes, as well as mandatory certification scheme for data intermediary. Under this package, the European Data Innovation Board would consist of an independent European body with legal personality, supported by a secretariat.
- **Policy Package 3** – Preferred policy options: this package is similar to Policy Package 1, to the exception that a compulsory authorisation is set-up for organisations offering data altruism schemes.

In a last step, this study conducted a macroeconomic analysis of these packages. To do so, the team has calculated the baseline scenario using the forecasts of the European Data Market Monitoring Tool, corrected for the impact of the Covid-19 crisis. A top-down analysis of the policy packages was then performed, as well as a bottom-up analysis based on the cost-benefit results of the policy options. It found that by 2028, the value of the data economy could increase from EUR 533.51 billion:

- To EUR 540.5 billion – 544.04 billion with the lower intensity regulatory intervention;
- To EUR 542.65 billion – 547.33 billion with the higher intensity regulatory intervention; and
- To EUR 540.73 billion – 544.43 billion with the mixed regulatory intervention.

Yet, while Policy package 2 yields the highest impact on the total economic value of the data economy, Policy package 3 ranks highest when including other assessment criteria such as particularly coherence, and legal/political feasibility and proportionality.

3 Measures to foster data sharing and re-use

This chapter provides the assessment of key issues identified as part of the challenge to foster data sharing and re-use in the EU. The problems, its causes and effects are explored, based upon which the policy objectives and options are set out to address these. These options are then assessed along five main criteria as part of a multi-criteria analysis to determine the preferred option in four key areas. Finally, the macro-economic impacts are derived.

3.1 Background and problem assessment

This section contains the problem assessment of issues related to Business to Government Data Sharing (B2G) for the public interest, citizen empowerment, rights over co-generated data, and conflict of laws at the international level.

3.1.1 Measures to encourage Business-to-Government data sharing for the public interest

3.1.1.1 Background

3.1.1.1.1 Context

Business-to-government data sharing refers to **privately held data being made accessible to (or shared with) public authorities** to enable them to fulfil a public interest purpose.

With the exponential increase in data over recent years, most of the world’s data is controlled by relatively few private companies.[212] The Covid-19 crisis has shown the **essential role of data use for crisis management and for informed decision making** by governments. In the wake of the public health crisis following the coronavirus outbreak and the subsequent governmental measures imposed around the world, EU Member States and the European Commission sought anonymised mobility data from mobile network operators in order to **help public authorities** track issues including the spread of the virus and the effectiveness of social distancing policies.

Beyond the role of B2G data sharing to **solve societal issues**, and as pointed out by the JRC[213], **there is an emerging market for B2G data sharing**. There is a plethora of ways in which data held by private organisations could be used in a transparent and proportionate way by public authorities where clearly necessary for purposes that serve the public interest. As the Commission’s communication on the European Green Deal notes, ‘It will be important to ensure that across the EU, investors, insurers, businesses, cities and citizens are able to access data and to develop instruments to integrate climate change into their risk management practices.”[214] Such data can be put to use

[212] https://www.digitalinformationworld.com/2018/12/here-is-what-the-big-tech-companies-know-about-you.html; https://www.weforum.org/agenda/2021/03/europe-digital-sovereignty/
[213] JRC (2020), The economics of Business-to-Government data sharing. See: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf
[214] COM/2019/640 final

for objectives from **making transport and energy systems more environmentally friendly**, to providing **more effective regional and urban planning**, to **improving education**. Recent studies provide rough estimates suggesting that data access and sharing can help generate social and economic benefits worth between 0.1% and 1.5% of gross domestic product (GDP) in the case of public-sector data, and between 1% and 2.5% of GDP (it can go as high as 4% of GDP) when also including private-sector data.[215]

Five dimensions have been proposed to measure the ways in which data shared by private organisations can provide value to public sector organisations:

- Discovery of **new insights** (situation analysis and cause and effect) leading to better understanding of problems and opportunities within the society;
- **Faster and more accurate decision-making** (based on improved situational awareness, and on a better linkage between cause and effect);
- Increased prediction accuracy and proactivity in preventing crises before they occur;
- **Optimised process efficiency and coordination** (leveraging rapid experimentation and impact assessment, but also better and often real-time monitoring and evaluation of a policy); and
- Increased public service **delivery and innovation**.

The Commission in its Data Strategy stated that 'making more data available and improving the way in which data is used is essential for tackling societal, climate and environment-related challenges, contributing to healthier, more prosperous and more sustainable societies.'[216] As a further example, the **United Kingdom's National Data Strategy identifies five opportunities** from greater B2G data reuse: boosting productivity and trade, supporting new businesses and jobs, increasing the speed, efficiency and scope of scientific research, driving better delivery of policy and public services, and creating a fairer society for all.[217]

Increased B2G data sharing would bring **benefits not only to government, but also to the private sector**. A clear framework for data sharing would provide legal certainty and reduce the cost, inconvenience and disruption of multiple, unclear and uncoordinated requests for data from different arms of government. Data holders may decide to share their data on the understanding that they would **reciprocally gain access to other data domain expertise, insights resulting from the data analysis to make better business decisions**, reputational benefits the data sharing could bring or the business opportunity to sell their data should it be the case. Even when this is not the case, allowing external researchers to analyse their data can enable data holders to **benefit from analytical skills they do not possess** and may not be in a position to acquire. Sharing data with the public sector can also improve a private sector organisation's **reputation and brand image**, resulting in potentially increased visibility and media attention, and in turn in potentially increased attractiveness to users, customers, employees, and investors. Lastly, B2G data sharing can in some cases help private sector organisations achieve their **social corporate responsibility** (or philanthropy) goals.[218]

Realising these benefits, however, **will require increased business-to-government data sharing and reuse practices** compared to the current situation. This will require **more systematic, sustainable and responsible methods of B2G data sharing** for the public interest.

[215] OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd-ilibrary.org/sites/276aaca8-en/index.html?itemId=/content/publication/276aaca8-en

[216] COM(2020) 66 final

[217] UK Department for Digital, Culture, Media and Sport, National Data Strategy. See: https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy

[218] A. Young & S. G. Verhulst, Data Collaboratives. See: https://link.springer.com/referenceworkentry/10.1007/978-3-030-13895-0_92-1

While currently there are a range of existing models (see section 3.1.1.1.3.), these tend to be ad hoc rather than systematic and sustainable.[219]

To achieve this increase in data access and reuse activities, however, *"a more active public policy stance may be required"* to overcome the **barriers to B2G data sharing and reuse.**[220] These are outlined in the subsection below.

3.1.1.1.2 Ecosystem

This section provides a preliminary overview of the ecosystem. It first identifies the types of stakeholders and of datasets categories concerned, before providing an overview of existing models for B2G data access and reuse. Lastly, it provides a list of possible case studies and interviewees.

Data is constantly produced by citizens and companies acting as co-producers. This data is then hosted by the company and constitutes the supply side of the B2G data ecosystem. In parallel, governments seek access to data for a variety of purposes, such as to inform policymaking for the public interest or design more targeted services, or to respond to public emergencies. This constitutes the demand side of the B2G data market. Demand and supply meet through a variety of data collaboratives, i.e. methods for B2G data sharing which are explored in the subsection below.

The table below provides an indicative list of the main stakeholders in scope of Business-to-Government (B2G) data sharing for the public interest.

Table 57 - Stakeholder scope (data value chain mapping)

Domain	Data holder	Data re-user	Personal data?	Purpose
Business-to-Government (B2G) data sharing for the public interest	Private sector organisations	Public sector: national executive government (e.g. statistical offices), regional and local government (e.g. municipalities), legislative branch (e.g. parliamentary research services).	Sometimes	Public interest (undefined, may relate to health, the environment, education, economy, transport, etc.)

The first type of stakeholders are **data holders**, or defined by the GDPR as data controllers, "*a legal or natural person, an agency, a public authority, or any other body who, alone or when joined with others, determines the purposes of any personal data and the means of processing it.*".[221]

In this study, data holders are private organisations that hold data either as their main activity (such as satellite companies), or as a by-product of their main activities (any company that has undergone a digital transformation).

The second type of stakeholders are **data (co-)producers**, defined as a software service, organisation (public or private), natural person, or combination thereof, providing data to a data storage system.[222]

[219] Micheli, M., Accessing privately held data: Public/private sector relations in twelve European cities. See: https://zenodo.org/record/3967044#.X2Hy7WgzaUn

[220] *Ibid*.

[221] GDPR. https://www.gdpreu.org/the-regulation/key-concepts/data-controllers-and-processors/

[222] Information Management, Glossary. See: https://www.information-management.com/glossary/d.html

In this study, data co-producers refer to both private companies producing data as their main activity or as a by-product of their main activities, and to citizens to whom the data relates. The latter are **out of scope** of this study. The former is in scope, but are referred to in this study as data holders instead.

A third category of stakeholders are **data re-users** in the public sector, which can be defined as "*generating the social value of data sharing*" through their use of the data.[223] They may be public bodies, or researchers and accordingly may use data to generate insights that can support public interest missions and decisions including (but not limited to):

- Prevention and mitigation of the effects of **climate change**;
- Official statistics;
- Urban planning;
- **Health** epidemics/pandemics management;
- Enabling scientific research and technical development;
- Better understanding of **migration** patterns;
- **Poverty and inequalities** eradication and reduction;
- Support agricultural policies;
- Improving **tourism** management; and
- Improving **mining and industrial policy,** coordination and control.

The last stakeholder category is **data intermediaries**, which primarily "*enable data holders to share their data, so it can be re-used by potential data users*", although they may also provide other services such as processing services.[224]

As the next section will show, data intermediaries are relevant in one specific type of B2G data collaborative. However, they are not in all other five and such intermediaries are currently rare. As a result of the market of data intermediaries being undeveloped[225], they **are excluded from this study's scope**.

Data held by the private sector is extremely varied and may originate from a variety of business activities. Most commonly, the following types of data may in certain circumstances be assessed as necessary for specific purposes that serve clear public interest goals. :

- Data on efficiency and circularity of products and materials
- data on types, volume and location of waste material
- Data emerging from producing products and goods (e.g. crops and seed data, manufacturing data, maintenance data, etc.); and
- Data emerging from managing infrastructures and natural assets (e.g. energy grid performance, water supply, forest data, broadband performance, etc.).
- Data resulting from consumption, commercial and financial transactions (e.g. payment transactions, ATM data, credit ratings, stocks and asset-related information etc.);
- product supply chain/ logistics data and anonymised and data on travel patterns of products (e.g. trip data, location data, motorway toll collection, check-in data, etc.);
- Data resulting from media and entertainment consumption (e.g. overall subscription trends, online sharing of content, viewing patterns, apps and games usage, browser cookies, etc.);

[223] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm
[224] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm
[225] An estimation of the total number of data intermediaries active in the European market could include an average number of 150 organisations, while the number of data users or data holders affected could entail any European company or individual wishing to buy or sell data through the intermediaries.

- Data resulting from **home assistance and other in-home, IoT, sensor devices** (e.g. energy use, water use, room temperature, purchase habits, visitors' log, security sensors, CCTV, etc.);

A more granular view of the types of datasets falling under each category, along with additional information, is available in Annex II.

3.1.1.1.3 Existing models for B2G data access and reuse

As noted by the High-Level Expert Group on Business-to-Government Data Sharing, **B2G data-sharing collaborations are mostly one-off pilots** at this point in time. These tend to take **different forms according to their context and objectives**, and can be described according to where they are on an '**open-restricted access' continuum**:

- On the 'restricted access' end of the continuum, the data holder analyses its data and shares the insights from that analysis;
- The data holder hosts researchers on-site to analyse data and share the findings from that analysis;
- Several data holders work together to share data and/or insights among themselves and with a public authority;
- A data holder shares data with trusted public authorities; and
- On the 'open access' end of the continuum, a data holder allows the public sector direct access to some of its datasets.

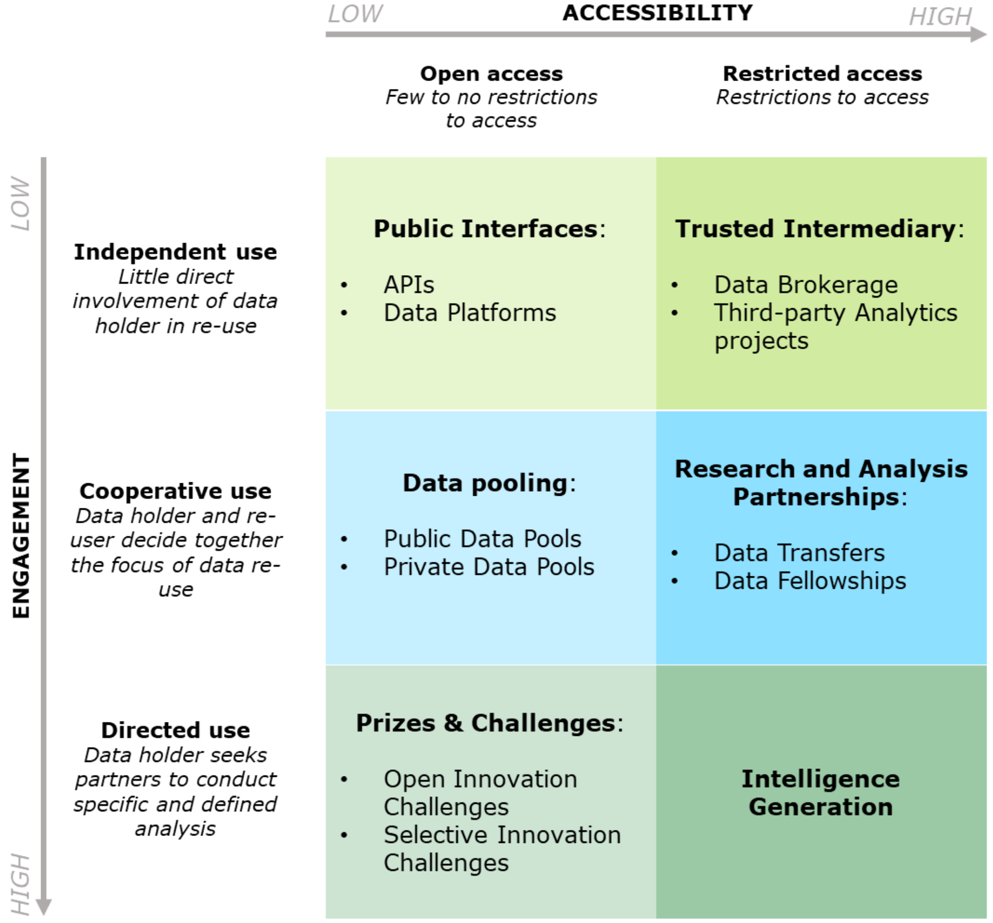
A **typology of six categories of data collaborations** [226] – in which private sector organisations provide access to their data for the public good – has been put forward. In addition to the **level of accessibility**, i.e. the 'open-restricted access' continuum, they **also differ in their level of engagement**, i.e. the degree to which private sector data holders and public sector data re-users co-design the reuse of the data. The six categories of data collaborations, graphically represented in the figure below, are:

- **Public interfaces**, where a company provides public access to pre-processed data for independent use by re-users. One the one hand, **Application Programming Interfaces** (APIs) publish data automatically and near real-time, either under certain conditions on acceptable use, or on an open access basis. On the other hand, **data platforms** make private sector data accessible through web or mobile applications – often with a target group in mind – at a lesser cost in terms of data or software-development expertise.
- **Trusted intermediary**, where data re-use remains independent from the data holder, but where access to data is restricted to their intended recipient. This can take the shape of **data brokerage**, where third parties match the supply and demand of data on a purpose-bound and time-bound basis. Alternatively, **third-party analytics projects** consist in trusted intermediaries (e.g. research organisations and non-profits) accessing and analysing private sector data, but sharing only the resulting insights with the public sector. Access to data is thereby more restricted, but this approach brings external expertise that may not be available through direct collaboration.
- **Data pooling**, where access to data is usually open and where data uses range from independent to highly cooperative. Approaches include **public data pools**, in which data from multiple data holders are made available on the web. While usually intended for contributing partners, access tends to be open and free. **Private data pools**, by contrast, are only available to approved partners.

[226] GovLab (2019), Data Collaboratives, Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices. Available at https://datacollaboratives.org/static/files/existing-practices-report.pdf

- **Research and Analysis Partnerships**, a cooperative practice whereby private sector companies transfer data to public sector partners (e.g. statistical offices) for targeted analysis of interest to the company. This can be done via **data transfers** (sometimes referred to as ‘data philanthropy’) in which access to data is restrictive in terms of who can access it, and of why and how the data can be used. **Data fellowships** allow specific individuals or parties to access and analyse data for a set period of time.
- **Prizes and Challenges**, i.e. competitions providing open access to certain datasets and encouraging competitors to address challenges and/or opportunities defined by the organisers. These challenges can be **open innovation**, where barriers to entry are lower and independent use of the data falling out of the intended scope are possible. **Selective innovation** challenges, by contrast, involve a more restricted access to data and pre-approved uses of the data to address a given issue. Because of these restrictions, the data shared may be more sensitive.
- **Intelligence Generation**, where the private sector company analyses its data and makes available only the insights of the analysis. It is similar to third-party analytics projects, but without the third party.

Figure 24 - Six categories of data collaboratives



	Open access *Few to no restrictions to access*	**Restricted access** *Restrictions to access*
Independent use *Little direct involvement of data holder in re-use*	**Public Interfaces**: • APIs • Data Platforms	**Trusted Intermediary**: • Data Brokerage • Third-party Analytics projects
Cooperative use *Data holder and re-user decide together the focus of data re-use*	**Data pooling**: • Public Data Pools • Private Data Pools	**Research and Analysis Partnerships**: • Data Transfers • Data Fellowships
Directed use *Data holder seeks partners to conduct specific and defined analysis*	**Prizes & Challenges**: • Open Innovation Challenges • Selective Innovation Challenges	**Intelligence Generation**

Source: Compiled by Deloitte from the GovLab (2019), *Leveraging Private Data for Public Good.*

In addition to these six, **data scraping** is a practice where users download publicly available data and analyse it, sometimes with a view to generating and sharing insights of public value. This is done, in some cases, without any participation of the data holder, and raises legal (where there is no legal basis for processing personal data) and ethical questions (for instance, when the data was not generated to be further analysed). Because data holders do not play any role however, data scraping is excluded from the above framework.

3.1.1.2 The problem, its magnitude and the stakeholders affected

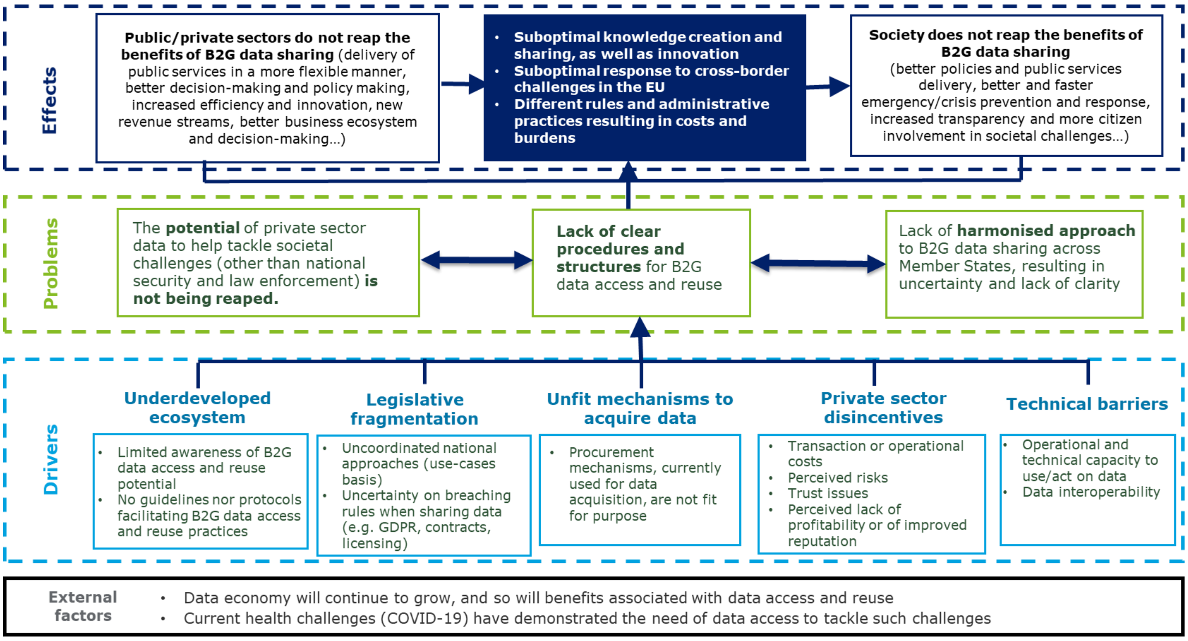
This section identifies the problem, its causes and its effects – all graphically represented in the problem tree below.

The lack of a harmonised approach to B2G data access and reuse, with clear procedures and structures, results in the **public sector having limited access to private sector data**. Due to this, the **potential of private sector data to help tackle societal challenges is not being reaped** and the ability to face cross-border challenges in the EU is limited. In addition, a **lack of a harmonised approach to B2G data access and reuse** also results in companies being subject to different rules and administrative practices, affecting competition in the EU internal market, and in several requests for the same or similar data from different authorities making it very time consuming. This was exemplified during the early days of COVID-19 where different national and local agencies in different member states sought to access aggregated location data from often the same MNOs generating both uncertainty and substantial transaction costs for both data holders and data users.

As pointed out in a 2020 JRC Technical Report,[227] "*there are likely to be significant potential economic benefits from additional B2G data sharing operations*". For instance, if one party collects the data and this is shared with many other users, such as public bodies, substantial cost savings for society could result, such as avoiding repeated collection of the same data. Additionally, different users may produce new and innovative outputs with the very same set of data, increasing the value of the datasets.

[227] JRC (2020), The economics of Business-to-Government data sharing. See: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf

Figure 25 - Business-to-Government (B2G) data sharing for the public interest Problem tree



Effects

Public/private sectors do not reap the benefits of B2G data sharing (delivery of public services in a more flexible manner, better decision-making and policy making, increased efficiency and innovation, new revenue streams, better business ecosystem and decision-making…)

- **Suboptimal knowledge creation and sharing, as well as innovation**
- **Suboptimal response to cross-border challenges in the EU**
- **Different rules and administrative practices resulting in costs and burdens**

Society does not reap the benefits of B2G data sharing (better policies and public services delivery, better and faster emergency/crisis prevention and response, increased transparency and more citizen involvement in societal challenges…)

Problems

The **potential** of private sector data to help tackle societal challenges (other than national security and law enforcement) **is not being reaped.**

Lack of clear procedures and structures for B2G data access and reuse

Lack of **harmonised approach** to B2G data sharing across Member States, resulting in uncertainty and lack of clarity

Drivers

Underdeveloped ecosystem

- Limited awareness of B2G data access and reuse potential
- No guidelines nor protocols facilitating B2G data access and reuse practices

Legislative fragmentation

- Uncoordinated national approaches (use-cases basis)
- Uncertainty on breaching rules when sharing data (e.g. GDPR, contracts, licensing)

Unfit mechanisms to acquire data

- Procurement mechanisms, currently used for data acquisition, are not fit for purpose

Private sector disincentives

- Transaction or operational costs
- Perceived risks
- Trust issues
- Perceived lack of profitability or of improved reputation

Technical barriers

- Operational and technical capacity to use/act on data
- Data interoperability

External factors

- Data economy will continue to grow, and so will benefits associated with data access and reuse
- Current health challenges (COVID-19) have demonstrated the need of data access to tackle such challenges

3.1.1.3 The causes of the problem

This section describes the drivers behind the problems identified in the subsection above.

3.1.1.3.1 Underdeveloped ecosystem

A first, fundamental element causing limited B2G data-sharing is the **limited awareness** among data holders, data re-users, and the general public, of the potential and value of private sector data for the public good.[228] And when there is demand it is often poorly defined or widely fragmented making it hard for the supply side to respond meaningfully.

Related to this, **there is currently no clear well-defined community of practice and expertise**: supply and demand of data are "*often widely dispersed*", resulting in a situation where "*those who need data do not know where to find it, and those who release data do not know how to effectively target it at those who can most effectively use it*".[229] An OECD report identifies reinforcing trust and empowering users via stakeholder engagement and community building as one of three key challenges to be addressed in order to facilitate, encourage and enhance data access and sharing "*for the benefit of all*".[230]

Moreover, the absence of industry-specific guidelines and protocols guiding collaboration, as well as the difficulty in identifying partners or opportunities to collaborate, are barriers to the **scaling-up** of existing B2G projects.[231]

[228] Hidalgo-Sanchis, P., Verhulst, G, S., Opinion: The promises – and challenges – of data collaboratives for the SDG's. See: https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082

[229] A. Young & S. G. Verhulst, Data Collaboratives. See: https://link.springer.com/referenceworkentry/10.1007/978-3-030-13895-0_92-1

[230] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd-ilibrary.org/sites/276aaca8-en/index.html?itemId=/content/publication/276aaca8-en

[231] Hidalgo-Sanchis, P., Verhulst, G, S., Opinion: The promises – and challenges – of data collaboratives for the SDG's. See: https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082

3.1.1.3.2 Legislative uncertainty

Given this unsatisfactory situation, a **number of countries** have already **introduced legislation** to ensure access to data for reasons of public interest. France, for example, included provisions enabling access to certain private sector data in its **Law for a Digital Republic,**[232] while the **Finnish Forest Act** places obligations on forest owners to share data on the management of the forest.[233] Meanwhile at **EU level**, the Green Claims Initiative will aim to improve the quality of claims on the environmental performance of companies and products to allow market actors to take greener decisions[234] EU-level **sectorial initiatives** are moving forward, with the Payment Services Directive[235] establishing data sharing mechanisms within its specific domains.

This context of national and sectorial initiatives creates a **risk of fragmentation across multiple dimensions**, including the type of data that can be collected, the manner it should be collected in, and the purpose for which this can be done.

Defining the 'public interest'

The concept of "public interest" occurs throughout EU law although there is no single legal definition and its interpretation may therefore differ across Member States. For instance, the abovementioned French Law for a Digital Republic does not define the term. The High-Level Expert Group on Business-to-Government Data Sharing considered that "while 'public interest' broadly refers to the welfare of individuals in society, its exact boundaries remain largely undefined, being heavily dependent on socioeconomic, cultural and historical factors".

Of relevance for the Data Strategy is the General Data Protection Regulation (GDPR), according to which 'processing … necessary for the performance of a task carried out in the public interest' may be a valid legal basis for the processing of personal data, and that this must have 'a basis in Union or Member State law' (Recital 45 and Article 6). The GDPR furthermore provides for derogations to some of its provisions on the basis of consideration for the public interest – i.e. in cases where exercising a data protection right would interfere disproportionality with the public interest (e.g. article 17.3, article 20.3, article 23.1.e, article 49.1.d, article 89.1, and others).

The 2004 White Paper on services of general interest states that:

- "Services of general economic interest", mentioned in, but not defined by the Treaties or secondary legislation, refers to economic services "which the Member States or the Community subject to specific public service obligations by virtue of a general interest criterion". A few examples are given, but that criterion is left undefined.
- "Services of general interest" covers "both market and non-market services which the public authorities class as being of general interest and subject to specific public service obligations."

Thus, there is no universal legal definition of the 'public interest', so any framework for business to government data sharing would require clarity, consistent with existing laws, notably the GDPR, of

[232] Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. Available at https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id

[233] High-Level Expert Group on Business-to-Government Data Sharing (2020), Towards a European strategy on business-to-government data sharing for the public interest. Available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

[234] https://ec.europa.eu/environment/eussd/smgp/initiative_on_green_claims.htm

[235] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015L2366

the basis for compliance with a request or to challenge the request. Clarity is also necessary for ensuring accountability and transparency.

Confusion also exist as to the lawfulness of the **re-use of personal data** collected by the private sector (with the GDPR and other rules). Lack of clear guidance similar to that provided by the European Data Protection Board for the re-use of data for Covid-19[236], may hamper public private engagement.

Similar fragmentation and confusion exist regarding the **contractual and data licensing** aspects of B2G data access and reuse. The lack of standardised templates for data access agreements (including data use; transfer and access agreements) leads to increased transaction costs for establishing data collaboratives[237]. Similarly, existing data licensing regimes often fail to anticipate the complexities and varieties of B2G data sharing.

3.1.1.3.3 Current mechanisms to acquire data are not fit

The public procurement of data as such (whether raw, pre-processed or processed data, or data-driven insights) **is often not the most appropriate nor cost-efficient approach to acquire data.**[238] Firstly, in several cases, access to numerous datasets from different companies over time is required in order to obtain meaningful value and to ensure unbiased public policies. Multiple parallel procurement procedures have a negative impact on the timely access to the data and on the costs of each individual procured dataset. Secondly, the cost of providing the data can oftentimes not be estimated *ex ante*, which prevents data holders and the public sector from fixing a fair compensation for providing the data. Since data markets are currently underdeveloped, there is no guarantee that a significant number of qualified and interested contractors would respond to the tenders or that tenders would be of a sufficient size to justify the resource-intensive procurement procedures. The lack of an active market for the type of data that public-sector bodies may seek to reuse makes public procurement a challenge. Lastly, if the data sought is unique or very specific and there is only one provider, the holder of the data may demand unreasonably high price as a condition for providing the data. If the data provider denies the access to the data, this could result in social welfare losses.

Due to these reasons, data procurement seems to be an ineffective and inefficient mechanism to enable B2G data access and reuse practices for public interest purposes.

3.1.1.3.4 Disincentives for private sector organisations

This fragmentation will potentially leave organisations **uncertain of their obligations** and increasing the burden on them. **Additional factors of uncertainty and perceived risks** include potential data leaks benefiting competitors, penalties from regulators, and reputation loss should customers be wary of B2G data sharing.[239] Other anticipated risks may relate, for instance, to the technical implementation of the data sharing agreement or to security and confidentiality. Private sector firms may also **fear a negative impact** on them from sharing data with the public sector, for instance if their competitors do not share their data as well.[240]

[236] See https://edps.europa.eu/data-protection/our-work/subjects/covid-19_en

[237] See https://contractsfordatacollaboration.org/

[238] High-Level Expert Group on Business-to-Government Data Sharing (2020), Towards a European strategy on business-to-government data sharing for the public interest. Available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

[239] Hidalgo-Sanchis, P., Verhulst, G, S., Opinion: The promises – and challenges – of data collaboratives for the SDG's. See: https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082

[240] JRC, The economics of Business-to-Government data sharing. See: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf

Already, **high transaction costs** and **perceived risks** constitute an important barrier to increased B2G data sharing (in particular for smaller entities): private sector data holders anticipate costs from, *inter alia*, negotiating contractual agreements and pre-processing data to be made available to government. For instance, the commercial bank of a big-sized EU Member State mentioned during the interviews conducted for this domain, that negotiating with a public authority, to achieve the partnership, took approximately 4 months with 12 employees involved (20% of their time). Making data accessible to public authorities can also represent costs for the private sector, particularly when the conditions and requirements are distinctive for every public authority; or only benefits a single data user.

Furthermore, **absence of trust** is also a barrier to B2G data sharing, as the relationship between the private sector, civil society and governments is currently uneasy in regard of collaboration on data. This distrust also stems from **the public which may be uncomfortable or concerned** with private companies sharing their data or having their data being misused for political and other reasons.[241] In part the distrust results from a lack of engagement by government with citizens on the re-use of private data for public interest purposes.

Finally, and perhaps more importantly, the absence of a clear "**business case**" or perceived lack of profitability, that can either recover costs or provide for a return on investment and risk-taking, corporations are reluctant to invest and engage in long term B2G data sharing arrangements. Indeed, developing sustainable business models and coherent incentive mechanisms is another key challenge that needs to be overcome, according to the OECD.[242]

3.1.1.3.5 Technical barriers

Lastly, there are also **technical barriers** to increased B2G data sharing.

Capacity for the processing, analysis and use of big data – on both the supply side and the demand side – is limited. The resources (e.g. IT equipment, data analytics skills, and capacity to anonymise data) vary widely across different governments and companies: while some can invest in IT in-house capabilities, others lack behind.[243] The OECD highlights the need for investment in data-related skills and infrastructures.[244] In particular, advances need to be made with regards to:

- Data preparation, including new ways for de-identification, anonymization, aggregation and cleaning;
- Data transmission, including secure cloud or safe sandbox modalities; and the emergence of federated data systems;
- Interoperable data standards, including data portability standards; and
- Data access and audit technologies.

3.1.1.4 The effects of the problem

As a result of the lack of clear procedures and structures for Business-to-Government data access and reuse, both the private and the public sectors are **not reaping the potential benefits of B2G data sharing.** For instance, delivery of public services in a more flexible manner, increased efficiency

[241] Hidalgo-Sanchis, P., Verhulst, G, S., Opinion: The promises – and challenges – of data collaboratives for the SDG's. See: https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082

[242] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd-ilibrary.org/sites/276aaca8-en/index.html?itemId=/content/publication/276aaca8-en

[243] Hidalgo-Sanchis, P., Verhulst, G, S., Opinion: The promises – and challenges – of data collaboratives for the SDG's. See: https://www.devex.com/news/opinion-the-promises-and-challenges-of-data-collaboratives-for-the-sdgs-94082

[244] OECD, Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

and innovation, better decision-making and policy making, are all benefits that at the moment are not being reaped.

The problem also affects areas such as knowledge creation or response to cross-border challenges in the EU, which at the moment are suboptimal or constrained by the lack of available data. Moreover, different rules and administrative practices result also in costs and burdens for both the private and public sectors since the partnerships are done in an ad-hoc basis, sometimes translating into building a complex and costly infrastructure for a specific use case.

Consequently**, society does not reap the benefits of B2G data sharing**, which could be better policies and more efficient public service delivery; faster and more targeted emergency/crisis prevention and response; increased transparency and more citizen involvement in societal challenges; among others.

3.1.2 Measures supporting citizen empowerment (‘human-centric data economy’)

3.1.2.1 Background

3.1.2.1.1 Context

In terms of **market**, this can be divided alongside three categories, although there are significant overlaps.

Smart appliances are typically all household items that today include software and sensors. Some producers such as Samsung have built their own platform, SmartThings. Other companies, such as Schneider, Siemens and Johnson are focussing more on smart buildings, the infrastructural elements such as heating and energy consumption. The smart home market is expected to grow from 78 billion to 135 billion dollars over the next five years.[245]

Fitness trackers gather data from our physical activity, including positioning and health data such as heartbeat. They are a different market segment. They are typically overlapping with smartwatches under the wearables’ category. There are many different devices, and they typically connect with an app and with the fitness platforms, namely Apple Health and Google Fit. In addition, there are data sharing platforms such as Strava, which gather data from such devices. However, the level of data sharing and integration between the device and platforms varies. When it comes to health monitoring systems, there is a clear overlap with the smart home market.

Voice assistants are yet another product, but it is overlapping with the previous two. Alexa, Google Assistant and Apple Siri are an interface and a platform that works with a huge variety of devices: voice assistants can be self-standing, such as Amazon Alexa, and/or integrated in third party speakers, smartphones and smartwatches. As such, they can be considered as a platform, as they connect with smart appliances and typically can be the main interface, so that home appliances often work best with one ecosystem (Google, Apple and Amazon). In many cases voice assistants are considered as the platform for smart home.

These differences and overlaps between the categories are important. For instant, voice assistants would benefit from data portability of smart appliances as a re-user, but would have to provide data as a data holder to other service providers.

A recent market report provides the following comprehensive market definition.

Category	Product

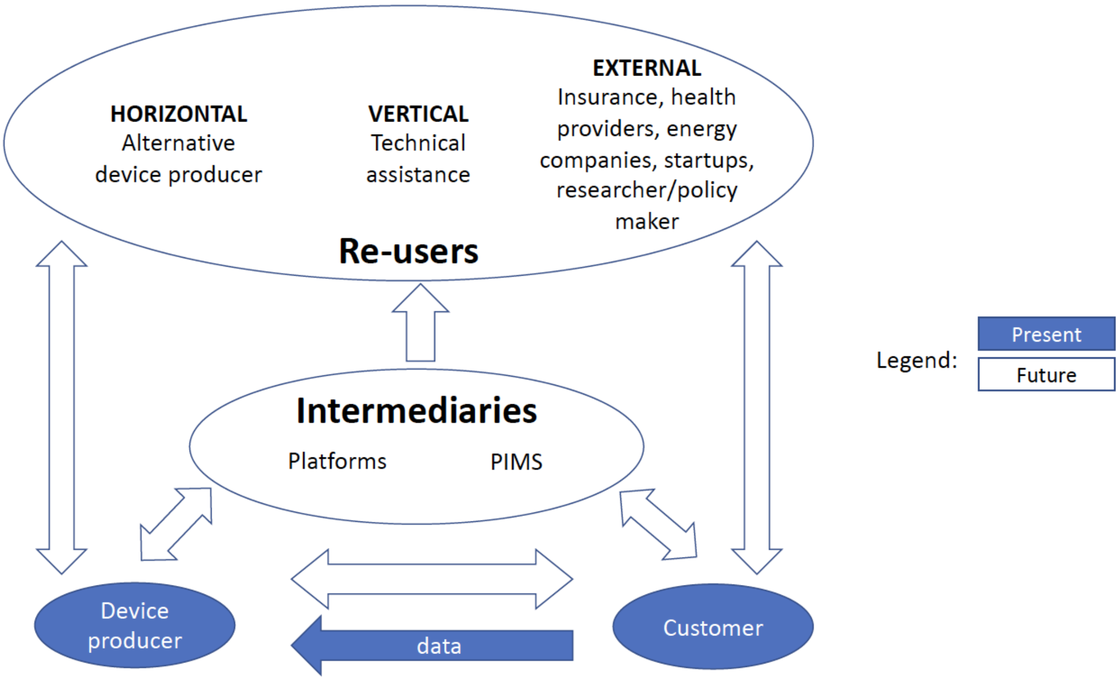
[245] Markets and Markets, "Smart Home Market with COVID-19 Impact Analysis by Product (Lighting Control, Security & Access Control, HVAC Control, Entertainment, Home Healthcare), Software & Services (Proactive, Behavioural), and Region - Global Forecast to 2025," *in Smart Home Market*, June 2020.

Lighting Control	Relays Occupancy Sensors Daylight Sensors Timers Dimmers Switches Lighting Control Accessories and Other Products
Security and Access Control	Video Surveillance Systems Hardware Cameras Storage Devices Monitors Servers Accessories and Others Software/Video Analytics Access Control Systems Biometric Access Control Facial Recognition Iris Recognition Fingerprint Recognition Others (Include Vein Recognition and Voice and Speech Recognition) Non-Biometric Access Control
HVAC Control	Smart Thermostats Sensors Used in HVAC Applications Control Valves Heating and Cooling Coils Dampers Actuators Pumps & Fans Smart Vents
Entertainment and Other Controls	Entertainment Control Products Audio, Volume, & Multimedia Controls Home Theater System Controls Touchscreens and Keypads Other Controls Smart Meters Smart Plugs Smart Hubs Smart Locks Smoke Detectors
Home Healthcare	Health Status Monitors Physical Activity Monitors
Smart Kitchen	Smart Coffee Makers Smart Kettles Smart Dish Washers

	Smart Ovens Smart Cooktops Smart Cookers
Home Appliances	Smart Refrigerators Smart Washers Smart Water Heaters Smart Vacuum Cleaners
Smart Furniture	Smart Tables Smart Desks Smart Stools & Benches Smart Sofas Smart Chairs

3.1.2.1.2 Ecosystem

Currently, the typical data flow goes from the user to the device manufacturer. In the future, ideal scenario of large-scale adoption of data portability for this sector, data will be accessible (upon consent of the consumer) to a much wider set of stakeholders, as illustrated in the figure below.



But who exactly are these other players? The table below provides an indicative list of the main stakeholders identified for each Domain.

Table 58 - Stakeholder scope (data value chain mapping)

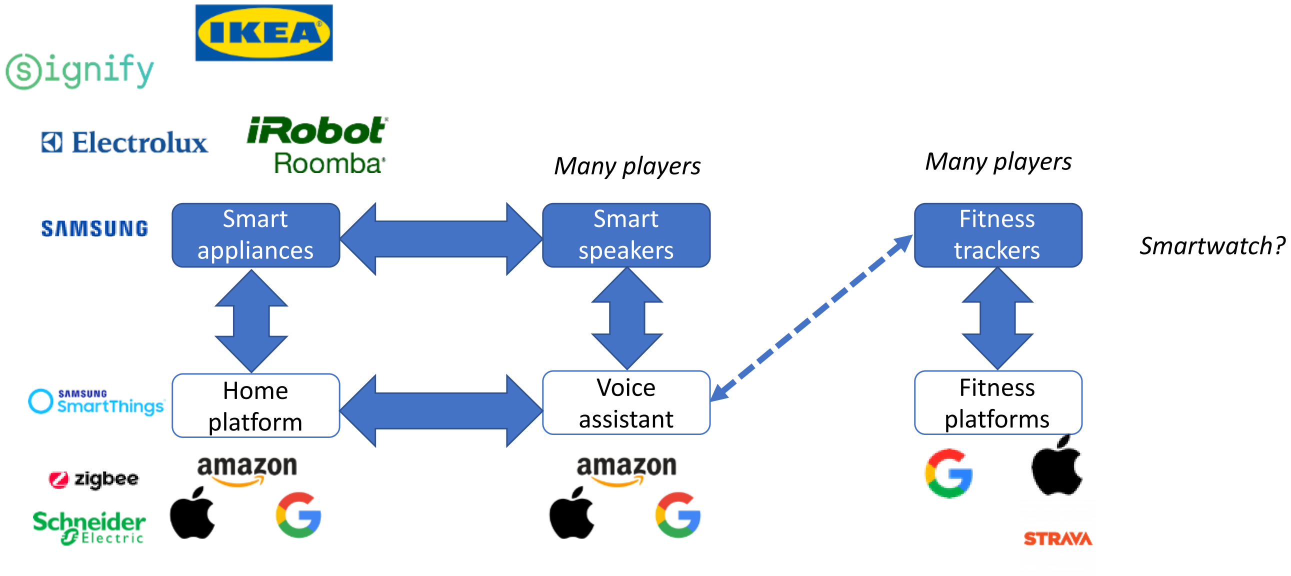
Domain	Data holder	Data (co-) producers	Data re-user (whole dataset)	Data re-user (individual	Intermediaries	Personal data?

				data portability)		
Measures supporting citizen empowerment ('human-centric data economy') – fitness	Device producer	Owner of device (individual)	Researchers Platforms (Google Apple Strava)	App developers Insurance companies Health providers Other device producers	Platforms (Google Apple Strava) PIMS	Y
Smart home	Producer (Electrolux) Energy companies	Owner of device (family)	Platforms (Google Amazon Apple Samsung IFTTT) Energy companies	Repair shop App developers Insurance companies Platforms Other device producers	Platforms (Google Amazon Apple IFTTT) Produced led platforms (Schneider, Johnson, Siemens, Samsung, Philips) PIMS	Y

The question is now what data and devices we are referring too. As we illustrate in the chart below, we can distinguish two layers: one for the hardware, and another one for software and platforms. Hardware includes smart appliances such as smart fridges, smart speakers such as Sonos, and fitness tracker such as Fitbit. There are clearly lots of players in these markets.

The other layer refers to platforms and voice assistants. There are relatively few players in these domains, notably Apple, Amazon, Google and Samsung. Voice assistants are different from smart speakers, insofar few voice assistants power a wide variety of smart speakers – for instance, Sonos works with both Alexa and Google Assistant. To complicate things further, platforms players also produce hardware devices in all three segments – from Samsung smart fridges to Google Nest speakers to Apple iWatch.

The present study will focus mostly on the upper layer, the hardware.



3.1.2.1.3 Existing models for personal data portability

Different countries are adopting measures related to data portability, and think tanks have come up with different possible solutions. The UK government has launched the "smart data review" dedicated to regulated markets (hence with a different scope from this analysis), building on the experience of the "midata" initiative in 2011. Several policy reports (listed in the next methodology section) have been published with recommendations to implement data portability and certainly the PSD2 Directive[246] represents the most ambitious regulatory provision enforcing portability.

Wider data access has been enforced in specific sectors through regulation, namely:

- Banking, through Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)[247], which mandates making account data accessible through APIs to third party payment service providers and account information service providers (with the consent of the account holder).
- Energy, through the Directive 2012/27/EU[248] on Energy efficiency, which mandates the use of intelligent metering systems (e.g. smart meters) to enhance energy saving and support the development of energy networks (smart grids). The smart grids will enable the new market actors, such as aggregators and energy service companies, to offer new types of services to consumers, enabling them to regulate their energy consumption, compare offers and switch suppliers. The Directive (EU) 2019/944[249] mandates that data shall also be made easily and securely available to final customers at no additional cost, through a standardised interface or through remote access, in order to support automated energy efficiency programmes, demand response and other services.

[246] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L2366

[247] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC

[248] Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC

[249] Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU (recast)

- Automotive, through the Regulation (EC) No 715/2007[250], which establish the rights to unrestricted and standardised access to vehicle repair and maintenance information to independent operators, in a non-discriminatory manner compared to the access granted to authorised dealers and repairers. The Regulation (EC) No 2018/858[251] (Annex X), amends the requirements with the availability of the direct vehicle data stream through the serial data port on the standardised data link connector (paragraph 6.5.1.4 of Appendix 1 of Annex 11 to UN Regulation No 83 and paragraph 4.7.3 of Annex 9B to UN Regulation No 49).

These sector specific measures can provide useful data points on the costs and benefit of each actions.

[250] Regulation (EC) No 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information

[251] Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC

Measure	Data	Timeliness	Format	Cost	Purpose	Security
Regulation (EU) 2016/679 on General Data Protection Regulation	Personal data (provided and observed)	Without undue delay and in any event within one month, except for complex requests (additional two months)	Structured, commonly used and machine-readable	Free of charge (beside exceptional cases)[252]	Protect fundamental rights and freedoms of natural persons and ensure the free movement of personal data	
Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)	Payment services data	Near real-time access to data payments	Standard format, common and secure open standards of communication, APIs included	Free of charge (for customers and registered re-users)	Service provision and innovation in the payment services market (innovative online and mobile payments, more secure payments and better consumer protection)	Strong customer authentication (SCA) Secure encryption, Licensed/ registered TPP
Directive 2019/944 on common rules for the internal market for electricity	Metering, consumption energy data, data required for customer switching, demand response and other services	Near real-time access to data about consumption (for customers)	Easily understandable harmonised format (for consumption data)	Reasonable and duly justified fees for access to data for eligible parties; free access to own data for customers	Service provision and innovation in the energy market (competitive, consumer-centred, flexible and non-discriminatory electricity markets)	
Regulation (EC) No 715/2007	Vehicle repair and maintenance information	Daily, monthly, and yearly basis data availability on manufacturers websites	Standardised format (e.g. OASIS, a common standard agreed with stakeholders), Machine readable and electronically processable datasets	Reasonable and proportionate fees for data access	Service provision and innovation (effective competition on the market; free movement of goods, freedom of establishment and freedom to provide services)	https//SSL-TLS (RFC4346) (cryptographic protocols for communications security), security certificates - ISO 20828
Singapore's Review of Personal Data Protection Act 2012 (Data Portability and Data Innovation Provisions)	User provided data, user activity data	As soon as reasonably practicable from the time of request	Structured, commonly used machine-readable format	Reasonable fee to recover the cost of providing the data portability service	Service provision and innovation	

[252] In exceptional cases, the data controller could charge the data subject a reasonable fee based on administrative costs or refuse comply with the request. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

The Regulation (EU) 2016/679 on General Data Protection Regulation (GDPR) aims ensuring the free flow of personal data and harmonising the protection of fundamental rights and freedoms of natural persons. The measure mentions data portability rights provisions for data provided by the user (both personal and observed data, from using the service). Regarding timeliness requirements, it establishes a maximum delay time of one month to comply with the request. In particular cases, this delay can be extended with another two months, with prior information of the data holder. The data holder should not bare any costs associate with the request, as the service should be free of charge. However, in some particular cases, the data controller has the possibility to apply a fee (based on administrative costs only) or refuse to fulfil the request. No specific security details are included, leaving this option at the latitude of the economic actors. The regulation targets personal (provided and observed) data, and, in terms of data formats, includes some general standards provisions such as structured and machine-readable, without other details.

In the banking sector, the PSD2 Directive targets service provision and innovation in the payment services market, aiming to encourage development of innovative online and mobile payments, more secure payments and better consumer protection. In terms of data, it narrows down to payment services data, with near real-time accessibility to payment data, and requires standard formats and common and secure open standards of communication. Data should be accessible free of charge for registered third-party providers for both payment service and account information service via APIs. The regulation focuses on two types of services - payment (initiation and execution) services and account aggregation services. When it comes to security, several provisions are included such as strong customer authentication (SCA), secure encryption, with the need of registration / licensing for third party providers.

In the energy sector, the Directive 2012/27/EU aims improving the service provision and innovation in the energy market, supporting the development of competitive, consumer-centred, flexible and non-discriminatory electricity markets. The data covered by the directive concerns the metering devices, consumption energy data, data required for customer switching, demand response and other services. It requires near real-time access to data about consumption for customers, with no charges for them. However, for other eligible parties, there is the possibility to charge reasonable and duly justified fees for data access. The regulation has no specification in what regards data security, and when it comes to data format the provisions mention only to use an easily understandable harmonised format for data consumption.

In the automotive sector, the Regulation (EC) No 715/2007 aims to improve the aftermarket services and innovation, by supporting an effective competition on the market, with free movement of goods, freedom of establishment and freedom to provide services. The data covered by the regulation concern the vehicle repair and maintenance information that should be provided in a standardised format, commonly agreed with the stakeholders, machine-readable and electronically processable. Regarding timeliness, the provisions refer to only data availability on the manufacturers' websites on daily, monthly and yearly basis, no real-time provisions included. In 2018, the Regulation (EU) No 2018/858 amends the previous regulation, bringing in several updates on the technical requirements for the data access and sharing. The data access requires specific security such as cryptographic protocols for communications security and security certificates, ISO 20828. Costs provisions mention that manufacturers can charge reasonable and proportionate fees for data access, under non-discriminatory conditions for all participants.

The Singapore Data Portability and Data Innovation Provisions targets both user provided data and user activity data, and aims enhancing both service provisions and innovation developments. Data should be available in a structured, commonly used machine-readable format. There are no specific security provisions and when it comes to timeliness, it is only mentioned that data should be available as soon as reasonably practicable from the time of request.

3.1.2.2 The problem, its magnitude and the stakeholders affected

This section identifies the problem, its causes and its effects – all graphically represented in the problem tree below.

With ubiquitous connectivity and the accelerated digitisation resulting from the COVID-19 pandemic, and the proliferation of the Internet of Things, where almost any product generates data, the data economy is growing exponentially. This includes personal data, gathered through smart home appliances such as fridges, voice assistants, and fitness trackers.

These data would potentially be very beneficial, if accessible to a wide variety of stakeholders in a standardised format, however, this rarely happens. While data portability provisions in GDPR in theory could pave the way to user centric data sharing, the reality is that data portability is rarely applied. Recent studies show that of 230 requests for data portability, only 163 were actually fulfilled.[253] But perhaps more worrying is that very few users even perform such requests: a recent survey by Sitra in four European countries found that only 21% knew of a right to data portability, against 58 % for the right to deletion of personal data and a right to know how and for what purpose their data is used. On a similar note, applications that help users manage their data, such as Personal Information Management Services (PIMS) or Personal Data Spaces (PDS) have struggled to encounter large-scale adoption.[254]

In most cases, with such devices, the data are held by the producer of the device and developer of the firmware – the Original Equipment Manufacturer –such as Electrolux, Amazon or Fitbit. But there are many differences depending on the type of product.

Data from home appliances is typically not shared – they are accessible only to the manufacturer and to the customer through proprietary services, such as in the case of Samsung Family Hub.

Data from voice assistants can be partially accessed and downloaded through dedicated export services. For instance, Google allows users to export activity through the "home assistant".

Data from fitness trackers are typically exportable and allow for easy integration with data integration services such as Strava, Apple Health and Google Fit. However, the format and granularity of exporting varies, as well as the integration with different platforms and services.

Therefore, the current models and initiatives for data portability are mainly market based. They are revolved on few platforms, with less focus on standards. Only few devices allow for portability and, in some cases, the technical aspects make the portability not effective (lack of standards, for both data and systems' interoperability, no real time access, and limited data points available).

Figure 26 - Overview of the problem

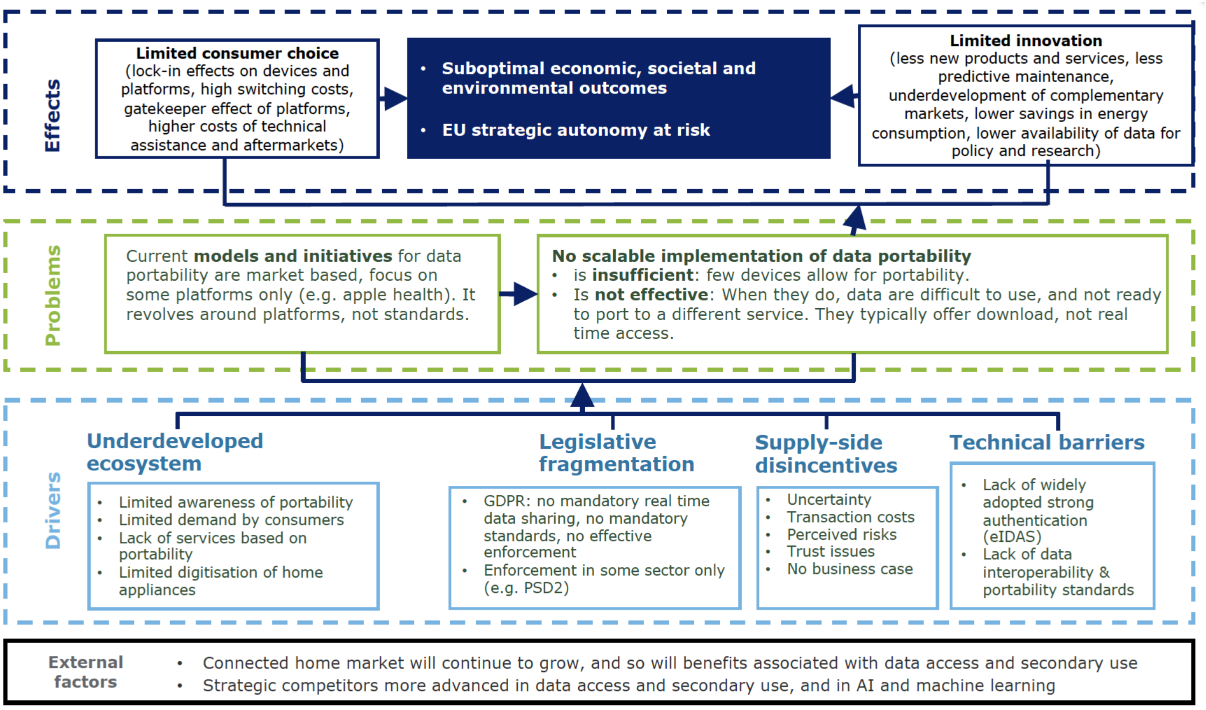
Measure	Barriers	Problem	Impact
Stimulate availability of better tools	Lack of offering of usable, scalable and largely adopted data portability tools	Lack of adoption of data portability as a way to	Lack of competition

[253] Janis Wong and Tristan Henderson, 'How Portable Is Portable?: Exercising the GDPR's Right to Data Portability', in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18* (presented at the 2018 ACM International Joint Conference and 2018 International Symposium, Singapore, Singapore: ACM Press, 2018), pp. 911–20 <https://doi.org/10.1145/3267305.3274152>.

[254] Ilves and Osimo. A Roadmap for a Fair Data Economy (Helsinki: Sitra, 2019)

New regulatory measures to enhance article 20 GDPR	Lack of supply of data portability by smart devices producers	promote data reuse	Lack of innovation

Figure 27 - Problem tree



Effects

Limited consumer choice (lock-in effects on devices and platforms, high switching costs, gatekeeper effect of platforms, higher costs of technical assistance and aftermarkets)

- **Suboptimal economic, societal and environmental outcomes**
- **EU strategic autonomy at risk**

Limited innovation (less new products and services, less predictive maintenance, underdevelopment of complementary markets, lower savings in energy consumption, lower availability of data for policy and research)

Problems

Current **models and initiatives** for data portability are market based, focus on some platforms only (e.g. apple health). It revolves around platforms, not standards.

No scalable implementation of data portability
- is **insufficient**: few devices allow for portability.
- Is **not effective**: When they do, data are difficult to use, and not ready to port to a different service. They typically offer download, not real time access.

Drivers

Underdeveloped ecosystem
- Limited awareness of portability
- Limited demand by consumers
- Lack of services based on portability
- Limited digitisation of home appliances

Legislative fragmentation
- GDPR: no mandatory real time data sharing, no mandatory standards, no effective enforcement
- Enforcement in some sector only (e.g. PSD2)

Supply-side disincentives
- Uncertainty
- Transaction costs
- Perceived risks
- Trust issues
- No business case

Technical barriers
- Lack of widely adopted strong authentication (eIDAS)
- Lack of data interoperability & portability standards

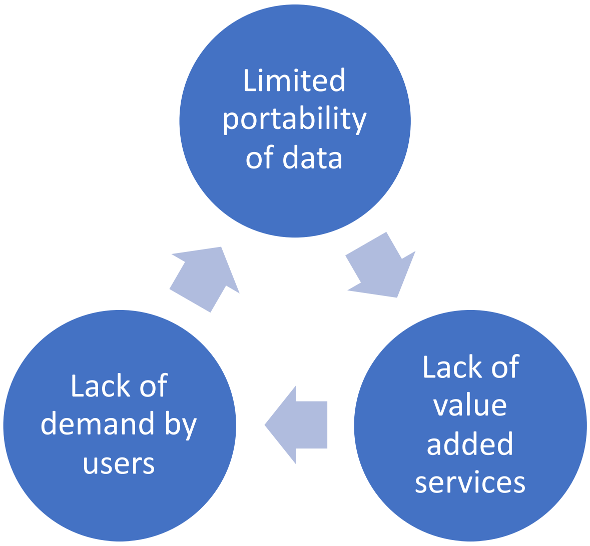
External factors
- Connected home market will continue to grow, and so will benefits associated with data access and secondary use
- Strategic competitors more advanced in data access and secondary use, and in AI and machine learning

3.1.2.3 The causes of the problem

GDPR allows data portability for data provided by the users, including both data actively and knowingly provided (such are address) and observed data. Broadly speaking, IoT home devices deal with observed data[255] and rely on consent or contract for the lawfulness of the processing of personal data therefore fall under GDPR Article 20.

However, this right is not yet widely used. There is no requirement for real time data to be shared, and consumers are less aware of the lost opportunity because of the absence of value-added services. It’s a vicious circle where lack of data leads to no services which leads to limited demand by users.

[255] https://gdpr-info.eu/art-20-gdpr/



In this sense, sector specific regulations aim to address some of these aspects by stipulating specific access rights. In the banking sector, there is the Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)[256] that includes provisions for data access/sharing via APIs to third party providers. In the energy sector, there are Directive 2012/27/EU[257] Energy efficiency and Directive (EU) 2019/944[258] includes some provisions regarding data access, standards and security issues. And in the automotive sector, the Regulation (EC) No 715/2007[259] and the Regulation (EC) No 2018/858[260] (Annex X) includes provisions related to data sharing and data standards. However, these regulations remain sector specific and the development of the cross-sector interoperability lags further behind.

In addition, observed use of existing limited data portability is also due to genuine concerns about data protection issues – data portability must be secure.[261] Thus, even if the customers are aware of the data portability provisions, they are often reluctant to share their own data due to lack of trust towards the producer of the devices or the data intermediaries. Security risks, uncertainty in data management and transaction costs remain important aspects that impacts data portability. Therefore, with low levels of data availability and high level of customers' concern over data protection, the services built based on portability remain few and far between.

Limited digitisation of the home appliances makes the data portability less compelling and valuable. And when devices' digitisation does not seem to be an issue, it is the data standards and systems' interoperability that prevent portability to become functional.[262] Additionally, the lack of widely

[256] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC

[257] Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC

[258] Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU (recast)

[259] Regulation (EC) No 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information

[260] Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC

[261] Stakeholder's interview.

[262] Stakeholder's interview.

adopted strong authentication systems together with customers' concerns about data protection issues constitutes important factors that also hamper the data portability developments.[263]

3.1.2.4 The effects of the problem

The lack of effective data portability leads to **limited choices for consumers** when it comes to products and services. They risk often to remain locked-in on specific devices and platforms, and the high switching costs make it difficult to break out.

When it comes to aftermarkets and assistance, the **limited offers also increase the costs for customers and often increase waste**, as it is cheaper to buy a new device than to repair the old one. This is also a consequence of the difficulties the independent repair shop to repair home appliances have to join the markets and increase the competition. In the end, the complementary markets remain underdeveloped as new players often cannot join due to limited access, resources and high costs.

Lack of data availability also impacts the level of innovation, as new products and services (such as predictive maintenance) remain difficult to develop, and the possibility to reuse data for health purposes remains limited.

Overall, the **lack of effective portability negatively impacts the EU strategic autonomy and the economic, social and environmental outcomes remain suboptimal.**

3.1.3 Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

3.1.3.1 Background

3.1.3.1.1 Context

The approaches and challenges for access and reuse of co-produced and business data can vary according to the sector and position on the value chain of the actors concerned. This is the reason why it is important to enshrine a sectorial dimension in the methodology put forward for this domain in order to cover different scenarios and situations and the impact that policy measures would have on those. This study focuses on the following four sectors:

- **Agriculture**: this sector is characterised by a high number of players involved in B2B co-generated data sharing (including apart from farmers, both upstream market players, such as seeds, fertilisers, machinery manufacturers, as well as downstream market players, such as food production and processing companies) and by a recent important development with regard to this topic (the above mentioned code of conduct signed in 2018 by 11 key stakeholders of the sector).
- **Construction**: Even though construction sector is not the fastest sector in IoT adoption, recent statistics show that it's only a matter of time before IoT in construction becomes a must-have technology. According to studies this sector represents approximately 13% of the global GDP.[264]
- **Manufacturing:** This sector presents interesting observations as Industrial IoT co-generated data, used inside the factory and within the supply chain, enable manufacturers to scale up different equipment capable of distant observing and servicing, as well as to have a proper estimation of customers' needs.
- **Transport and logistics:** access to data along the transport and logistic value chains is becoming an increasingly burning topic also due to the huge digitalisation efforts carried out by this sector in the past decade[265]. Access to third parties' data is now the norm but legal and

[263] Stakeholder's interview.
[264] IoT in construction industry, Digiteum, https://www.digiteum.com/iot-construction-industry/
[265] See for instance the activities of the Digital Transport and Logistics Forum organised by the European Commission, https://www.dtlf.eu/

organisational solutions are still found on case-by-case basis. This sector alone constitutes up to 5,2% of the EU GDP[266].

According to recently published statistics, the total installed base of IoT connected devices worldwide is projected to amount to 30.9 billion units by 2025, a sharp jump from the 13.8 billion units that are expected in 2021.[267] The real value of the IoT is linked to the data and their insights. IoT is unlocking significant value for companies by enabling smart factories and connected supply chains as well as the ability to monitor products and deliver new services. In asset-heavy industries, the proliferation of IoT data is fundamentally shifting the customer value proposition from goods to services, and this shift is leading companies to adopt new business models that require new capabilities. The majority of IoT solutions today are built around internal applications such as predictive maintenance, fleet management, factory optimization, supply chain automation, and improved product design.[268]

A recent study explains that by collaborating with new business partners, including industry incumbents and players in other sectors, companies can form new data ecosystems. These ecosystems give their participants access to valuable collective data assets as well as the capabilities and domain expertise necessary to develop the assets into new data-driven products and services. Data ecosystems will play a critical role in defining the future of competition in many B2B industries. They enable companies to build data businesses, which are valuable not only because they generate high-margin recurring revenue streams but also because they create competitive advantage. New data-driven products and services deliver unique value propositions that extend beyond a company's traditional hardware products, deepening customer relationships and raising barriers to entry. They also build highly defensible positions, thanks to natural monopolies rooted in economies of scale and scope (similar to monopolies based on claims on IP protection or *de facto* control over information by smart machinery manufacturers). Companies that secure advantaged positions in data ecosystems will generate significant value and competitive advantage across their entire business, including their traditional hardware offerings.[269]

In recent years, industrial OEMs are increasing their focus on aftermarket services (i.e. the provision of parts, repair, maintenance, and digital services for the equipment they sold). When exploring aftermarket value pools, industrial OEMs are often tempted to prioritize data-driven advanced services enabled by digital innovation and the IoT.[270] The definition of an aftermarket or secondary market is linked to goods or services that are complements to a long-lasting primary product and that are typically bought after acquisition of the primary product. Aftermarkets are a common feature in innovative market environments. This may lead to several legal **questions around competition**

[266]https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Transportation_and_storage_statistics_-_NACE_Rev._2#Structural_profile

[267] Statista, (2021), IoT and non-IoT connections worldwide 2010-2025, https://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/

[268] Quoting: Russo, M., Albert M., (2018), How IoT Data Ecosystems Will Transform B2B Competition, BCG, https://www.bcg.com/publications/2018/how-internet-of-things-iot-data-ecosystems-transform-b2b-competition

[269] Russo, M., Albert M., (2018), How IoT Data Ecosystems Will Transform B2B Competition, BCG, https://www.bcg.com/publications/2018/how-internet-of-things-iot-data-ecosystems-transform-b2b-competition

[270] wold A., Brotschi A., Forsgren M., Kervazo F., Lavandier H., Xing J. (2017) Industrial aftermarket services: Growing the core, McKinsey &Company https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/Industrial%20aftermarket%20services%20Growing%20the%20core/Industrial-aftermarket-services-Growing-the-core-final.pdf?shouldIndex=false

law when innovators keep these lucrative aftermarkets to themselves. Regularly, this is achieved through intellectual property (IP) protection, leading to what is called a proprietary aftermarket.[271]

Advances of IoT brings opportunities for aftermarket services around IoT-connected products, becoming more and more attractive for innovators and investors. In this context, further opportunities and business models are emerging linked to data analytics services for performance optimisation, availability and security of IoT-connected products. Opportunities for such aftermarket services can be found in many industry sectors including, among others, power, transportation, construction, agriculture, oil and gas, healthcare and aerospace. One analysis done across 30 industries showed that average profit margins for aftermarket services were 25%, compared to 10% for new equipment.[272] However, in order to be able to consider the importance of aftermarkets in relation to the primary market, different factors should be taken into consideration, including product lifetime and the average annual services price. These factors significantly vary from sector to sector. When comparing for example the transport with other industry sectors, an analysis showed that the aftermarket lifetime value associated with heavy-duty trucks, which is estimated to be 30% of sales price is much lower than other sectors such as power-generation equipment, with aftermarket lifetime value estimated to 75% of sales price.[273]

A study focused on the construction equipment industry[274] confirms that European OEMs see value-creating opportunities, among others, in aftermarket services. The European market has traditionally been complex, with a variety of local specifications, and fragmented, with OEMs focused on national markets. Although the market has been consolidating for years as demand and standards globalized, Europe remains more complex and fragmented than other markets. Today more than 450 OEM groups[275] are active in Europe, and even though almost 75% have revenues of less than €100 million, across the board, they have more than 40% of their business in other continents. Furthermore, the European demand situation has been challenging for several years because construction activity is low. According to a survey conducted in the frame of this study, aftermarket services have been highlighted by European OEMs as the second most important trend creating primary value opportunities (with a percentage of 56% of the respondents). These players also see emerging-market competition as a key threat and will not be able to compete with them on price in most cases.[276]

Where digital ecosystems that evolve in the IoT are significantly driven by user data, the **lock-in effect** for users may be particularly strong, and it may extend to a broad variety of services and hence aftermarkets. Also, user data can provide a competitive advantage not only in markets for

[271] Robertson V., (2020), Competition Law's Innovation Factor: The Relevant Market in Dynamic Contexts in the EU and the US, Bloomsbury Collections

[272] Nistor G., (2018) Aftermarket services are an IoT opportunity, Deutsche Telekom https://www.b2b-europe.telekom.com/blog/2018/01/22/aftermarket-services-are-an-iot-opportunity

[273] Ambadipudi A., Brotschi A., Forsgren M., Kervazo F., Lavandier H., Xing J. (2017) Industrial aftermarket services: Growing the core, McKinsey &Company, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/Industrial%20aftermarket%20services%20Growing%20the%20core/Industrial-aftermarket-services-Growing-the-core-final.pdf?shouldIndex=false

[274] Sjodim E., Granskog A., Guttman B., 2016, Toward a customer-centric construction-equipment industry, McKinsey &Company, https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/toward-a-customer-centric-construction-equipment-industry

[275] According to the Committee for European Construction Equipment - CECE, The sector counts around 1200 companies that employ about 300,000 people directly and indirectly. Their annual revenues amount to ca. 40 billion euros, https://www.cece.eu/about

[276] Sjodim E., Granskog A., Guttman B., 2016, Reengineering construction equipment: from operations focused to customer centric, McKinsey &Company, https://www.mckinsey.com/~/media/mckinsey/industries/automotive%20and%20assembly/our%20insights/toward%20a%20customer%20centric%20construction-equipment%20industry/reengineering-construction-equipment.pdf

secondary goods, but also at the time of the replacement of the primary product.[277] In the transport or construction sector, according to an article of the "IoT business magazine", a smart machine generates large volumes of data, but within a closed ecosystem which only the manufacturer can access to deliver services such as telematics. That means any third-party solutions providers may have to install additional hardware if they want to provide services. The interests of original equipment manufacturers (OEMs) and purchasers are often not aligned: the OEM wants to maintain a **closed ecosystem** where they alone can provide services, while the owner wants freedom to choose their service providers. Closed ecosystems point to the importance to OEMs of maintaining long-term relationships with equipment buyers, especially given the revenue and profits generated by maintenance and spare parts.[278] Furthermore, the US-based Caterpillar, ranked as world's number one construction equipment OEM, can be used an example of construction OEM generating additional revenues from aftermarket services. A study of 2018 explains that Caterpillar's suite of digital offerings, known as "Cat Connect", includes services such as asset health monitoring and automated grade assist. These services not only add value for customers, they also **increase switching costs**, as other brands of equipment are not integrated into Caterpillar's product and service ecosystem.[279] According to the stakeholders interviewed for this study, the AEMP Telematics Standard in the construction equipment industry, makes the playing field is quite equal, because every OEM has access to more or less the same level of information from other OEMs. The challenge arises from the fact that the level of this information is not very deep, and most OEMs are not willing to provide any more information than what is foreseen by the AEMP Standard.

Taking that even further, even the distinction of what constitutes an open or closed eco-system might not be easy sometimes. Using as an example heavy commercial vehicles in the transport sector - which seems to be more mature compared to other sectors examined in this study as far as B2B data sharing is concerned-, in Europe most OEMs comply with an open system, the Fleet Management Standard (FMS). However, according to the managing director of a European telematics device manufacturer and developer "in order to access onboard data a fleet owner may have to purchase a gateway that can cost as much as €500 per vehicle, which may be more expensive than simply installing additional third-party sensors to use in a telematics system."[280] Similarly, in manufacturing sector, IoT technology is allowing OEMs to take advantage of the aftermarket segment rather than ceding the parts and service market to other suppliers. Visibility provided by IoT connectivity that allows OEMs to see all the way to the factory floor is invaluable in identifying trends to capture further parts sales and to develop new services at higher margins.[281]

In the agricultural sector, a recently published JRC technical report of the European Commission identified similar competition problems and market failures linked to **monopolistic data lock-ins** in the sector.[282] The authors of the report explain that "agricultural machinery manufacturers can design the machine in such a way that they have exclusive access to sensor data and actuators inputs. Once a farmer buys a particular machine, he is locked into the data channels controlled by

[277] Quoting: Schweitzer H., & Robert Welker R., (2019), Competition Policy for the Digital Era, http://robertwelker.bplaced.net/Schweitzer-Welker-CPI.pdf

[278] Overdahl S., (2020), IoT and Data Ownership: Your Device your Data?, The IoT Business Magazine, https://www.smart-industry.net/iot-and-data-ownership-your-device-your-data/

[279] Russo, M., Albert M., (2018), How IoT Data Ecosystems Will Transform B2B Competition, BCG, https://www.bcg.com/publications/2018/how-internet-of-things-iot-data-ecosystems-transform-b2b-competition

[280] Overdahl S., (2020), IoT and Data Ownership: Your Device your Data?, The IoT Business Magazine, https://www.smart-industry.net/iot-and-data-ownership-your-device-your-data/

[281] Quoting: How Manufacturing Data is Affecting OEMs, https://www.machinedesign.com/automation-iiot/article/21837613/how-manufacturing-data-is-affecting-oems

[282] Can A., Bertin M., (2020) Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU, JRC European Commission, https://ec.europa.eu/jrc/sites/jrcsh/files/jrc121337.pdf

the machine manufacturer. The manufacturer may use this monopolistic position in upstream data collection, or in access to the downstream implementation of data-driven agronomic services, to leverage his position in downstream services markets. The manufacturer's exclusive control over the data channels, combined with the lack of interoperability between data formats and devices from different manufacturers, contributes to this monopolistic market structure. Farmers cannot multi-home between different data service providers. Agricultural machine producers and service providers deliberately segment the data standards in order to increase switching costs. It creates a lock-in effect that makes farmers dependent on manufacturers/service providers and possible monopolistic pricing of their services". Therefore, the lack of interoperability and common data protocols, formats and standards leads to switching platforms-related difficulties and costs. This means that, even if even in theory farmers keep the "ownership" of the industrial non-personal data and data portability is contractually allowed, in practice this might become ineffective from a technical point of view, due to the incompatibility of such data with other platforms.[283]

Similarly, an OECD study of 2020 on issues around data governance in the digital transformation of agriculture identifies issues around the choice in the servicing of farm machinery ('right to repair') and data portability. [284] The study highlights that there is often a lack of clarity on whether farmers are able to transmit data generated by a service provider on their farm to other service providers. Due to the loss of historical data when changing machinery brand or service supplier, farmer's choice of equipment and service provider as well as the possibility of switching provider, is reduced, as the accuracy in services requiring as an input historical data is limited. Additionally, another issue according to the study is the ability of farmers to access the data and software needed to repair their digital farm equipment. OEMs' software programs, terms of use of the technology contracts that accompany the software as well as digital locks (or technological protection measures) to protect their rights in software and intellectual property rights prevent farmers from being able to access the software for the purposes of repair.[285]

Even though the automotive sector is not the focus of this study, this industry sector also presents interesting observations related to competition and innovation situations in aftermarkets. A working paper focused on the example of connected cars explains that "through controlling the access to the data of connected devices as well as to the device itself, the manufacturer of a device gets into the position of a monopolistic gatekeeper to the entire ecosystem of services and products that can be offered through or in combination with this device. Due to the investment in the connected device (and other sunk costs) the users can have large switching costs (lock-in). Such an exclusive control of the access to the data of the connected device and/or the technical access to the device can be used by the manufacturer of the device (as a primary product) for foreclosing all in-dependent providers of services on the markets for aftermarket and complementary services (secondary products), as far as such an access is necessary for providing these services and entering these markets ("essential resources"). Therefore the manufacturer can leverage this monopolistic gatekeeper position to all markets for those services which depend on this access, and therefore can control these markets."[286] Furthermore, exclusive de facto control of data/access to connected car

[283] Can A., Bertin M., (2020) Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU, JRC European Commission, https://ec.europa.eu/jrc/sites/jrcsh/files/jrc121337.pdf

[284] Jouanjean, M., et al. (2020), "Issues around data governance in the digital transformation of agriculture: The farmers' perspective", OECD Food, Agriculture and Fisheries Papers, No. 146, OECD Publishing, Paris

[285] Jouanjean, M., et al. (2020), "Issues around data governance in the digital transformation of agriculture: The farmers' perspective", OECD Food, Agriculture and Fisheries Papers, No. 146, OECD Publishing, Paris

[286] Quoting: : Kerber, Wolfgang (2019) : Data-sharing in IoT ecosytems from a competition law perspective: The example of connected cars, MAGKS Joint Discussion Paper Series in Economics, No. 21-2019, Philipps-University Marburg, School of Business and Economics, Marburg, https://www.econstor.eu/bitstream/10419/204816/1/1676213953.pdf

can lead to hold-up situations, high access prices, discrimination, and finally **limited innovation** for products/services offered in connected car.[287] Literature sources show that the business impact of automotive aftermarkets is considerable as aftersales generate 20% of revenue, and ~50% of profits for OEMs.[288]

On the other side, despite the above mentioned findings which show monopolistic issues and competitive advantage of machinery manufacturers in aftermarkets, another study of 2019, focused on the US market, shows that construction and agriculture OEMs are not flourishing in aftermarket services**,** compared to other sectors including automotive, industrial services, and aerospace —their penetration rates are far below the 75% seen in other industry sectors such as data storage and wind turbines.[289] This was confirmed by the machinery manufacturers interviewed for this study explaining that, as far as data sharing is concerned, the relevance between agricultural aftermarkets and other industries is low, as in agriculture, there is no sizable independent aftermarket distribution chain, as is the case for example in the automotive sector even if we talk about connected vehicles in both cases.

Finally, it should be mentioned that a similar characteristic between construction and agriculture is that both industry sectors present **low levels of digitalization and associated productivity gains.** The low productivity and digitalization levels in the construction equipment industry was highlighted as well by the stakeholders of the sector interviewed for this study. Additionally, according to the agricultural machinery manufacturers interviewed, approximately only 10% of farmers are currently using agricultural software at the moment. This is being confirmed also by studies that have looked at the state of digitalization in sectors across the US economy and found a large and growing gap between sectors, with construction and agriculture being the least digitalized.[290] The question that arises is whether this **limited digitalization** of certain industry sectors is linked to trust issues due to the lack of clarity on rights over co-generated data. This assumption could apply for example for the agricultural sector, where there have been observations of increasing concerns among farmers that sharing agricultural data may not return the expected benefits for them, which is seen to be hindering some of the opportunities for agricultural data collection, sharing, and use in the sector.[291] The views of the interviewed stakeholders vary on this matter. Some of them agreed that lack of trust and legal clarity over co-generated data rights is one of the factors keeping behind certain sectors from digitization, while others did not agree with this assumption linking the lack of clarity over co-generated access and usage rights to the limited digitalization, but rather attributed this problem to conservative perceptions, to the psychological factor, as well as to workforce restrictions, which could be addressed by better manners of communication and "training" of IoT solution users.

A key challenge in relation to co-generated data is that there is currently no horizontal legislation that regulates access and use of non-personal co-generated data specifically. As a result, there is no legal framework governing the rights and obligations of data holders, co-producers, intermediaries, or aspiring re-users in relation to co-generated data in a consistent cross-sector manner. Contractual terms are therefore the predominant instrument that govern the legal interests and possibilities of

[287] Kerber W., Frank. S. (2017), Data Governance Regimes in the Digital Economy: The Example of Connected Cars, https://www.tilburguniversity.edu/sites/tiu/files/download/Tilburg_Kerber_Frank_12102017_01_2.pdf
[288] Optimizing Industry 4.0 Aftermarket Services (2020), Pega and AWS, https://www.pega.com/system/files/resources/2020-06/optimizing-industry-aftermarket-services.pdf
[289] Johnson S., Laczkowski K., Paadhi A., Sandrone P., (2019), For OEMs in the United States, the aftermarket is fertile ground, McKinsey &Company, https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/for-oems-in-the-united-states-the-aftermarket-is-fertile-ground
[290] Gandhi P., Khanna S., Ramaswamy S., (2016), Which Industries Are the Most Digital (and Why)?, Harvard Business Review, https://hbr.org/2016/04/a-chart-that-shows-which-industries-are-the-most-digital-and-why
[291] Jouanjean, M., et al. (2020), "Issues around data governance in the digital transformation of agriculture: The farmers' perspective", OECD Food, Agriculture and Fisheries Papers, No. 146, OECD Publishing, Paris

the stakeholders along the value chain. This can obviously have suboptimal outcomes from a policy perspective, since freedom of contracting enables flexibility and innovation, but in a manner that favours parties with the largest market power.

This is not to suggest that contractual freedom is unlimited, of course. Several EU level legal frameworks clearly can and do impact the rights and obligations in relation to co-generated data, even though this is not their principal or exclusive focus, and sector specific legislation can further narrow down the flexibility for contracting parties. By way of examples (non-exhaustive):

- Depending on the context, some co-generated data may contain personal data as defined under the **GDPR**, if the co-generated data would allow any party (including the participants in the value chain collectively) to link specific data elements to an identifiable natural person. In that case, the parts of the co-generated data that qualify as personal data must be processed in accordance with the terms of the GDPR, or in the likely scenario that the data is mixed in a manner that doesn't allow personal and non-personal data to be distinguished in a reasonable manner, all of the co-generated data should be processed in accordance with the terms of the GDPR. In terms of establishing rights to co-generated data, this is however not a particularly unambiguous outcome. Firstly, for some machine-generated data, it is not clear whether it is reasonably possible to link that data to a natural person, and therefore whether the data qualifies as (or contains) personal data. As a result, that the applicability of the GDPR can be uncertain. Secondly, the rights that the GDPR grants aim to safeguard the fundamental right to data protection and to support the free flow of personal data in the internal market, in a way that's conducive to promoting economic growth and competition in the EU. It is however not intended to create data access and usage rights to personal data merely because this is economically beneficial to the development of the internal market; for any such rights, a specific legal basis would need to be available under Article 6 of the GDPR. And thirdly, the rights granted in the GDPR (such as the right to data portability, which could be relevant to co-generated data) are accorded to the data subjects, i.e. the natural persons identified in the data. These are not generally the stakeholders in the value chain who could benefit from access to data. Thus, the GDPR's applicability to the co-generated data is uncertain and not geared towards resolving the policy problem examined here.
- Legislation in relation to **intellectual property rights** generally will not be immediately relevant, given that co-generated data in this study is interpreted mainly as machine-generated data. This excludes the applicability of **copyright** in virtually all cases due to the absence of originality. The application of the so-called *sui generis* **database rights** in the sense of the Database Directive 96/9/EC is more ambiguous. Based on a series of rulings from the European Court of Justice in 2004[292], *sui generis* rights do not apply to databases that are the by-products of the main activity of an organisation. This is commonly interpreted to mean that *sui generis* protections would not apply broadly to machine-generated data and IoT devices, since (and to the extent that) such data is principally a by-product of a device's principal functionality that hasn't been the object of a substantial separate investment (that is, separate from the investment made in the main product or service). However, there is no perfect consensus on this interpretation. Furthermore, even if it was universally accepted that machine-generated data could not be covered by *sui generis* rights rights on account of the CJEU case law, the exclusion would still only apply to the extent that co-generated data is generated as an increasingly important but none the less ancillary by-product of a main product or service that wasn't the subject of a substantial separate investment (machine equipment, vehicle, manufacturing device,

[292] Fixtures Marketing Ltd v. Oy Veikkaus Ab (C-46/02, 9/11/2004), Fixtures Marketing Ltd v. Svenska Spel Ab (C-338/02, 9/11/2004), British Horseracing Board Ltd v. William Hill (C-203/02, 9/11/2004), and Fixtures Marketing Ltd v. OPAP (C-444/02, 9/11/2004)

etc…). This leaves a gap in situations where data is not merely a by-product, but rather a key component or even the main outcome of using a specific product or service for which the database-maker invested substantially for obtaining, verifying or presenting the data. Thus, some ambiguity remains. In cases where *sui generis* database rights do apply, the Directive generally protects the makers of databases and lawful users, without specific consideration of co-generators of databases.

- Arguably, in specific cases co-generated data could be considered a **trade secret** under the Trade Secrets Directive (EU) 2016/943, provided that the data is secret (i.e. not generally known or readily accessible to relevant stakeholders), commercially valuable because of its secrecy, and reasonably protected to keep it a secret[293]. The qualification obviously could not apply to data that is readily accessible to a user (since that implies that it is not secret), but especially in cases where data is automatically co-generated, captured and protected by e.g. an OEM manufacturer without revealing it to a third party, a qualification as a trade secret is possible. Again however, this does not solve the issue of establishing rights and obligations to the data, since these are not addressed by the Directive. The Directive indeed lays down the rules on the protection against the unlawful acquisition, use and disclosure of trade secrets. The Directive does not create any exclusive right, as do other IP rights. This also appears from the provisions of the Directive, for instance when it states that an "*acquisition of a trade secret shall be considered lawful when the trade secret is obtained by […] observation, study, disassembly or testing of a product or object that has been made available to the public or that is lawfully in the possession of the acquirer of the information who is free from any legally valid duty to limit the acquisition of the trade secret*" (article 3.1 (b)).
- Finally, and perhaps most relevantly of course, there is a broader body of **sector specific legislation** establishing some rights to co-generated data, without however consistently naming it or addressing it as such. Vehicle Emissions Regulation No 715/2007 on type approval of motor vehicles and on **access to vehicle repair and maintenance information** (Type-Approval Regulation) regulates vehicle emissions for small passenger and commercial motor vehicles and lays down rules to ensure that independent operators have access to vehicle repair and maintenance information (**RMI**)[294]. Such data obviously falls within the scope of the concept of co-generated data, and access and usage rights are thus granted in specifically tailored scope and for specific purposes under this Regulation. Similarly, the **Electricity Directive** 2019/944 requires the deployment of smart metering systems, and supports the introduction of standards and rules for sharing such data with specific services providers, including via implementing acts. In the context of mobility and logistics, the **ITS Directive** 2010/40/EU targets the collection, access and use of static road data, dynamic road status data and traffic data, some of which can be (and increasingly is) co-generated data as well; and the Vessel Traffic Monitoring and Information System Directive 93/75/EEC (**VTMIS Directive** 2002/59/EC) similarly supports maritime data collection and sharing.

The landscape is thus fragmented, in the sense that there is no clear common perspective on rights and obligations for the various stakeholders in co-generated data ecosystems. Some legislation may apply, but the impact is generally unpredictable and uncertain, and not targeted towards resolving the policy challenges examined here.

In such an environment, the impact of contractual terms between stakeholders in the value chain is usually decisive. Based on a review of a sample set of IoT related contracts (focused on mobility, agriculture and electricity), the following trends clearly emerge:

[293] See article 2 of Directive 2016/943.

[294] As amended by Regulation (EU) 2018/858

- Virtually all contracts define **subsets of co-generated data**, with diverging rights and constraints depending on the subset. Terminology differs widely, as could be expected; but (leaving aside purely administrative data, for which not notably particularities can be observed), a distinction is often made between data that was measured or inferred in relation to a specific customer; data measured or inferred in relation to the equipment; and insights generated by a service provider in relation to the prior two categories. The objective of such distinctions is to enable separately defined usage and retention rights, where the first category of data (relating to an individual customer) benefits from stronger access and usage rights for the customer, and weaker retention and re-use rights for the service provider.
- **Ambiguity in relation to personal data** is not universally addressed by contractual terms. Data protection language is nearly universally implanted in the examined contacts, but the clauses are often phrased conditionally (*"To the extent that such data should be qualified as personal data",* or *"It is possible that our services involve the processing of personal data. In this case, …"*). This approach is not conducive to a predictable business environment, or to scoping the rights of other stakeholders. Furthermore, reference is commonly made to external data processing agreements, making it harder for customers to obtain a clear overview of rights and restrictions.
- **References to property claims in relation to the co-generated data (including but not limited to intellectual property rights)** are uncommon, except for the more commercial statement that some of the data "belongs" to the customer or is "owned" by them (without specifying any legal consequences to this statement or clarifying what it means in practice). Database rights are not referenced under that name (or as *sui generis* rights). This would seem to suggest that the emerging common practice focuses on the explicit contractual definition of access and usage rights, rather than on the application of existing legislation (including copyright and database rights).
- Contracts commonly reserve **significant rights to analyse co-generated data for the OEMs, and for service providers who offer platforms to collect, aggregate, analyse, interconnect and disseminate data**. These rights to analyse include the right to process data for purposes including the evaluation and improvement of products and services, creation of new products and services, or market and trend analysis (without further constraints).
- Provisions in relation to the **accessibility of data for other purposes than the principal functions provided by the IoT manufacturer** are rare, and only encountered in agreements that relate to a service that was initially designed as an ecosystem that should enable and facilitate accessibility and use of data to other stakeholders that the parties who are co-generating the data. This does not imply that data access and usage requests from other service providers would not be allowed, but rather that such requests are not built into the original agreements, and that case-by-case negotiations to access the data would be required.
- Nearly all examined contracts include **revision clauses** that allow OEMs and service providers to modify terms of the agreements in substantive ways (including by changing usage rights to the data) without explicit opt-in consent of the data co-generating party. While this is in line with how a part of the data economy operates currently and while this approach is conducive to flexibility and innovation, legal certainty of course also suffers.

The general conclusion emerges that the current legal framework is not particularly well suited to addressing the legal aspects of the policy problem.

In particularly, when mapping the observed characteristics of the examined agreements against the principles promoted by the European Commission in the Communication 'Towards a common

European data space' for contractual agreements in business-to-business (B2B) data sharing[295], several weaknesses emerge. Transparency and predictability of contractual arrangements suffer from ambiguous language and revision clauses, and there is little to no consideration for shared value creation (except in the specific context of data sharing platforms that are designed to enable a user to share their data with third parties). Lock-in risks are also real since the accessibility and usability of data (including technical and operational issues such as interfaces and standards) are in the current state of play almost universally absent from contractual arrangements.

While these problems could conceptually be resolved purely through sufficiently tailored legal arrangements, there is no commonly accepted best contractual practice at this stage. Rights and obligations to co-generated data thus remain fragmented, subject to the contractual power of participants in the ecosystem, and therefore not necessarily optimally suited to supporting innovation and competition.

3.1.3.1.2 Ecosystem

In the frame of this study, **co-generated data** are perceived as machine-generated data, i.e. the data that is recorded, collected or produced by a connected device, network or asset independent of any direct human intervention, which cannot be directly nor indirectly linked to a natural person and therefore are non-personal data at the time of collection.[296] This sub-study focuses in particular on co-generated data produced in the context of IoT ecosystems. According to the stakeholders interviewed in the frame of this study, a potential classification of the different types of industrial machine-generated data, could be the following:

a) **Internal machine operations data**, linked to the technical functionalities of the machines (e.g. machine health and operation technique data linked to efficiency of engine, life-cycle and operating hours of the machine, fuel usage, tire pressure data, equipment function and reference data etc.), for which IoT solution providers (e.g. Original Equipment Manufacturers or OEMs) are the data holders and controllers. In several cases, these data or data sets might be considered as *"trade secret"* of the IoT object manufacturers. Such data might be claimed to be protected by IP law, being invisible to the customer/smart machinery user, who cannot access and use it. However, this is not always the case, as it depends on the exact data points and the contractual agreement among the parties. The interviewees further explained that there are different levels of this data set, including among others:
   (i) Data sets linked to machine's performance, which can be interpreted into predictive maintenance and repair of the machine. Such data can be made available to the user and/or third parties, depending on the contractual agreement and the industry sector.[297]
   (ii) Data gathered at sensor level. This includes granular technical details which is controlled and used by the IoT solution provider for deep technical analyses; Access to these data from the "user" side is limited.
   (iii) Software related data, protected by intellectual property rights, preventing users to have any visibility, access and usage rights on such data.

[295] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Towards a common European data space", COM/2018/232 final; see https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN

[296] Realising the economic potential of machine-generated, nonpersonal data in the EU, Report for Vodafone Group, p.14, https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

[297] It was observed during the interviews that in the construction equipment, manufacturing or transport and logistics industry sectors, these data can be made accessible to the user, while this is not always the case in the agricultural sector

b) **Machine-generated data, produced in a "site context" i.e. in the field of the customer/smart machinery user** (e.g. agronomic data or data produced in the construction site); the access and usage rights of such data is regulated by bilateral contractual agreements. In many considered cases, this type of data might be perceived as "*trade secret*" or "*commercially sensitive information"* of the customer/user, who might , therefore, be the data controller or co-controller of such data sets.[298]

c) **Anonymised aggregated data**, arising from the processing of several industrial data by the IoT solution providers (e.g. OEMs), who are considered as the *data processors*, without necessarily being the data holders of the data sets described in category b. This type of data cannot be re-traceable from the user side.

The following table summarizes the above-mentioned machine-generated data classification, based on the information collected during the interviews.

Type of data set	Short Description	Data holder
Internal machine operations data	Data linked to the technical functionalities of the machines. There are different levels of them including among others: a) Data sets linked to machine's performance (predictive maintenance related data); b) data gathered at a sensor level; c) software related data, protected by IPR	IoT Solution provider (e.g. OEMs, smart machinery, connected vehicles manufacturers)
Machine-generated data, produced in a "site context"	Data produced in the field of the customer/smart machinery user (e.g. agronomic data or data produced in the construction site)	IoT Solution user (e.g. farm corporations; construction companies; transport and logistic companies)
Anonymized aggregated data	Data arising from the processing of several industrial data by the IoT solution providers,. This type of data cannot be re-traceable by the user or other third parties.	IoT Solution provider (e.g. OEMs, smart machinery, connected vehicles manufacturers)

Even though this sub-study focuses on machine-generated data, beyond personal data protection questions, it should be noted that a common issue identified by interviewed stakeholders in several industry sectors (agricultural, transport and construction), is that the distinction between industrial and personal data is difficult or even impossible. The reason is that there is no clarity in legislation on this topic, leaving a big margin of interpretation. This big margin of interpretation creates hesitancy within the legal communities of companies and constitutes a factor significantly limiting B2B data sharing and monetisation options for industrial non-personal data. This issue refers mainly to the data sets of category b - i.e. data, produced in a "site context", but in some cases even to data sets of category a – i.e. internal machine operations data, linked to predictive maintenance and repair information. In the transport sector for example, the majority of data is connected to the vehicle identification number and the driver. Therefore, even deep internal vehicle data might be considered as personal and subject to GDPR's provisions.

[298] In the agricultural sector, examples of such data sets might include, among others, land data, planting data, crop seed data, soil and fertility data, livestock data, disease and pest management data, supply chain data, farm operations and management data,

The stakeholder ecosystem consists of data holders, data co-producers, data re-users and data intermediaries. The table below provides an indicative mapping of the main stakeholders affected within the value chain of the Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.

Table 59 - Stakeholder scope (data value chain mapping)

Sector	Data holder (or data controllers)	Data co-producers (or data "co-owners")	Data re-user	Intermediaries	Personal data[299]	Purpose
Horizontal / Cross-sector approach	Private sector companies: IoT Product/ service providers (i.e. OEMs, smart machinery, connected vehicles manufacturers)	Private Sector Companies: IoT product/ service users (i.e. farm corporations, construction companies)	Private Sector Companies: independent service providers (i.e. data analytics companies, data platforms, competitors)	Potentially - B2B Data Intermediaries (i.e. data marketplaces, industrial data platforms, trusted third parties, data collaboratives, data trusts)	Sometimes	Business; R&I; Public Good
Agriculture	IoT Solution providers (i.e. Farm equipment manufacturers e.g. smart machinery, sensors providers)	Farmers	Third parties/ Independent service providers	B2B Data Intermediaries	Sometimes	Business; R&I; Public Good
Manufacturing	IoT Solution providers (e.g. smart machinery, sensor providers)	Factories/ Manufacturers	Third parties/ Independent service providers	B2B Data Intermediaries	Sometimes	Business; R&I; Public Good
Transport and Logistics	IoT Solution providers (e.g. smart devices, sensors providers)	Transport and Logistics Companies	Third parties/ Independent service providers	B2B Data Intermediaries	Sometimes	Business; R&I; Public Good
Construction	IoT Solution providers (i.e. construction equipment manufacturers, telematics, tracking	Construction Companies	Third parties/ Independent service providers	B2B Data Intermediaries	Sometimes	Business; R&I; Public Good

[299] This is questionable, as some specific machine-generated data sets could be linked to identifiable person but there are industrial data that may be broadly considered non-personal data. The study focuses on machine-generated non-personal data.

	sensors providers)

Data holders are defined in the OECD report on 'Enhancing Access to and Sharing of Data'[300] as "a party who, according to domestic law, is competent to decide about the contents and use of (personal and non-personal) data regardless of whether or not such data are collected, stored, processed or disseminated by that party or by an agent on its behalf." (In alignment with the definition of *"data controller"* provided by the OECD Privacy Guidelines (OECD, 2013))[301]. They are sometimes considered *"data owners"*, even though they may not have any legal "ownership" rights over the data they control. For this reason, they are sometimes called *"data stewards"*. Data holders are among the most critical actors for data sharing and re-using because without their active contributions there would be no data available. Therefore, properly aligned incentive mechanisms that target data holders without discouraging their data-related investments are crucial for a well-functioning data-sharing ecosystem. The effectiveness of incentive mechanisms will depend on the extent to which data holders can benefit from data sharing and be protected from risks. The availability of sustainable business models, IPR and privacy protection, and mediation through trusted intermediaries are among the most crucial factors for incentivising and facilitating data sharing across society, but challenges remain.[302]

The data holder or data controller determines the purposes for which and the means by which data is processed. Therefore, the term should be differentiated from the **data processor** who processes the data on behalf of the controller (for example an OEM might be processing data without being the exclusive data "owner").

As described in the previous section, it is not always clear whether "de facto exclusive control" over machine-generated industrial non-personal data in IoT ecosystems aligns with a smart product/service provider or with the user, as this issue is regulated by bilateral contractual agreements between the parties, on a case-by-case basis. According to the Association of Equipment Manufacturers (AEM), **the "owner" or controller of machine-generated data is the entity that holds title to the device that recorded the data**. The same association however explains that the situation gets more complex in the case of lease holdings, or in cases data are "owned" by one party but controlled by another one. Possession of machine-generated non-personal data is linked to control and does not necessarily equate to title which is linked to "ownership".[303] In particular, it is not clear whether data rights are "exhausted" at the point of sale or lease of the sensor device. (i.e. if rights are exhausted at the point of sale of the device, the buyer and user of the device could acquire all rights to the data collected during use; if not, the manufacturer may retain rights to the data. A similar principle could be applied to leasing contracts, where rights could remain with the owner of the rented equipment, or be transferred to the user of the equipment).

Taking as an example the **agricultural sector**, contracts often regulate the relationship between farmers and machinery and service providers, all of which, once they enter into a business relationship with a farm, become potential stakeholders in the agricultural data generated on and

[300] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2, https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[301] OECD (2013), Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, amended on 11 July 2013, OECD, Paris, https://legalinstruments.oecd.org/public/doc/114/114.en.pdf.

[302] Quoting: OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2, https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[303] AEM, Who owns the data generated by machines? https://www.aem.org/news/who-owns-the-data-generated-by-machines

about that farm.[304] This means that data access and usage rights related to data "ownership" might be different from case to case depending on the contract terms. On the other side, according the EU Code of conduct on agricultural data sharing by contractual agreement, rights regarding data produced on the farm or during farming operations are granted to ("owned by") the farmer and may be used extensively by them.[305] However, given that this is a recent development, it is not clear to what extent this code of conduct is being respected by the various market players and transformed into contractual clauses. In the USA, according to a similar code of conduct, farmers have the right to decide on the use of information generated on their farming operations.[306] Nevertheless, farm equipment manufacturers have developed a system of agreements with a high level of transparency to enable agricultural machine-generated data to flow freely (i.e. transparency regarding the purpose of use of farmer's data).[307] Apart from Europe and the US, a similar code of conduct has been developed in New Zealand (Farm Data Code of Practice).[308]

In the **transport and logistics sector**, according to a private company in the field, IoT solution providers usually act as the gatekeepers for commercial vehicle data, but open interfaces are giving fleet managers more control, "the ability to collect data from mixed fleets will allow fleet managers to assume a more prominent role in partnerships with OEMs, who can only collect data from the vehicles they manufacture".[309] Similarly, regarding smart manufacturing in the **manufacturing sector,** the IoT data might not stay on the factory premises in some cases, "as is often the case in IoT, the sensors and machines are constantly sending data to their mothership, i.e., the vendor".[310] However, according to the interviewed stakeholders representing the manufacturing sector, the customer/user is usually the data holder and able to decide whether data produced by the machine will leave the factory premises or not, while machine manufacturers remain the data holders of datasets that aim to protect their "know-how". When comparing manufacturing with other industry sectors, fewer issues related to access and usage rights on machine-generated data were identified. This is presumably linked to the fact that manufacturing is among the highly digitalised industry sectors and advanced in terms of data sharing, where best practices are already being followed to a certain extent.

In the **construction** sector, a construction machine generates large volumes of data, but within a closed ecosystem which only the manufacturer can access to deliver services such as telematics.[311] The CECE report of 2019 on "Digitalising the Construction Sector" presents an innovative model similar to a "Banking Model", currently adopted by CNH Industrial, which could be used as an example for data sharing in the construction sector. Based on this model, the OEM provides the customer with a digitally equipped machine that can collect a wide array of data about its functioning and usage; when signing the contract, the customer has the right to sign a legal disclaimer that establishes which data can be accessed by the OEM and for what purposes (i.e. monitoring and predictive

[304] Jouanjean, M., et al. (2020), "Issues around data governance in the digital transformation of agriculture: The farmers' perspective", OECD Food, Agriculture and Fisheries Papers, No. 146, OECD Publishing, Paris

[305] EU Code of conduct on agricultural data sharing by contractual agreement,, https://copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf

[306] Privacy and Security Principles for Farm Data, https://www.fb.org/issues/innovation/data-privacy/privacy-and-security-principles-for-farm-data

[307] DRAFT Data Privacy and Use Whitepaper, http://s3.amazonaws.com/aggateway_public/AgGatewayWeb/WorkingGroups/Committees/DataPrivacySecurityCommittee/Data%20Privacy%20and%20Use%20Whitepaper%20v3.5.pdf

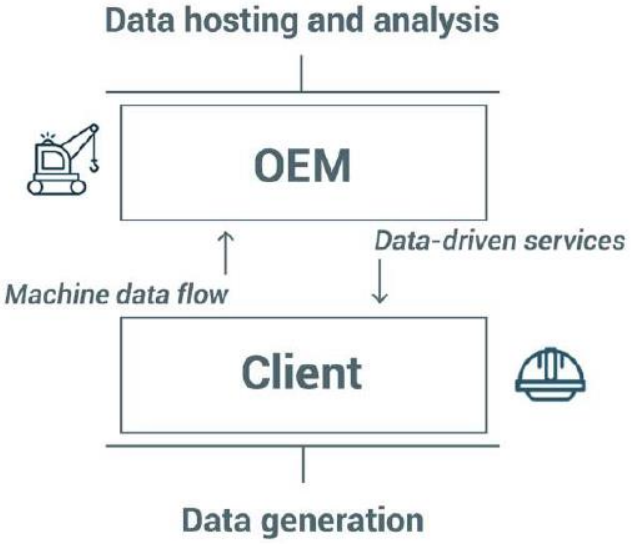
[308] New Zealand (Farm Data Code of Practice), http://www.farmdatacode.org.nz/wp-content/uploads/2016/03/Farm-Data-Code-of-Practice-Version-1.1_lowres_singles.pdf

[309] Quoting: "IoT trends in transport and logistics" https://www.telenorconnexion.com/industries/transportation-logistics/iot-trends/

[310] Quoting: "To Share Or Not To Share – Data Ownership & Other Challenges Of Smart Manufacturing" https://biztekdotblog.wordpress.com/2018/11/10/to-share-or-not-to-share-data-ownership-other-challenges-of-smart-manufacturing/

[311] Overdahl S., (2020), IoT and Data Ownership: Your Device your Data?, The IoT Business Magazine, https://www.smart-industry.net/iot-and-data-ownership-your-device-your-data/

maintenance, energy consumption analysis). Legal "ownership" of data stays in the hands of the customer as he is considered the originator of the data. The originator of the data is the only subject that can decide what to do with his data (i.e. granting access to third parties). The OEM hosts the customer's data that is analysed in order to improve the efficiency of the product. In exchange for the right to access data, OEM provides the customers with data-driven services that are linked to the results of data analysis.[312] This model presents an example for data sharing in the construction sector, and does not necessarily represent the current situation on the distribution of rights on co-generated data in this industry sector, as this might be different from case to case, depending on the contractual agreement between the parties.



Source: ID Consulting and S.I.RI.O.

The aforementioned information show that machine-generated data might be controlled by different market players, depending on the exact type of data, industry sector and the contractual agreement between the parties. For this particular sub-study, **data holders are understood as the large private sector companies, which are the IoT product or service providers** (i.e. Original Equipment Manufacturers, smart machinery or connected vehicles manufacturers). This is based on the assumption that, currently, in the majority of cases and data sets, data is being controlled by them, as such large companies appear to often have de facto control over the data due to their market position.

Data co-producers in the value chain are understood as **players having an important role in the generation of data from which data access and usage right might derive**. Based on the above-mentioned assumption, those include private sector companies which are IoT product/service users (i.e. farm corporations, construction companies, etc).

Data re-users are independent service providers and include players interested in accessing data (e.g. business/service providers' competitors and same sector down-stream providers) and players

[312] Quoting: CECE report, 2019, "Digitalising the Construction Sector" p. 24

interested in re-using data (e.g. data analytics companies, data platforms, etc).[313] In both cases, data re-users do not necessarily contribute to the production of the data.

Data intermediaries might also be involved in the value chain. The OECD report 'Enhancing Access to and Sharing of Data' defines 'data intermediaries' as organizations that "enable data holders to share their data' which 'may also provide additional added-value services such as data processing services, payment and clearing services and legal services, including the provision of standard-license schemes".[314] B2B data intermediaries can include, among others, data marketplaces, industrial data platforms, trusted third parties, B2B data collaboratives, B2B data trusts.[315]

3.1.3.2 The problem, its magnitude and the stakeholders affected

This section identifies the problem, its causes and its effects – all graphically represented in the problem tree below.

Data is produced at an exponentially increasing speed and by larger and larger groups of players. In the business environment and across all sectors, the number of situations in which data is co-generated by multiple players and/or in which data access is necessary for third parties to carry out their activities is also increasing exponentially. The main problem identified under this domain is the **lack of clarity in determining and disseminating access and usage rights on co-generated IoT data** in the economy. This problem is further accompanied by lack of trust, legal uncertainty and imbalances between economic actors, constituting **significant barriers to data sharing**.[316]

"Ownership" rights over both personal and non-personal data is generally considered incompatible with EU law[317]. A core issue affecting opportunities presented by IoT data is that in many sectors, there is no agreed approach yet for control, access and beneficial usage rights over machine-generated data. While there has been steady progress on the complex debate on personal data, for machine-generated non-personal data[318] there is little clear headway on what rights different parties have on the data, e.g. the device manufacturer or the device user.[319] The legal barrier of data "ownership" is also identified in the 2018 report on the economic potential of machine-generated, non-personal data in the EU[320] as a major barrier to data sharing of machine generated non-personal data, cited in several industry sectors (including i.a. healthcare, manufacturing). In several industry sectors, in the case where a machine vendor collects data, it is unclear whether access and usage rights over such data should be reserved by the machine vendor or the machine operator as a result of a de facto situation or superior market power. According to the study, further guidance from

[313] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[314] OECD Report (2019) 'Enhancing Access to and Sharing of Data', chapter 2, https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm

[315] For more details, see Impact Assessment on enhancing the use of data in Europe, Report on Part 1 – Data governance, p. 37-40, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=71220

[316] Data sharing within the lines of competition law principles, avoiding in particular potential anti-competitive effects, such exchanges of commercially sensitive data between competitors may have.

[317] This study focused on issues around the lack of clarity on rights and obligations with regard to the access and usage rights of machine-generated data. Terminology related to data "ownership" rights is being examined and used only as part of literature review, as there is no basis in EU law for data "ownership" rights.

[318] Machine-generated data can either be personal or non-personal data, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0009&from=GA

[319] https://www.deliveringvaluethroughdata.org/pdf/sections/4.5%20Ownership%20of%20Machine%20Data.pdf

[320] Realising the economic potential of machine-generated, nonpersonal data in the EU Report for Vodafone Group, July 2018, https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

regulators on broader regulatory/legal concerns related to the access and usage rights on data could help clarify and enable data sharing.[321]

Furthermore, the 2018 study on emerging issues of data "ownership", interoperability, (re-)usability and access to data, and liability[322] put forward a categorisation of business players involved and interested in (co-)produced data and business data in general:

- **Players co-producing data**: product/service providers (i.e. Original Equipment Manufacturers, telecommunication companies, sensor providers) and product/service users (i.e. airline or bus companies);
- **Players interested in accessing data,** without necessarily contributing to their production,such as providers' competitors and same sector down-stream providers;
- **Players interested in re-using data,** without necessarily contributing to their production, such as data analytics companies.[323].

The same study highlights that contractual barriers are impeding the sharing, access and re-use of data in the EU, with issues that are more important for 'data users' than for 'data producers'.[324]

In this context, the European Commission has launched several important initiatives aimed at supporting the development of fair practices for access and reuse of co-generated data and business data, all focused on providing clarity on possible contractual arrangements and sharing knowledge about the best solutions for data sharing. In the Communication "Towards a Common European Data Space"[325], the European Commission lists a number of key principles to be followed in any contractual agreement and notably: transparency, shared value creation, respect for each other's commercial interest, ensuring undistorted competition and minimised data lock in[326]. Furthermore, the European Commission established in 2019 a Data Sharing Support Centre[327] with the objective of spreading knowledge and best practices around this topic.

Specific legislative and non-legislative measures that have been taken in the past could be considered as enablers that contribute in unlocking access to valuable data in certain sectors, even though their objectives might not be directly linked to data sharing. Examples of such legislative measures include, among others, the Regulation on the free flow of non-personal data,[328] the Payment Service Directive

[321] Realising the economic potential of machine-generated, nonpersonal data in the EU Report for Vodafone Group, July 2018, https://www.vodafone.com/content/dam/vodcom/files/public-policy/Realising_the_potential_of_IoT_data_report_for_Vodafone.pdf

[322] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[323] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[324] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[325] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "Towards a common European data space", COM/2018/232 final, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN

[326] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, "Towards a common European data space", COM/2018/232 final, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN

[327] https://eudatasharing.eu/about-us

[328]Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1807

2[329] in the financial sector; the Intelligent Transport Systems Directive[330] in the transport sector; the EU Regulations on Drones[331] in the aviation sector; the Medical Devices Regulation[332] in the healthcare sector. Within the agricultural sector, a Code of Conduct on Agricultural Data Sharing has been developed by eleven stakeholders[333] of this sector in 2018[334].

The question of access and reuse of co-generated and business data has also sparked a very interesting debate within the academia. The Max Planck Institute for Innovation and Competition for instance has set up a working group to answer questions such as what are legitimate reasons for regulatory intervention (e.g. to facilitate access to data) and how should such interventions be designed in order not to unnecessarily impair innovation and competition[335]. The European Law Institute is currently running a research project on data rights and transactions, which aims at determining the rights distribution of co-generated data[336]. The Federation of German Industries (BDI) has recently conducted a study on the biggest existing barriers to data use for companies. According to the results, 84,2% of 500 participating companies think that "legal uncertainty regarding data usage rights" is holding them back from data sharing, being considered as the third most important obstacle.[337] Similarly, according to another survey conducted in the frame of a study on data sharing between companies in Europe, legal uncertainty about rights over data constitutes the second most important obstacle to data sharing, within 54% of the respondents confirming this statement.[338]

Despite all the attention that has been given to this topic in the last years, in the European Data Strategy of 2020 it is mentioned that *"in spite of the economic potential, data sharing between companies has not taken off at sufficient scale"*[339]. A recent survey showed that dark data may be the biggest untapped resource in business today, as one third of the survey respondents reported that more than 75% of their organization's data is dark. Machine data in particular, which constitutes a major source of dark data, is growing much faster than traditional organizational data, with an accelerating importance to decision making and organisation success.[340] This is further confirmed by another study, which reveals that apart from a handful (8%) of elite companies, the vast majority of businesses are not capturing value from data but only eking out small gains across a few, isolated experimental use cases.[341]

[329] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015L2366

[330] Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport, https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32010L0040

[331] EU Aviation Safety Basic Regulation, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1139 and Commission Delegated Regulation (EU) 2019/945 of 12 March 2019 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems, https://www.consilium.europa.eu/media/40525/delegated-act_drones.pdf

[332] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745

[333] Including Copa-cogeca, CEMA, CEJA, EFFAB, FEFAC, CEETTAR, ESA, Fertilizers Europe, CLIMMAR, AnimalhealthEurope, and European Crop Protection

[334] https://www.copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf

[335] https://www.ip.mpg.de/en/projects/details/regulation-of-the-digital-economy.html

[336]https://www.europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/data-economy/

[337] BDI,Datenwirtschaft in Deutschland, Wo stehen die Unternehmen in der Datennutzungund was sind ihre größten Hemmnisse?, https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/

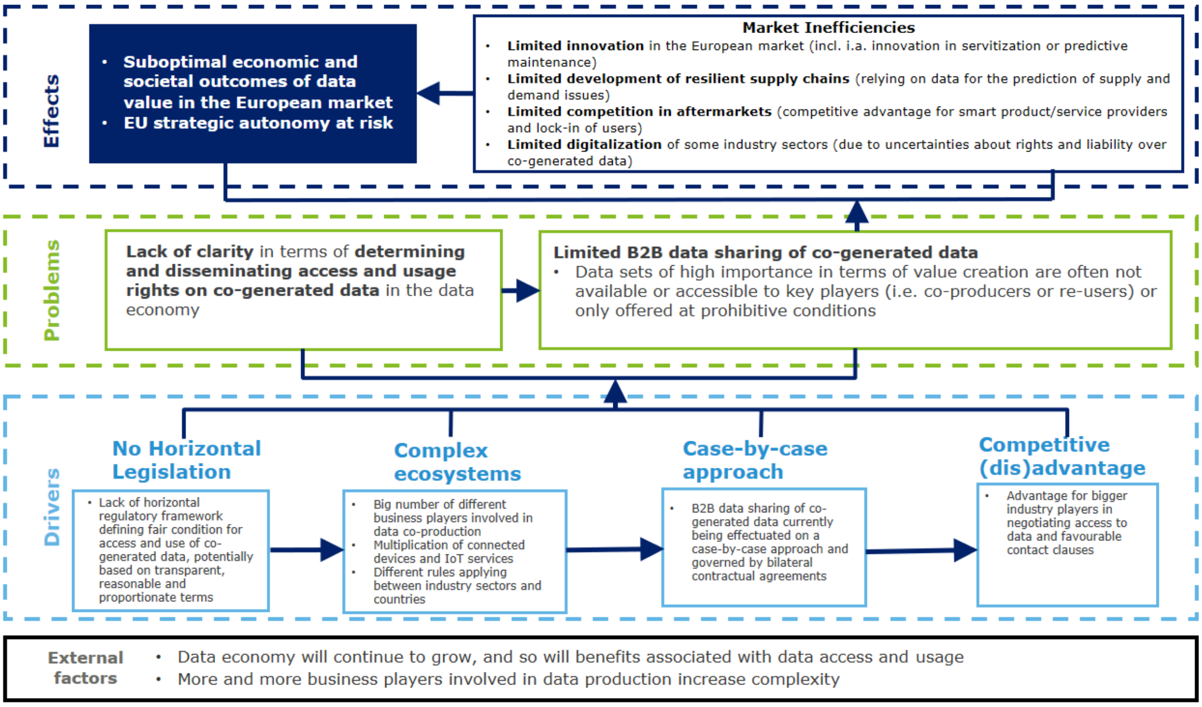
[338] European Commission, Study on data sharing between companies in Europe, 2018, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

[339] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, 2020, https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[340] The state of dark data, Splunk, https://www.splunk.com/pdfs/dark-data/the-state-of-dark-data-report.pdf

[341] Bisson P., Hall B., McCarthy B., Rifai K., Breaking away: The secrets to scaling analytics, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/breaking-away-the-secrets-to-scaling-analytics

Figure 28 - Task 2.3 Problem Tree


Effects
• Suboptimal economic and societal outcomes of data value in the European market
• EU strategic autonomy at risk
Market Inefficiencies
• Limited innovation in the European market (incl. i.a. innovation in servitization or predictive maintenance)
• Limited development of resilient supply chains (relying on data for the prediction of supply and demand issues)
• Limited competition in aftermarkets (competitive advantage for smart product/service providers and lock-in of users)
• Limited digitalization of some industry sectors (due to uncertainties about rights and liability over co-generated data)
Problems
Lack of clarity in terms of determining and disseminating access and usage rights on co-generated data in the data economy
Limited B2B data sharing of co-generated data
• Data sets of high importance in terms of value creation are often not available or accessible to key players (i.e. co-producers or re-users) or only offered at prohibitive conditions
Drivers
No Horizontal Legislation
• Lack of horizontal regulatory framework defining fair condition for access and use of co-generated data, potentially based on transparent, reasonable and proportionate terms
Complex ecosystems
• Big number of different business players involved in data co-production
• Multiplication of connected devices and IoT services
• Different rules applying between industry sectors and countries
Case-by-case approach
• B2B data sharing of co-generated data currently being effectuated on a case-by-case approach and governed by bilateral contractual agreements
Competitive (dis)advantage
• Advantage for bigger industry players in negotiating access to data and favourable contact clauses
External factors
• Data economy will continue to grow, and so will benefits associated with data access and usage
• More and more business players involved in data production increase complexity

3.1.3.3 The causes of the problem

Currently, there is no horizontal regulatory framework for accessing and co-using industrial co-generated data in the European market. In this context, B2B data sharing is being effectuated on a case-by-case approach. In the absence of horizontal legislation, B2B data sharing is regulated by bilateral contracts between the parties (IoT machinery or service users and providers). These types of contracts contain details, covering among others, the duration, value of the contract, types of data to be exchanged, access rights, type of data usage, technical aspects for data sharing, data protection.[342]

However, the lack of measures provides competitive advantages for bigger players in negotiating access to the data and favourable contract clauses. This assessment is confirmed by the 2018 study on emerging issues of data "ownership", interoperability, (re-)usability and access to data, and liability[343], according to which, the position of a business within its value chain, the sector in which it operates and its size are the three key factors determining the type of opportunities and challenges it faces for accessing and reusing data.[344] With the take up of IoT technologies and the multiplication of connected devices, the number of business players involved in data co-production is increasing and this entails an increased complexity of the relations between business partners, all interested in exploiting the value of data and getting their fair share.

[342]CECE report, 2019, "Digitalising the Construction Sector" p. 22 and "OECD, Issues around data governance in the digital transformation of agriculture: The Farmers' Perspective, https://www.oecd-ilibrary.org/docserver/53ecf2ab-en.pdf?expires=1606894801&id=id&accname=guest&checksum=46EACE368CAE08100C802080CEB6D6B5"

[343] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[344] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

As the abovementioned study found out, businesses rely mostly on bilateral negotiations and contractual arrangements to find pragmatic solutions to these questions[345]. This leads to a number of problems as solutions remain fragmented and smaller companies can suffer from lack of access to data and best practices.

3.1.3.4 The effects of the problem

The lack of regulatory framework and clarity of rights on co-produced data access and using rights leads to limited B2B data sharing and use of co-generated data. The reason is twofold. On the one side, businesses might often be reluctant to share data due to the lack of legal clarity and/or trust, which might further be accompanied by transaction and legal risk costs. On the other side, the market power of bigger industry players might be leading to anti-competitive situations (i.e. with big companies not offering access to data to other market players or only doing so at prohibitive conditions), in order to maintain a competitive advantage in the provision of aftermarket services.[346] As a result of both cases, co-generated data might often not be available or accessible for key-players, such as co-producers or (re-)users.

As a consequence, this situation might also further lead to different types of market inefficiencies in terms of lack of fair access and using rights over co-generated data. In particular, these market inefficiencies are linked to a) limited innovation in European primary and secondary markets (including, for example, innovation in servitization[347] or predictive maintenance); b) limited development of resilient supply chains (for the supply chains that rely on data for the prediction of supply and demand issues); c) limited digitalization of some traditionally conservative industry sectors (due to uncertainties about rights and liability over co-generated data); and d) limited competition in aftermarkets (with the unfair competitive advantage for smart product/service providers and lock-in of users). A more detailed description of such market inefficiencies in different industry sectors can be found in the following sections of this report.

As a final consequence, the economic and societal value of data is not maximised in the European market, putting the EU strategic autonomy at risk.

Table 60 - Intervention logic Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

Measure	Drivers	Problems	Effects
Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing	• Lack of horizontal regulatory framework defining fair conditions for access and use of co-generated data, potentially based on transparent, reasonable and proportionate terms	• Lack of clarity on access and usage rights on co-generated data in the data economy, further leading into: • Limited B2B data sharing of co-	• Market Inefficiencies linked to a) Limited innovation in primary and secondary markets**;** b) limited development of resilient supply chains; c) limited competition in aftermarkets; d) limited

[345] Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability, 2018, https://ec.europa.eu/digital-single-market/en/news/study-emerging-issues-data-ownership-interoperability-re-usability-and-access-data-and

[346] On the other side, it should also be noted that even though data-driven network effects can lead to competition-related issues, that should be traded off against the benefits of network effects for users.

[347] The term servitization refers to the gradual shift from product-centred value propositions to complex product-service systems offerings, has led many manufacturing companies to modify their business models and internal organization. Cf. (PDF) Definition of a conceptual scale of servitization: Proposal and preliminary results (researchgate.net)

	• Complex and different ecosystems with big number of different business players involved in data co-production; multiplication of connected devices and IoT services; and different rules applying between industry[348] sectors and countries • B2B data sharing of co-generated data currently being governed by bilateral contractual agreements • Competitive advantage for bigger industry players in negotiating access to data and favourable contract clauses	generated data: important data sets might often be not available or accessible to key players (i.e. co-producers or re-users) or offered at prohibitive conditions	digitalization of some industry sectors • Suboptimal economic and societal outcomes of data value in the European market • EU strategic autonomy at risk

3.1.4 Measures supporting companies in cases of conflict of laws at international level

3.1.4.1 Background

3.1.4.1.1 Context

Examples of legislation that gives rise to conflict of laws at international level

The core thesis of the problem definition is that conflicts of laws exist at the international level, where specific non-European legislation could enable access by non-European public authorities to European data, on terms that do not satisfy European legal and societal standards. In order to assess the validity of this thesis, this section will provide a summary description of several legal frameworks that could impair European data sovereignty.

The United States of America

With regard to potential extraterritorial powers, four instruments are most commonly referenced in literature. All of these permit extraterritorial data claims.

Executive Order 12333 provides a legal framework for US authorities to proceed to certain types of surveillance, by giving them the ability to collect data unilaterally and therefore without the cooperation of other parties, such as a cloud provider, under the supervision of the Attorney General. The term foreign intelligence is interpreted in a broad manner, and the possibility to proceed to surveillance exists regardless of where the data is located (i.e. including data held in the territory of the EU), and regardless of which parties are involved in the data operation (i.e. including cloud providers that are not subject to US regulations). E.O. 12333 also allows the NSA to access data 'in

[348] Even though several similarities (i.e. identical IoT objects among the different sectors) and common issues (i.e. lower bargaining power on the "user" side and SMEs, compared to big OEMs manufacturing IoT objects) are observed.

transit' to the United States, e.g. by accessing underwater cables on the floor of the Atlantic Ocean, and to collect and retain such data before arriving in the United States (where capturing it would be subject there to the terms of FISA, as discussed below).

Supervision is rather limited, and the instrument can certainly be used by intelligence services to capture data from a foreign entity stored in a cloud. The technical details remain confidential and unclear but include the exploitation of infrastructural vulnerabilities in cloud services (thus not necessarily requiring cooperation from the cloud provider). The Executive Order doesn't necessarily require targeted requests, and also allows bulk data collection. The only way to resist surveillance activities and attempts to access data by private individuals/groups is to take effective technical measures that make access difficult or impossible.

A second instrument is **Section 702 of the Foreign Intelligence Surveillance Act** (**FISA**). This federal law allows the government services to order certain service providers, including cloud service providers[349] a) to provide foreign intelligence information or b) to cooperate in obtaining such information when this is required to safeguard national security. The approach differs from the Executive Order, since:

- cloud providers under this law could be confronted with an order to submit foreign intelligence information;
- the cloud provider does not necessarily have to be a US company to be confronted with the injunction; a systematic link with the US is sufficient;
- US administrations can link the injunction to a confidentiality obligation, so that the cloud provider cannot notify the owner of the data of the submission;
- orders should not necessarily target a specific person; a general target (a general organisation or even a specific demographic group) is sufficient;
- supervision is mainly exercised by the Attorney General and the Director of National Intelligence; the role of the Foreign Intelligence Surveillance Court is rather limited.

In view of the foregoing, this law sets out very broad investigative data claiming powers that affect a very large part of the cloud market.

A third instrument to be analysed is the **Electronic Communications Privacy Act** (**ECPA**). This legislation mainly regulates powers that are comparable to the tapping powers in EU criminal procedure law. Under this Act, cloud providers can also be ordered to submit data or metadata, but only in the context of combating crime by the police and judicial authorities. The warrant may take the form of a search warrant, an administrative injunction or a specific court order. Different requirements apply to each of these forms, with the most relevant distinguishing criterion being whether or not the person concerned may or must not be informed of the order.

Finally, the U.S. also enacted the **Clarifying Lawful Overseas Use of Data ("CLOUD") Act**. The CLOUD Act has two main parts. The first clarifies that service providers under the US jurisdiction validly served with a subpoena or warrant must produce the information requested regardless of where it is stored. As such, it is a very direct and explicit example of legislation with a potential extraterritorial reach. The second directs the establishment of bilateral "executive agreements" to facilitate the sharing of data between the U.S. and other countries. The CLOUD Act amends the Stored Communications Act by adding a provision stating that "A provider of electronic communication service or remote computing service shall comply with the obligations of this section

[349] Specifically, FISA covers "*electronic communication service providers*". This term is defined by 50 USC § 1881(b)(4), and includes inter alia telecommunication carriers, electronic communication services, providers of remote computing services (cloud providers), as well as their employees, officers or agents.

to preserve, backup, or disclose the contents of a wire or electronic communication and any record or other information pertaining to a customer or subscriber within such provider's possession, custody, or control, regardless of whether such communication, record, or other information is located within or outside of the United States.”

Essentially, along with its other provisions, the CLOUD Act lays out the circumstances under which an electronic communication service or a remote computing service must comply with a U.S. law-enforcement order to disclose data within its “possession, custody, or control,” even when that data is “located outside of the United States.” Although the CLOUD Act expands the geographic scope of the Stored Communications Act, it does not change who is subject to such law enforcement orders or what type of data is covered.

For all of the instruments referenced above, under U.S. law, a U.S. federal court has the power to require the production of documents located in foreign countries if the court has in personam jurisdiction of the person in possession or control of the material. In personam jurisdiction can be based on either physical presence or minimum contacts. However, determining whether the United States may assert personal jurisdiction over a foreign corporation requires a two-step analysis: (1) a determination as to whether the party has established sufficient minimum contacts, and (2) that the assertion of personal jurisdiction is reasonable and comports with fair play and substantial justice.

Thus, these instruments may cover European companies, based on a case-by-case factual analysis to determine whether in personam jurisdiction is appropriate under U.S. law.

With regard to each of these instruments, it must be noted that the legal protection for foreigners in US law is significantly smaller than for US nationals. As a foreigner, one cannot count on the constitutional protection offered by the Fourth Amendment. Protections offered by statutory law are available, but less accessible in practice.

China

The principal relevant instruments in China that affect the possibility of extraterritorial data claims are the Decision on Strengthening Information Protection on Networks of 2012, the Cybersecurity Law of 2017, and the National Intelligence Law of 2017.

The Standing Committee of the National People’s Congress adopted the **Decision on Strengthening Information Protection on Networks in 2012**. Its goal is to protect network information security, to protect the lawful interests of citizens, and to safeguard national security and social order. Although the decision caters principally to cybersecurity, some requirement concerning data privacy were included as well.

Following the adoption of this Decision, the following broader data sovereignty framework was implemented:

- Personal Financial Information Protection Technical Specification (2020)
- China’s Encryption Law (2020)
- Personal Information Outbound Transfer Security Assessment Measures [DRAFT] (2019)
- China’s Personal Information Security Specification (2018)
- Cybersecurity Law of the People’s Republic of China (2017)
- Consumer Protection Law and its amendments (2014)
- Decision concerning Strengthening Network Information Protection (2012)
- National Intelligence Law (2017)

- Guidelines and administrative measures

A sectorial approach to protecting data was thus chosen, creating several competent authorities who supervise information security per sector, including:

- The Cyberspace Administration of China (CAC) which has primary responsibility for the supervision and enforcement under the Cybersecurity Law. It has broad responsibilities and enforcement powers, particularly in relation to cybersecurity.
- The Public Security Bureau (PSB) which has investigatory powers and enforces the Cybersecurity Law at local level.
- The Ministry of Industry and Information Technology which oversees the supervision and protection of personal data by telecom operators and internet information services.
- The Ministry of Public Security (MPS) has wide investigatory and enforcement powers to combat cybercrimes and may carry out inspections and criminal investigations, which may include inspecting the servers and systems of critical infrastructure operators and network operators.
- Provincial communication administrations are tasked with overseeing the protection of PII in the telecoms and information services sector, including the supervision and administration of personal information of telecommunication and internet users.

General requirements are set in the Cybersecurity Law, which is (among other aspects) a data localisation instrument aiming to ensure that strategic data assets remain under exclusive Chinese control, thus establishing a high degree of data sovereignty. Article 37 requires that all personal information and important data gathered or produced within the mainland territory by critical information infrastructure operators must remain on the Chinese mainland. These operators are defined in a manner that exceeds traditional European perspectives on this concept, and include non-exhaustively "public communication and information services, energy, transportation, water resources, finance, public services, e-governance" under article 31 of the Cybersecurity law. Chinese data localisation requirements are extensive, and they would also apply to companies that do not have a physical presence in China, but which have operations that involve the collection of personal information of Chinese residents (assuming that such companies would be permitted to provide strategic services).. With respect to data claims towards (potentially) non-Chinese companies, **the 2017 National Intelligence Law** is particularly relevant, since it introduces a broad framework that allows Chinese authorities to compel cooperation of targeted companies and individuals, including notably:

- Article 7: "All organizations and citizens shall support, assist, and cooperate with national intelligence efforts in accordance with law, and shall protect national intelligence work secrets they are aware of. The State protects individuals and organizations that support, assist, and cooperate with national intelligence efforts."
- Article 11: "National intelligence work institutions shall lawfully collect and handle intelligence related to foreign institutions, organizations or individuals carrying out, directing or funding foreign or domestic institutions, organizations, or individuals colluding to carry out, conduct endangering the national security and interests of the People's Republic of China; so as to provide intelligence references and bases for preventing, stopping, and punishing the above conduct."
- Article 12: "In accordance with relevant State provisions, national intelligence work institutions may establish cooperative relationships with relevant individuals and organizations, and retain them to carry out related work."
- Article 14: "A national intelligence work agency may, when carrying out intelligence work pursuant to the law, require relevant organs, organisations and citizens to provide necessary support, assistance and cooperation."

The law contains no exclusions or limitations for the exercise of these competences towards non-Chinese companies. In addition, any natural persons, e.g. board member, directors, managers or employees or other engaged by such entities – who are Chinese citizens – would be bound by a duty to cooperate because of their Chinese citizenship, even when working for non-Chinese companies. Seizure powers are thus extensive under the National Intelligence Law, and while these are formally subject to court supervision, independence of the court system and the possibility of redress is not conclusively assured.

Russia

A central part of the Russia data protection framework is the Federal Law No. 152-FZ on Personal Data dated 27 July 2006. The law was amended in 2014 by Federal Law No. 242-FZ to require that all personal data operators store and process any personal data of Russian individuals within databases located in Russia. The law is applicable to any entity collecting personal data in Russia, irrespective of their field of activity. This instrument can thus be qualified as a data localisation law, rather than as a law enabling or supporting extraterritorial data claims.

Other relevant laws include:

- The Federal Law No. 149-FZ on Information, Information Technologies and Information Protection (2006) and
- Federal Law No. 187-FZ "On the Security of the Russian Federation's Critical Data Infrastructure", which introduces requirements for infrastructure security (the "CDI Law");
- Federal Law No. 276-FZ "On Amendments to the Federal Law "On Data, Information Technologies and Data Security", which regulates the technologies that can be used to access restricted websites in Russia (the "VPN Law"), and
- Federal Law No. 241-FZ "On Amendments to Articles 10.1 and 15.4 of the Federal Law "On Data, Information Technologies and Data Security", which introduces specific regulations for instant messaging service providers (the "IM Law")

The main Russian Data Protection Authority is the Federal Service for Communications, Information Technology and Mass Communications Supervision (Roskomnadzor). All operators (i.e. data controllers and processors) must register their processing activities with the Roskomnadzor, and operators processing personal data of Russian citizens are subject to a local storage requirement. Accordingly, they must ensure that the recording, systemisation, accumulation, storage, clarification (updating, modification) and retrieval of Russian citizens' personal data is conducted only through the databases that are physically located in Russia (Art.18(5) of the Personal Data Law. The Roskomnadzor must be notified of the physical address of the datacentre. It is possible to send a copy of the local database outside of Russia.

Beyond the data localisation requirements, data seizure is regulated principally through the Operational-Search Activities Act (Law no. 144FZ of 12 August 1995), which provides that investigating authorities may perform various operational-search measures, including "inspection of premises, buildings, constructions, plots of land and vehicles", bugging of telephone conversations, and "taking of information off the technical communications channels". These measures can be conducted anywhere on Russian territory, and the Law notes explicitly that citizenship, nationality, and place of residence shall not be an obstacle to the lawfulness of these measures. Operational-search measures involving interference with the constitutional right to, among other things, privacy of the home, may be conducted subject to judicial authorisation. Independence of such authorisation is however not ensured, as was also affirmed by a June 2020 ruling from the European Court of Human Rights[350].

3.1.4.1.2 Ecosystem

3.1.4.1.2.1 Value chain

The table below provides a short description of the main stakeholders affected by cases of conflict of laws at international level.

Table 61 - Stakeholder scope (data value chain mapping)

Domain	Data holder	Data (co-) producers	Data re-user	Intermediaries	Personal data?	Purpose
Measures supporting companies in cases of conflict of laws at international level	ICT (cloud) service providers and their customers	Complementary service providers – data intelligence and analytics	Public sector bodies (law enforcement, national security)	ICT service providers (often but not exclusively cloud based)	Yes (though not exclusively)	Public policy (including national security and law enforcement)

Data holders in this context refers to ICT companies in general, and cloud computing service providers in particular, as well as their customers. The providers and their customers – or rather: the data that they collectively manage - constitute the target of potentially conflicting data claims.

Data co-producers are any service providers that extend or otherwise enhance the data or its utility. This includes data intelligence and data analytics services, which can be an inherent part of a cloud service provider, or which may be provided by an external third party. These stakeholders are relevant since much of the value of the data is created through the intelligent exploitation of the original data.

Data re-users in this domain are the public sector bodies that, for whatever public policy reason, make claims against specific data sets held by the data holders. This can include, but is not limited to, law enforcement bodies, national security agencies, or sector specific supervisors.

Intermediaries in this context refers to the ICT service providers (not their customers) and data co-producers collectively since these are both intermediaries and targets for data claims. Indeed, it is precisely their intermediary role that makes them attractive targets for data claims, since it is this characteristic that ensures breadth and volume of data.

Personal data is likely but not certain to be included in data targeted by data claims. This is of course context specific – it is viable and possible for data to be targeted that does not include any personal data, although in the typical scenario personal data will be involved, as the description of the state of play below will demonstrate.

[350] KRUGLOV AND OTHERS v. RUSSIA, Applications nos. 11264/04 and 15 others – see https://hudoc.echr.coe.int/eng#{%22sort%22:[%22EMPTY%22],%22languageisocode%22:[%22ENG%22],%22documentcollectionid2%22:[%22JUDGMENTS%22],%22itemid%22:[%22001-200719%22]}

The purpose of cross border data claims – or of legislation prohibiting, limiting or encumbering such data claims – can be summarily described as public policy goals. These include law enforcement, national security or national sovereignty, as well as fundamental rights such as privacy and data protection.

In the sections below, we will examine in greater detail how this ecosystem is affected by the state of play.

3.1.4.1.2.2 Summary conclusions on public sector interests in data, and on elementary procedural safeguards and fundamental requirements in the EU

In the sections above, a few short examples were examined of non-European legislation that either contain **data localisation claims** that appear to go beyond what the EU would provide in the context of data protection law (i.e. fundamental rights protection), or that contain **extraterritorial data claims** competences towards ICT providers that do not appear to be explicitly constrained in a manner that would satisfy European expectations of lawfulness, proportionality and independent supervision.

As these examples in the preceding sections show, non-European legal frameworks can trigger conflicts of law, in particular where national laws provide the competence to local authorities to access data relating to non-nationals if such data are hosted with a service provider subject to those non-European legal frameworks. Such frameworks usually require that the access and usage of the data is driven by national security, law enforcement, the protection of critical infrastructure, or similar justifications.

As such, this is not problematic, as all countries – including all EU Member States – will have at least some legislation in place allowing law enforcement and/or national security authorities to target data held by local service providers. However, the legitimacy of this approach hinges on procedural and fundamental rights safeguards that must apply to these procedures. Key challenges are notably proportionality (whether bulk data collection is avoided[351]), independent supervision (prior authorisation by an appropriately independent authority, taking into account the state of the judiciary and its (lack of) independence in the affected country), and redress (the ability to challenge the lawfulness of the order before an appropriately independent authority). The frameworks discussed above do not consistently satisfy these requirements, thus creating a potential conflict of laws.

In the Schrems II decision[352], the European Court of Justice affirmed the importance and significance of these safeguards. The Court noted first that the GDPR – thus in relation to personal data only – only allows transfers to a third country if that third country ensures an adequate level of protection, which must "be understood as requiring the third country in fact to ensure, by reason of its domestic law or its international commitments, a level of protection of fundamental rights and freedoms that is essentially equivalent to that guaranteed within the European Union by virtue of the regulation, read in the light of the Charter". The level of protection essentially equivalent to that guaranteed within the European Union "must be determined in the light of EU law, in particular the rights

[351] Which seems dubious based on available statistical data on data request volumes - see e.g. https://www.apple.com/legal/transparency/us.html, https://www.microsoft.com/en-us/corporate-responsibility/us-national-security-orders-report?activetab=pivot_1%3aprimaryr2, and https://d1.awsstatic.com/certifications/Information_Request_Report_June_2020.pdf for the reports from Apple, Microsoft and Amazon respectively. It should be noted that this indicates only an approximate number of request; since it is unclear how much data is targeted by a single request, it is not possible to ascertain whether this qualifies as bulk data collection.

[352] Data Protection Commissioner v Facebook Ireland Limited, Maximillian Schrems (Case C-311/18, "Schrems II"); see https://curia.europa.eu/juris/document/document.jsf?text=&docid=228677&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=9745404

guaranteed by the Charter and/or the fundamental rights enshrined in the European Convention for the Protection of Human Rights and Fundamental Freedoms ('the ECHR'), or in the light of the national law of the Member States".

The Court noted that data access on the basis of national security, defence or law enforcement could be justifiable, but that the GDPR would require "minimum safeguards resulting, under EU law, from the principle of proportionality", requiring access that is limited to what is strictly necessary. Section 702 of the FISA and E.O. 12333 were both found to fail this test. Additionally, surveillance programmes based on Section 702 of the FISA and those based on E.O. 12333 did not grant "data subjects rights actionable in the courts against the US authorities, from which it follows that data subjects have no right to an effective remedy". The Court thereafter invalidated the Privacy Shield regime that facilitated such data transfers to the USA.

As a result, for personal data in particular, clear requirements and limitations have been formulated that determine the conditions for lawful data transfers to non-European public authorities. The same is not true currently in relation to non-personal data, resulting in an arguably incomplete or at least inconsistent policy framework, since some non-personal data should clearly also benefit from legal protections that are equivalent to what the EU would provide, irrespective of where the data (or the service provider hosting the data) is located.

3.1.4.2 The problem, its magnitude and the stakeholders affected

This section identifies the problem, its causes and its effects – all graphically represented in the problem tree below.

A key challenge in adopting and implementing any European policy in relation to the data economy, digital services, digital markets or digital innovation, is the reality that relevant service providers holding relevant data are not necessarily or exclusively established or active within the EU; and indeed that service providers with a dominant market position are often established or have a business nexus outside of the EU. To some extent, this is the outcome of liberalisation policies espoused under international trade law, including specifically the General Agreement on Trade in Services (GATS), within the framework of the WTO. According to the GATS Service Sectoral Classification List (W/120) and the UN Provisional Central Product Classification (CPC), data processing activities are categorized as falling under computer-related services, specifically under "(iii) data processing services" of W/120 . In practice, the largest category of such services is commonly referred to as 'cloud services'. Under that interpretation of the same GATS list, EU (and Member State) laws and policies must apply the 'most favoured nation' principle of the GATS to data processing such services, including cloud computing services as long as none of the exceptions under GATS is invoked. The market is therefore in principle liberalised to a significant extent. This inherent freedom to provide services, free from geographic constraints, can create complexities when public authorities with designated competences to make data claims under their own national laws exercise those legally defined competences towards a service provider in relation to data pertaining to a European customer, irrespective of the seat of establishment of that service provider.

The problem can be analysed from several different perspectives. Firstly, there is the challenge of extraterritorial applicability of national rules, i.e. the application of legislation in relation to service providers established in a country where that legislation generally does not apply. Conceptually, this type of extraterritorial jurisdiction is not necessarily a blocking point. From a European perspective, as has been witnessed in e.g. consumer protection legislation, anti-trust law and the GDPR, the acceptance of a legal nexus based solely on doing business in the EU can justify the applicability of European rules to foreign service providers; this is now a relatively common and accepted practice in these policy areas. The reverse situation – service providers established in the EU – being targeted by non-European legislation and non-European public authorities – is therefore conceptually also possible without necessarily violating GATS principles.

The more salient point however relates to the risk of non-EU jurisdictions exercising their national competences towards service providers established in the EU, or to daughter, sister and mother companies of such service providers in the EU, including local branches without legal personhood, without respecting procedural safeguards and fundamental rights that would be mandatory under European legal frameworks, or without applying the instruments and procedures available under existing international/bilateral legal frameworks (such as MLATs) that these jurisdictions and the EU have agreed to adhere to. In such situations, data pertaining to European customers or European activities could conceivably be accessed by public authorities in a manner that would be contrary under EU law, e.g. due to a lack of proportionality, a lack of independent judicial oversight, or a lack of redress mechanisms (to name but a few potential problems). As will be explained below, in some cases, different service providers including EU based providers may indeed become subject to extraterritorial legislation that requires them to share data pertaining to European customers or European activities with foreign authorities.

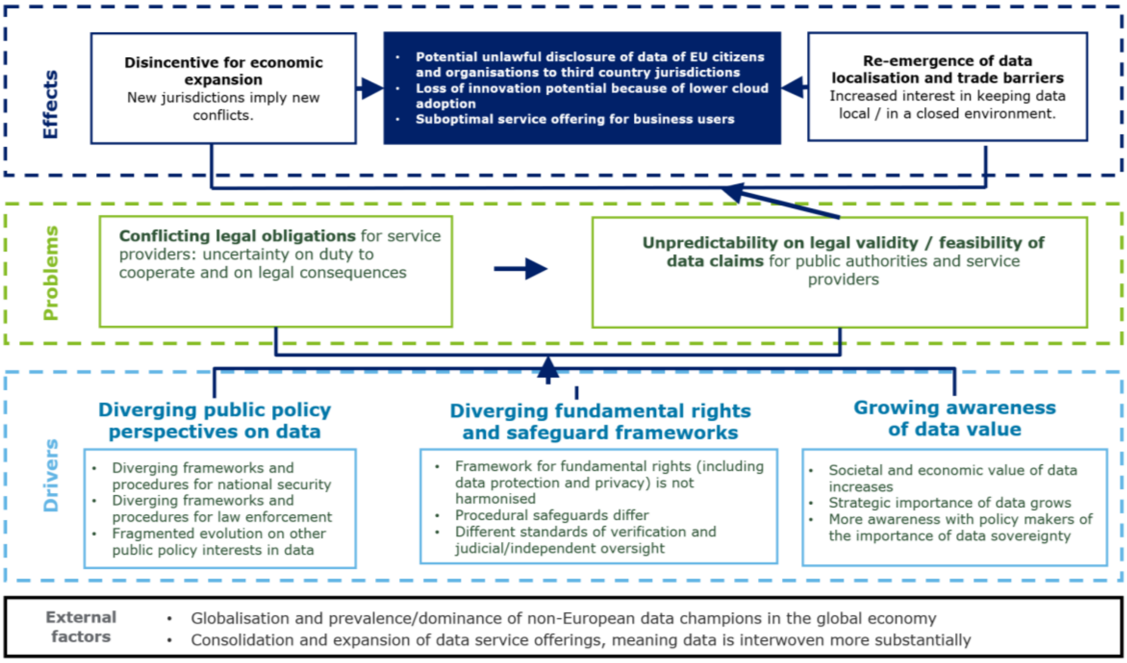
Such obligations can undermine the effectiveness of European policies, since they create a conflict of laws at the international level, with companies potentially being simultaneously obliged by non-European law to make data available to relevant authorities, and prohibited by EU law to do so. The outcome – and the problems under examination here - is **legal uncertainty for the targeted entities in relation to their duty to cooperate** with such access requests **and on the legal consequences of their decisions** in relation to such requests, and **a general lack of predictability on the legal validity and feasibility of such requests**. That uncertainty also extends to their customers who can no longer have justified trust in the protection of their data.

This problem is not entirely new, since the challenge of assessing the lawfulness of cross border data access requests has been examined at great length already in the context of EU data protection law. In that particular policy sphere, specific legal requirements that aim to mitigate the problem of cross border data access requests have been introduced in the EU in measures that enable (and also constrain) personal data transfers to third countries, notably via the General Data Protection Regulation. These requirements aim to ensure that personal data benefits from an equal level of protection, irrespective of the physical location of the service provider or its infrastructure. While these requirements also impact the liberalised approach adopted by GATS, their necessity as a measure to safeguard fundamental rights (notably the rights to privacy and personal data protection) ensures their legitimacy and their compatibility with international trade law.

It is however also clear that this data protection framework is not a conclusive answer to the problem identified above. Data relating to European citizens, businesses and public administrations may be confidential, sensitive, strategically important or economically valuable, even if it does not qualify as personal data. In those cases, the protection regime provided by the GDPR and related legal texts does not apply, and therefore the same third country transfer mechanisms and equivalence logic are inapplicable – nor, indeed, would they conclusively solve the problem, as can be seen in continued discussions on data protection risks in third country transfers.

This creates challenges in terms of data sovereignty, to be understood as the ability of European citizens and organisations to act independently in relation to their digital data[353], as stakeholders cannot be certain that their data is protected in accordance with the same high standards imposed by EU law if it is hosted or accessed from a third country. Thus, this problem is distinct from the solutions provided by EU data protection law.

Figure 29 - Measures supporting companies in cases of conflict of laws at international level Problem tree


Effects
Disincentive for economic expansion
New jurisdictions imply new conflicts.
Potential unlawful disclosure of data of EU citizens and organisations to third country jurisdictions
Loss of innovation potential because of lower cloud adoption
Suboptimal service offering for business users
Re-emergence of data localisation and trade barriers
Increased interest in keeping data local / in a closed environment.
Problems
Conflicting legal obligations for service providers: uncertainty on duty to cooperate and on legal consequences
Unpredictability on legal validity / feasibility of data claims for public authorities and service providers
Drivers
Diverging public policy perspectives on data
Diverging frameworks and procedures for national security
Diverging frameworks and procedures for law enforcement
Fragmented evolution on other public policy interests in data
Diverging fundamental rights and safeguard frameworks
Framework for fundamental rights (including data protection and privacy) is not harmonised
Procedural safeguards differ
Different standards of verification and judicial/independent oversight
Growing awareness of data value
Societal and economic value of data increases
Strategic importance of data grows
More awareness with policy makers of the importance of data sovereignty
External factors
Globalisation and prevalence/dominance of non-European data champions in the global economy
Consolidation and expansion of data service offerings, meaning data is interwoven more substantially

3.1.4.3 The causes of the problem

The problem stems from a combination of technological and market trends on the one hand, and an increasing recognition of the importance of data to support public policy goals on the other hand. The latter point in particular contains – increasing awareness of the policy interest in data sovereignty – creates some friction towards traditional policies that emphasise the benefits of liberalisation of data services, in view of the strong divergence in legal frameworks and political opinion between different areas of the world on the legitimacy of data claims (i.e. diverging perspectives on the ground for making data claims, and for opposing them).

In the background of this problem, there are the general technological and market trends, including notably the increased adoption of cloud computing, big data, IoT, and X-as-a-Service models. These trends have resulted in an ecosystem where a small number of service providers control an ever greater and continuously growing amount of data. Globalisation of ICT services ensures that market leaders are capable of rapidly capturing market shares, and the corresponding data, across the globe. As a result, such market leaders become increasingly powerful as repositories of source material and the resulting knowledge: the ability to access, use or simply measure such data confers significant

[353] Description derived from the European Parliamentary Research Service Ideas Paper on Digital sovereignty for Europe; see https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI%282020%29651992_EN.pdf

advantages to these companies, and to any entity – public or private – that can induce or compel them to grant access to their data.

Against this background of technological and market trends, it is also clear that governments across the world hold very different perspectives – and implement correspondingly diverging laws and policies – in relation to the lawfulness of their own claims to data, and on the protections that should available to citizens, business or public administrations from other areas of the world.

The most visible example in the EU are the Schrems cases before the European Court of Justice[354], which fundamentally relate to a conflict of laws between US national security and law enforcement legislation on the one hand, and European perspectives on the fundamental rights to data protection and privacy on the other hand, as will be examined in greater detail in the sections below. This issue is however not unique to the EU-US relationship, nor to the topic of personal data protection or to law enforcement and national security, since similar legislations can be found in other countries, such as Russia, China, India, and others; some of these will be briefly described below as well. While the potential lawfulness of data claims on the basis of national security or law enforcement, or on the basis of other public interests, is not in dispute, tensions exist on the one hand due to the EU's strong emphasis on procedural and fundamental rights safeguards for European citizens, business and public administrations; and on the other hand due to legal frameworks in other jurisdictions that don't adhere to these safeguards. This is particularly challenging in practice when considering the dominance of US based IT service providers – principally cloud services - in some segments of the market, which exposes European data to non-European jurisdictions on a routine basis.

This trend is coupled with the increasing awareness of regulators and policy makers of the importance of data in achieving public policy goals, and inversely of the significant strategic and operational risks that countries (including their citizens and companies) are exposed to when they lose control over such data by entrusting it to companies in jurisdictions with different legal norms and social values. In public policy, this debate is now routinely framed in terms of data sovereignty, i.e. the ability to ensure sufficiently exclusive control over critical data. While personal data protection was one of the earlier policy domains in which this discussion was addressed through specific European legal requirements on cross border data transfers, the concept of data sovereignty is significantly broader and also encompasses non-personal data. Indeed, given that society and the economy are both increasingly data driven – by both personal and non-personal data – a consistent policy framework requires that reasonable protections are available to all categories of data where a policy interest in a level playing field exists, not just for personal data.

Common examples of legal frameworks encountered outside the European Union that support data sovereignty to some extent can be encountered in data protection laws, information security and cybersecurity laws, and laws governing the activities of critical industries (e.g. financial services and health care), with the principal requirements being that data must remain within a certain jurisdiction (data location laws), that it may only leave a jurisdiction (or become accessible for a requesting entity outside of that jurisdiction) if certain strict legal prerequisites are met (such as prior approvals or commitments to respect legal safeguards), or inversely that data must be available and accessible to designated authorities upon request. Examples of such non-European legislation will be examined below.

[354] Data Protection Commissioner v Facebook Ireland Limited, Maximillian Schrems (Case C-362/14 "Schrems I" and Case C-311/18, "Schrems II"); see https://curia.europa.eu/juris/document/document.jsf?text=&docid=228677&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=9745404

3.1.4.4 The effects of the problem

The concrete effect of the problem is legal uncertainty and thus friction: when entrusting data to a service provider, a user should in theory (i.e. in order to make a rational and informed decision) be aware of the jurisdictions to which that service provider may be subject, and of the resulting risks to their data. Beyond the routine operational questions (the technical capacities, security measures, quality of service, etc.), this implies that a user needs to examine carefully which laws may apply to the service provider, and thus whether its data may become available or accessible to third parties (such as law enforcement bodies or third country supervisors) as a result of its choices.

This problem is exacerbated to a great extent by the fact that it may be impossible to determine precisely which jurisdictions could conceivably claim competence, on which grounds and to what extent, and whether any safeguards implemented by the service provider could simply be disabled by the service provider upon the insistence of an authority established in a third country.

From the perspective of service providers, the same problem manifests itself in a different manner: when choosing to operate across a more or less broad range of jurisdictions (i.e. when offering their services in a large number of countries, whether by creating an establishment there or merely by permitting usage of its service in those countries), they run the risk of becoming subject to foreign jurisdictional claims. It may be extremely difficult for a service provider, even when acting diligently and in good faith, to determine the legitimacy of such jurisdictional claims from third parties, and to assess whether complying with any claims for data would constitute a breach of contract towards their customers, a violation of a competing law (including provisions protecting fundamental rights), or both.

This friction, both for service providers and for their customers, currently can only partially be mitigated through significant expenditure, since the risks can only be assessed with specialised legal expertise for any relevant jurisdiction, which is out of reach for most companies. Even when relevant resources are available (i.e. even for large companies and well-funded governmental bodies), the outcome of extensive and expensive legal diligence may be ineffective as unpredictable court rulings may invalidate prior efforts.

The only certain way to avoid this friction in the current environment is market segmentation: a customer that chooses to rely only on a service provider which is established and active exclusively in their own jurisdiction is relatively well protected against third country jurisdictional claims. From the service providers' perspective, the same consideration acts as a disincentive for economic expansion: expanding to new markets inevitably creates a risk of extraterritorial competence.

However, such an approach is not conducive to finding the most efficient and advanced service providers, given the (for most jurisdictions) significantly smaller service offering when being confined to one's home market. As a result, companies either have to accept economically suboptimal service providers, incur substantial risk assessment costs, or simply accept the uncertainty in relation to their data sovereignty. The ultimate outcome is stress on international commerce since the free choice of service providers is under threat, and as a consequence a lower rate of cloud adoption than in other advanced economies.

3.2 Policy objectives and policy options

3.2.1 Policy objectives

The general objective of this initiative is to maximise the potential of data for the EU economy and society, in line with fundamental values, by **enhancing the use of data in the EU**. This entails that different types of current market inefficiencies will be addressed. Fair competition and innovation in the European market will be enabled to reach its full potential. Additionally, resilient supply chains that rely on data sharing for the prediction of supply and demand issues, will be further developed and digitalisation of certain industry sectors will be increased.

To reach this general objective, this initiative has **four specific objectives** with underlying operational objectives:

- To promote **business-to-government (B2G) data sharing** ensuring fair reliable and transparent, access to and use of (big) data sources held by private companies that can be valuable for innovative uses and the digital transformation of delivery of public services and better policymaking in a more flexible manner in full observance of the GDPR rules.

  Operational objectives:

  - To develop a B2G data access and reuse ecosystem that increases awareness of B2G data access and reuse potential, promotes transparency, and provides guidelines for the B2G data access and reuse activities. By developing a framework for responsible access to and use of such data sources that, due to their size and volume, could not be subject to a reporting obligation (big data) and that are better exploited for specific use-cases only; and where traditional procurement mechanisms might not work given the urgency and the uncertainty of who has what data that can be used for the purpose at hand;;
  - To address the legislative fragmentation and bring a more harmonised approach to B2G data access and reuse, providing legal certainty for both parties, by clarifying and bringing certainty to B2G data sharing, facilitating decisions on use of private sector data by government while complying with the existing legislation concerning protection of personal data, should such data be involved. This is important to strengthen the internal market, and prevent companies being approached by different authorities for the same or similar data;
  - To address supply-side disincentives by reducing uncertainty and risks, increasing trust, and lowering current transaction costs; and
  - To address technical barriers by increasing the readiness and operational capacity to use and act on data.

- To foster a **human-centric data economy** by **increasing competition**, by avoiding lock-in and enabling consumers to easily switch device, as well as to use the aftermarket service of your choice, and **enhancing innovation**, by enabling the provision of complementary services. This also includes the increased effectiveness of the so-called "right to repair" thanks to making devices data available to aftermarket players, with the deriving benefits in terms of circular economy.

  Operational objective:

  - To foster the effective portability of personal data held by home devices through the wide availability of portability by many devices, in continuous way and, possibly, in real time, directly between data holder and third parties, using as much as possible open standard format and at costs that do not hinder reuse.

- To promote fairness in **business-to-business (B2B) data sharing** contracts to further facilitate access to data and data sharing, which will benefit in particular start-ups and SMEs while ensuring compliance with EU competition rules as regards the data sharing. By improving legal certainty on access and use of co-generated (IoT) non-personal data, it will be possible to open up more opportunities for specific parties that have a close connection to the generation of IoT data to use data such objects. This would allow innovation through improved product design, to design additional services, but also to avoid lock-in effects. Improving portability of IoT data generated by individuals will allow consumers to have more choice with respect to the services around such objects, services that would depend on having access to certain data generated by these objects.

Operational objectives:

  - To establish a horizontal regulatory framework, ensuring fair, transparent, reasonable, proportionate and non-discriminatory terms for access and co-use of co-generated data in the European market;
  - To harmonize different rules that might apply between industry sectors, by identifying similarities and common issues;
  - To foster a common approach for access and usage of co-generated data and elimination of case-by-case approach regulated by bilateral contractual agreements; and
  - To enable negotiation among parties on equal terms and eliminate competitive advantage for bigger players in negotiating access to data and favourable contact clauses.

- To **reduce the risk of conflicts of laws at international level and the legal uncertainty they generate for service providers**, notably providers of cloud computing services, and establish clear safeguards and transparency for non-personal data of EU companies that may be subject to disproportionate foreign access requests.

  Operational objectives:

  - To clarify the position of data processing services subject to conflicting jurisdictional requirements for disclosure of data while respecting the EU's international obligations in the WTO and bilateral trade agreements including in the areas of services, investment and intellectual property rights.

3.2.2 Policy options

3.2.2.1 Policy options: Business-to-Government data sharing for the public interest

3.2.2.1.1 Policy option 0: Baseline scenario

Even **without EU intervention**, the **ecosystem would likely develop further**: it is likely that awareness would increase among stakeholders and the public of the potential value of private sector data for the public good – although this could be a lengthy process. As this awareness increases, and as B2G data sharing increases incrementally, a community of practice and expertise would likely emerge eventually. As this community emerges, industry-specific guidelines and protocols would likely be created within that community, although it would most likely remain voluntary and under-used. However, in the absence of EU intervention, this ecosystem, community, and guidelines and protocols may not embed the EU's values and fundamental rights.

Without any public sector policy intervention, none of the market inefficiencies identified would correct themselves. There is no reason to expect data provided from reporting requirements to suddenly be reused or combined with other datasets, nor would reporting requirements become more flexible. Monopolies would not be compelled to lower their prices. These barriers **could be addressed by national or sectorial legislation – but this would amplify the existing legislative fragmentation across sectors and among Member States**. This would further **hinder the development of common European data spaces and the completion of the internal market** – with negative consequences for the EU's overall competitiveness and innovation, including in AI and machine learning. In turn, this is likely to **reduce the EU's strategic autonomy** now and, in the future, denting the objective of a Union that is more active on the international stage – as a player rather than as a playground. Thus, formulating a clear set of rules on B2G data sharing would streamline intra-EU legislation and coordination, and would enable the EU to streamline international coordination beyond the EU. Limiting regulatory fragmentation within the EU would enable markets to scale up, thus reducing transaction costs.

In the absence of public sector intervention, **transaction costs associated with B2G data sharing would remain, as would fears of negative impacts**. **Trust may form gradually** as more data is shared with the public sector on an ad-hoc basis. Conversely, whatever level of trust currently exists **may break down** precisely because data sharing is done on an ad-hoc basis rather than sustainably and responsibly. Lastly, the **technical barriers to data sharing** – namely insufficient capabilities – **may be exacerbated** as the volume of data held by the private sector continues to grow exponentially. Lastly, the already growing **data divide** – between those that do have access to data and those who do not – may increase further, generating increased **asymmetries of insight and opportunities**.

3.2.2.1.2 Policy Option 1: non-regulatory intervention

The first policy option would entail a Recommendation to Member States, detailing the following:

- Member States are encouraged to set up governance structures to oversee and give advice on access and reuse in the public interest of data held by businesses.
- The Recommendation would include a set of criteria to help determine whether and to what extent a given purpose is in the public interest.
- Depending on the degree of public interest, different compensation models for businesses that share their data would be proposed (e.g. free of charge, marginal costs, return on investment, market price, non-monetary incentives (e.g. CSR programme)).
- Member States would be encouraged to increase their readiness and operational capacity to use and act on data.
- Private, public and civil-society organisations would be called on to create and promote the function of a data steward.

3.2.2.1.3 Policy Option 2: low-intensity regulatory intervention

The second policy option would entail a Regulation or Directive, detailing the following:

- The legal act would define four types of B2G data sharing:
  - compulsory, free of charge data sharing for a very limited set of public interest purposes defined in the EU level legislation (e.g. disaster response);
  - compulsory data sharing for certain data that is scarce, unique or needed by public authorities to ensure compliance with existing laws; preferential treatment would apply for public authorities to access the data below market price for this category;
  - a set of criteria based on which Member States would determine their national public interest purposes, for which data sharing would be compulsory and the same preferential treatment regime would apply;
  - voluntary data sharing.
- Obligations on the public sector to ensure veracity of results and independence of public-sector action (e.g. audit procedures).

- Obligations on private companies to identify the data that can be valuable for the public interest purposes identified following a request from a public authority.
- Each Member State would be required to designate a national structure(s) to convene public sector bodies interested in certain data as well as private sector data holders.[355] Their mandate would be to provide guidance on what public interest is and on the B2G data collaboratives resulting from this. These national structures, funded by public resources, would include a decision-making body composed of both public and private parties tasked with:
  - assisting public-sector organisations and private companies or civil-society organisations in entering into new data access and reuse partnerships and facilitating the sharing of good practice. Over time, such structures could become trusted third parties between the public and private sectors, by bringing the relevant players together.
  - overseeing responsible B2G data access and reuse practices and ensuring that when a public-sector body uses data obtained from the private sector, it does so legally and responsibly, without causing harm to the general public or the private-sector partner(s).
  - providing for an initial dispute settlement (arbitration) mechanism.
  - maintaining a public registry of B2G data access and reuse activities, ensuring transparency.

Each Member State would decide whether an existing entity would be appointed to take on these additional responsibilities or a new body created.

The private, public and civil-society organisations would be called on through a recommendation or other non-regulatory measure to create and promote the function of a data steward.

3.2.2.1.4 Policy Option 3: high-intensity regulatory intervention

The third option would entail a Regulation or Directive detailing the following:

- Obligations as for the lower intensity option.
- There would be an obligation for the private sector to share data for the public interest purposes identified, under preferential conditions:
  - free of charge – for very high public interest purposes (e.g. health emergencies, disaster response);
  - marginal costs for dissemination – for high public interest purposes (e.g. combating climate change and biodiversity loss, official statistics, public service delivery);
  - marginal costs for dissemination + fair return on investment (ROI) – for all other public interest purposes (e.g. tourism management).
- This requires clarity on ‘public interest’, in line with existing provisions in EU law and jurisprudence.
- Obligation to create a data steward function in all public and private sector organisations over a certain size. Based on research and analysis of hundreds of B2G data sharing initiatives[356] (including data collaboratives and public-private partnerships), one factor seems to stand out as determinative of success above all others—whether there are individuals or teams within data-holding and user organisations who have a mandate and professionally recognised function to consider access to private data for the public interest. We call these individuals and teams “data stewards.” The High-Level Expert Group on Business-to-Government (B2G) Data Sharing also noted the need for and recommended the creation of data stewards to enable responsible, accountable data sharing for the public interest. In their report[357], they write: “A key success factor in setting up sustainable and responsible B2G partnerships is the existence, within both

[355] This structure could, for instance, be the One-Stop Shops under the Data Governance Act.
[356] See https://datacollaboratives.org/explorer.html
[357] European Commission (2020). See https://digital-strategy.ec.europa.eu/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more

public- and private-sector organisations, of individuals or teams that are empowered to proactively initiate, facilitate and coordinate B2G data sharing when necessary. As such, 'data stewards' should become a recognised function."

- The national structure, funded by public resources, would have the same tasks as the lower intensity option; however, the decision-making body would be composed of public parties only.

3.2.2.2 Policy options: Measures supporting citizen empowerment ('human-centric data economy')

3.2.2.2.1 Policy Option 0: Baseline scenario

The baseline scenario is developed based on the current provisions: the application of the article 20 of GDPR combined with a set of sector specific provisions. In the current regulatory context, the data portability is a right, but not yet widely exercised. It seems to be mostly limited to data provided by the user, although this is not fully clear yet. The right does not include the provision in a continuous or real time format. The costing framework is not clear, although portability should not entail additional costs for the consumers.

There are ad hoc provisions for specific sectors, namely banking, smart meters and car.

3.2.2.2.2 Policy Option 1: non-regulatory intervention

There are a wide variety of non-regulatory measures that can be taken to address the key problem drivers identified above and foster data portability.

The proposed solution is a voluntary scheme similar to the US Green Button initiative (data portability for energy). The latter is an emulation of the Blue Button initiative implemented in the health sector, which enhance portability of personal health data, in the energy and utilities sectors. It is a label for data portability that certifies compliance with a certain standard format and effective, real time data portability.

The scheme could be applied first at smart home appliances (white goods) and fitness tracker, but potentially extendible to all sectors and companies.

The scheme would establish a technical standard and a label to identify services that provide fully fledged data portability beyond the formal requirements of GDPR (including standard format, API, continuous, real time but also secure etc.). It would be a recognizable brand that raises awareness about data portability and generates users' expectations and demand for data by re-users. Crucially, the scheme would have to be accompanied by a monitoring mechanism that monitors compliance, including usability.

The development and adoption of the standard could be entirely voluntary, or follow a more stringent path similar to the open banking initiative in the UK, where the largest banks were forced to collaborate in developing and adopting the standard through a dedicated semi-public organization (Open Banking ltd). Just as for the Green Button and Open Banking, governance will be crucial to ensure buy in by all parties.

In addition, based on the experience of Consumer Data Rights in Australia, this option will include a reciprocity clause. This clause entails that any entity wishing to take part in the scheme and access personal data would also be required to offer portability of equivalent data used in the delivery of a similar services. In other words, portability would not only be from Original Equipment Manufacturers to third party service providers such as repair shop, but also the other way around. Specific exemptions could be introduced for small business to avoid excessive costs.[358] For example, micro- and small businesses (with less than 50 employees) will be exempt from complying with the reciprocity clause requirements. This will reduce the level of costs for these businesses, but they will still be able to benefit from data portability effects.

There is a need for a dedicated technical entity spearheading the initiative and supporting the definition of standard – similar to the role of the Open Automated Data Exchange Task Force (for the Green Button Alliance)[359], the Data Standards Body for Consumer Data Rights in Australia[360], or the Open Banking ltd. in the UK). The board could also be in charge of monitoring and redress, as proposed in by the Singapore data portability proposal and related answer to the consultation: "the PDPA (Personal Data Protection Authority) will provide PDPC (Personal Data Protection Commission) with powers to review an organisation's (i) refusal to port data; (ii) failure to port data within a reasonable time; and (iii) fees for porting data, pursuant to an individual's data porting request."

Additional measures could include:

- Support to the creation of trust-based data scheme and frameworks. This refers to initiatives such as Gaia-X/IDSA, IHAN and iShare, which provide a self-contained service for data sharing that covers not the format of the data, but how it is shared, with whom and for which purposes.
- Financial support for pilots implementing data portability-based services, and experimenting on business models. One such example is the IHAN pilots. Financial instruments such as Horizon 2020 appear far-fetched, because of the length of the necessary process to set up the work programme. More agile instruments (including some already in place such as the SME instruments) could be used, for instance by developing hackathons such as the EU datathon (https://op.europa.eu/en/web/eudatathon).

Options	**Relevant Drivers**	**Precedent/Illustration**
Support the creation and adoption of data interoperability standards	Lack of technical interoperability	IDSA, Connectedhomeip.com, IHAN rulebook, PSD2
Guidelines and codes of conduct, sector based, on data portability conditions	Uncertainty	Singapore sectoral Codes of Practice
Data sharing frameworks, including trust-based mechanisms to ensure who reuses data for what purposes	Lack of trust	Singapore data sharing framework; IDSA; iShare; IHAN
Label for companies who provide full data portability (real time, API, standard format)	Lack of business case	Data-agri.fr, Green Button
Support pilots of services based on data portability under Horizon	Lack of services based on data portability	IHAN pilots

[358] Scott Farrell et al., Inquiry into Future Directions for the Consumer Data Rights, Australian Government 2021.
[359] https://www.greenbuttonalliance.org/technical-committee
[360] https://consumerdatastandards.gov.au/about

Europe innovation actions, possibly SME instrument		
Monitoring of effective data portability enforcement, including review of refusals	No enforcement of GDPR art. 20	Singapore proposal for Personal Data Protection Commission

Overall, this policy option would consist of issuing non-binding recommendations or guidelines encouraging Member States to foster the development of a market for data portability.

The creation and adoption of data interoperability standards will help address the lack of technical interoperability, while the development of guidelines and codes of conduct, sector based, on data portability conditions aims addressing uncertainty for users, data holders and data re-users. At the same time, the data sharing frameworks, including trust-based mechanisms, will help addressing the lack of trust by focusing on who reuses data and for what purposes rather that what type of data is shared and in which format. Therefore, these options will help enhancing data portability through clear rules and principles for both data holders, data re-users and the final end-users.

The monitoring and evaluation mechanisms are other important aspects of a policy recommendation. This process will measure the effective data portability enforcement, and it should also include the review of refusals of data portability. Understanding these refusals will also help further improve and adjust the policy recommendations by addressing the identified gaps of the system. However, as there is no one size fit all solution, there is the need to develop custom made solution and guidelines for different situation, and the monitoring mechanism helps further tailor these solutions to fit each situation encountered.

3.2.2.2.3 Policy Option 2: low-intensity regulatory intervention

Policy option 2 refers to the adoption of a legal instrument complementing the portability right under Article 20 GDPR by requiring companies selling smart home appliances and wearables to ensure data portability within additional, enabling conditions to what is stipulated by GDPR.

The policy option target all data holders and aim addressing some of the key problem drivers identified previously. The recommendations for data portability refer to a limited set of basic data categories, both provided and observed that should be provided in a structured, commonly used and machine-readable format. In this way, the policy option aims addressing the lack the technical interoperability existing currently in certain sectors (e.g. smart home appliances) by providing a set of standards for exchanging/sharing the data sets. At the same time, the availability of these sets of data is not time restricted, recommending only a fixed maximum delay, but underlying the need of continuously and API based provisions. Similar provisions were made also for the high-value datasets under Public Sector Information Directive and are also included in the current legislative proposal of the Digital Market Act.

There is no particular rule regarding the costs for data portability within these policy recommendations. While a fair, reasonable, and non-discriminatory approach is suggested for data portability, there are no limit of costs imposed. A similar approach for data access/sharing was used in the case of the Directive 2012/27/EU on Energy efficiency.

The policy option states that the access to the data is allowed only to the specially accredited re-users and only for direct service provisions. This approach was inspired by the Directive (EU) 2015/2366 on Payment services in the internal market (PSD2) that also included these types of provisions.

Conditions	Baseline (GDPR)	Policy option 2	Precedent
Data type	Data provided actively and knowingly by user and observed data, by virtue of the use of the service or device[361]	Limited set of basic data categories, both provided and observed	Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)
Timeliness	Without undue delay, but within one month from the request (additional two months extension possible for complex requests)	Maximum delay fixed, continuous, API based	High Value Datasets under PSI directive Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM/2020/842 final
Format	Structured, commonly used and machine-readable	Structured, commonly used and machine-readable	Directive 2012/27/EU on Energy efficiency
Cost	Free of charge (beside exceptional cases)[362]	No limit or FRAND	Directive 2012/27/EU on Energy efficiency; Singapore's Review of Personal Data Protection Act 2012 Regulation (EC) No 715/2007
Purpose	Consent-based or direct service provision	Direct service provision	GDPR
Type of re-user	No requirement	Only specially accredited re-users	Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)

The policy option 2 will include a reciprocity clause, based on the experience of Consumer Data Rights in Australia. This clause entails that any entity wishing to take part in the scheme and access personal data would also be required to offer portability of equivalent data used in the delivery of a similar services. In other words, portability would not only be from Original Equipment Manufacturers to third party service providers such as repair shop, but also the other way around. Specific exemptions could be introduced for small business to avoid excessive costs.[363] In this case, micro- and small businesses (with less than 50 employees) will be exempt from complying with the reciprocity clause

[361] European Commission, Directorate General Justice and Consumers, "Guidelines on the right to data portability under Regulation 2016/679," WP242 rev.01 (Brussels: European Commission, 2017).

[362] In exceptional cases, the data controller could charge the data subject a reasonable fee based on administrative costs or refuse comply with the request. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[363] Scott Farrell et al., Inquiry into Future Directions for the Consumer Data Rights, Australian Government 2021.

requirements, thus, reducing the costs burden for these businesses, but still allowing them to benefit from data portability effect.

Even more than in policy option 1, underlying these measures, there is the need to reinforce the capacity to support the implementation of data portability, through the creation of a dedicated support service. This is similar to what was recommended in the Furman report under the name of "Digital Markets Unit" and inspired by the UK Open Banking Implementation Entity (Open Banking ltd), the public body in charge of implementing open banking by supporting standardisation and fostering innovation, or the Data Standards Body for the Consumer Data Rights in Australia. This could fit as well under the EU support centre, the European Data Protection Board (EPDB) or a new separate body.

3.2.2.2.4 Policy Option 3: high-intensity regulatory intervention

Policy option 3 refers to the adoption of a legal instrument enhancing the portability right under Article 20 GDPR by requiring companies selling smart home appliances and wearables to enhance data portability within additional conditions to what is stipulated by GDPR, and stricter than what is defined in policy option 2.

Compared to the previous policy option, this one widens the recommendations for data portability. The data targeted are no longer limited to specific sets of data categories, but concerns all the data with exceptions, both provided and observed. Also, the requirements regarding the format that should be provided go one step further by demanding specific or open standards format. In addition to continuous and APIs based availability of data, the data holders should also real-time data access for the data re-users and final end-users. The Directive (EU) 2015/2366 on Payment services in the internal market (PSD2) is a positive example of using real-time data access to improve products and services provided to end-users. This real-time data access is significantly important for the re-users where data freshness is a sensitive aspect of their business.

The policy option recommends that data should be portable for free or at marginal costs, similar to the provisions included in the Public Sector Information Directive. Therefore, the costs will not represent a burden for the re-users that want to use data portability. Moreover, the re-users are no longer restricted by accreditation rules. Opening up the data access brings forward more opportunities for new and complementary businesses development. Opportunities for new and diversified products and services will no longer be sector specific limited as cross-sector use of data becomes a real possibility also enhancing further innovation.

Conditions	**Policy option 3**	**Precedent**
Data type	All data with exceptions, both provided and observed	PSI directive
Timeliness	Real time and continuous, including historical data, API based	Directive (EU) 2015/2366 on Payment services in the internal market (PSD2)
Format	Specific standard or open standard	INSPIRE directive
Cost	Free or marginal costs	PSI directive
Purpose	Direct service provision and product or service innovation	Singapore's Review of Personal Data Protection Act 2012 (Data Portability and Data Innovation Provisions)

Type of re-user	Any re-user, without previous accreditation	PSI directive

In this policy recommendation, the reciprocity clause, based on the experience of Consumer Data Rights in Australia, no longer apply. Thus, the data re-users wishing to take part in the scheme and access personal data will no longer be required to also provide the equivalent set of data used in the delivery of a similar services. In this case, the portability requirements would concern mainly the data holders (Original Equipment Manufacturers) with limited to no impact on third-party service providers.

Again, and more than in policy options 1 and 2, underlying these measures, there is the need to reinforce the capacity to support the implementation of data portability, through the creation of a dedicated support service. This is similar to what was recommended in the Furman report under the name of "Digital Markets Unit" and inspired by the UK Open Banking Implementation Entity (Open Banking ltd), the public body in charge of implementing open banking by supporting standardisation and fostering innovation. This could fit as well under the EU support centre, the European Data Protection Board (EPDB) or a new separate body.

3.2.2.3 Policy options: Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

3.2.2.3.1 Policy Option 0: Baseline scenario

Policy option 0 represents the status quo. This means that no action will be taken at EU level to clarify and potentially further develop rights over co-generated IoT data access and usage.

This lack of measures signifies that limited clarity in terms of determining and disseminating using rights on data in the economy will persist. B2B data sharing of co-generated data will remain at low levels, while data sets that might have cross-sector relevance and importance will continue to not be available or accessible to key players, such as data co-producers or re-users.

This scenario is associated to several market failures, as consequences. Due to limited access and usage of industrial IoT data, innovation in the European market will remain at the same levels and will not reach its full potential (i.e. innovation in servitization or predictive maintenance). Current competition issues in aftermarkets linked to competitive advantages of IoT solution providers will persist, creating high costs and lock-in situations for the users and preventing new players and third parties from entering the market. Lack of clarity on access and usage rights over co-generated data also signifies limited development of resilient supply chains that rely on data for the prediction of supply and demand issues. Finally, limited digitalisation of certain industry sectors due to lack of trust and uncertainties about rights and liability over co-generated data will persist, preventing these sectors from having economic benefits associated to digitalisation. Therefore, the economic and societal value of data will not be maximized in the European market.

3.2.2.3.2 Policy Option 1: Non-regulatory intervention

The first policy option would entail the creation of a **multi-stakeholder forum or expert group** (multisector group) of the European Commission, with industry representatives from different industries, aiming to discuss and exchange views on the problem of lack of clarity over co-generated non-personal data access and usage rights as well as on its drivers and consequences.

The purpose of this would be the creation of an **industry-driven self-regulatory framework for co-generated data**. This framework could include: a) a **standardisation** exercise; b) rules and/or initiatives aiming to **clarify rights over co-generated data access and (co-)usage**. The outcome could include the establishment of a horizontal or various sector-specific **Codes of Conduct**.

The initiative would be similar to the EU Code of conduct of 2018 on agricultural data sharing by contractual agreement; [364] however, in this case the self-regulatory framework will not be focused only on one sector having vertical but horizontal/cross-sector approach applicability.

3.2.2.3.3 Policy Option 2: Low-intensity regulatory intervention

The low-intensity regulatory intervention could entail the **adoption of a legal instrument** (a Regulation or a Directive), aiming to bring legal certainty and promote **contractual fairness** for accessing and (co-)using IoT co-generated data. This legal instrument would:

- clarify in legislation who has rights to access and (co-)use IoT co-generated non-personal data.
- establish specific **transparency obligations** for manufacturers of IoT objects on rights to access and (co-)usage of co-generated data that would oblige them to include in their terms and conditions a description of the technical and contractual access of users of such IoT objects to any data which they generate as part of the use of the IoT object.
- allow the development of approved **model contractual terms** for IoT co-generated data use, which would comprise fair, transparent, reasonable, proportionate and non-discriminatory conditions of potentially mandatory data accessing and co-using for the stakeholders involved in this value chain, including both data holders and co-producers. (e.g. in terms of price and economic share, data access, control and portability, data quality, liability in case of misuse, information about intellectual property rights, information about privacy and security of data).

3.2.2.3.4 Policy Option 3: High-intensity regulatory intervention

The high-intensity regulatory intervention could entail the **adoption of a legal instrument** (a Regulation or a Directive), laying down **specific provisions that would clarify and determine access and using rights for co-generated IoT data**. In particular, the provisions of the legal instrument would define:

- the type of stakeholder eligible to exercise access and usage rights over data;
- the type of data over which such rights are determined;
- the purpose, terms and conditions of data access and (co-)use, guaranteeing fair, transparent, reasonable, proportionate and non-discriminatory access and use of data for all stakeholders involved, in compliance with competition law (i.e. certain circumstances under which access to data is compulsory)
- A potential dispute resolution mechanism as a solution to cases where parties are not able to come up to an agreement.

The aforementioned legal instrument would be of horizontal nature, harmonising provisions that could address common issues in various sectors, leaving, though, space for modalities to be further specified in sector-specific legislation where needed.

All three policy interventions would aim to tackle key drivers of this problem assessment, including the current lack of a horizontal regulatory framework defining fair, transparent, reasonable, proportionate and non-discriminatory terms for accessing and using co-generated IoT data. These policy options would also help harmonising different rules that might exist currently among different industry sectors and countries. Furthermore, the case-by-case approach for B2B data sharing, which is currently regulated by bilateral contractual agreements would be eliminated. This would further

[364] https://copa-cogeca.eu/img/user/files/EU%20CODE/EU_Code_2018_web_version.pdf

result in the elimination of the unfair competitive advantage for bigger industry players in negotiating access to data and favourable contact clauses.

3.2.2.4 Policy options: Measures supporting companies in cases of conflict of laws at international level

3.2.2.4.1 Policy Option 0: Baseline scenario

Under the baseline scenario, the status quo is maintained. This implies that uncertainties on data sovereignty and data claims continue to exist, thus maintaining the friction in global commerce and the risk of suboptimal service offerings. Solutions to mitigate the severity of the problem can still be explored, as this is already part of the status quo, with examples being the creation of data protection safeguards to facilitate cross border data flows (such as standard contractual clauses), and the negotiation/creation of international cooperation schemes such as Mutual Legal Assistance Treaties (MLATs). However, as the overview above shows, these are not optimally efficient to resolve the problem, nor do they address the issue of access to non-personal data, for which no clear solution framework exists at the EU level. The baseline scenario can also comprise multilateral international alignment initiatives that would aim to establish a common understanding between key jurisdictions (e.g. within the seat of the G20) with respect to required safeguards in cases of extraterritorial data claims.

3.2.2.4.2 Policy Option 1: soft non-regulatory options focusing on transparency and/or operational changes

1.a. Focus on transparency

The first policy option aims to create and facilitate transparency towards all stakeholders in the data economy, with a particular emphasis towards customers of service providers whose data might become subject to extraterritorial data claims. While this policy option does not eliminate the potential of conflicts – and therefore also does not eliminate the negative impacts of the status quo fully - it does ensure that customers are fully aware of the risks to their data, allowing them to make more informed decisions when selecting a service provider.

At a first level, this policy option entails the creation of an EU level **data sovereignty knowledge centre**, i.e. an informative register, created and maintained by the European Commission, which identifies potentially conflicting legal frameworks from outside the EU, which could be triggered by the seat of establishment of a service provider location or by the location where data is hosted. The knowledge centre could allow customers to determine with some degree of accuracy whether choosing a specific service provider (or a specific service) creates certain risks.

The knowledge centre could be created **without regulatory intervention**.

1.b. Focus on operational changes

The transparency-oriented approach described above provides clarity, but does not fundamentally solve or reduce the extraterritorial data claims problem. Or rather, any beneficial impact on this point is indirect: transparency might create a disincentive for choosing a service provider subject to foreign data claims. As a result, less extraterritorial data claims might occur, but at the cost of creating data flow barriers and harming international trade, which may be an imprudent trade-off.

In this second non-regulatory option, the emphasis is on providing **templates and best practices that aim to bring about operational change,** i.e. that actually reduce risks (rather than just increasing their visibility). Usage of these templates and best practices would of course be voluntary, given that this policy option is non-regulatory.

This can be done by creating and promoting **model contract clauses/standard contractual clauses**, that would define specific **legal, technical and operational safeguards to mitigate data sovereignty risks** and that aim to elevate the level of protections available to the customers to a reasonable level. This is in line with ongoing discussions in Europe following the Schrems II case. Essentially, the contractual templates would allow providers to voluntarily commit to the adoption of such measures in a way that would be in line with existing best practices.

In addition to contractual terms, it would be possible to **create and promote technical/organisational solutions** that reduce or eliminate the possibility of extraterritorial data claims. These can include e.g. **advanced encryption and key management solutions** that ensure that extraterritorial data claims are technically not possible without the support of the customer, who is solely in charge of encryption key management.

These measures too could be presented and promoted **without regulatory intervention**.

3.2.2.4.3 Policy Option 2: soft regulatory option focusing on transparency

It is also possible to enhance transparency through a specific **low-impact regulatory measure** that e.g. obliges service providers to notify the Commission of all different extraterritorial laws of non-EU jurisdictions to which they are subject (based on their best knowledge and on prior experience), thus allowing the repository to be expanded and maintained more easily over time. If a low-impact regulatory approach is chosen, it would be possible to complement this approach by obliging service providers to notify customers of:

- **Potentially conflicting legal frameworks**, i.e. mirroring the information from the knowledge centre on their own websites in a manner that is tailored to their service offering. This answers the question for any given customer *whether there is a potential risk to their data by choosing this service provider*.
- **Statistical information on data claims** made on the basis of an extraterritorial framework, i.e. communicating not whether there is a potential risk, but specifically *whether a risk has actually materialised for any European customers of the service provider,* without providing information on an individual customer.
- **Data claims against that customer ('s data)**, i.e. a specific notification obligation that requires an individual European customer to be personally informed *whether a risk has indeed materialised for that specific customer*, to the extent that this is legally permissible.

Each of these options would not remove conflicts of law, but would increase transparency in relation to them, and allow customers to get greater clarity of their specific situation. The effectiveness of this measure may be reduced in practice if foreign legislation supporting data claims reduces the legal right of service providers to create transparency; for this reason too, the policy option is not likely to bring about optimal policy results, but none the less it can be considered a step forward compared to the status quo.

3.2.2.4.4 Policy Option 3: high impact regulatory intervention - focus on operational change

The third policy option focuses on reducing or eliminating actual international conflicts of law to a significant extent (not just increasing transparency), by limiting the cases in which they can occur or in which specific data from and EU based customer might be targeted. This would be done by targeting the service providers (i.e. cloud service providers).

Concretely, this option would require cloud providers established (or active) in the EU to:

- **Implement legal, technical and organisational solutions that reduce or eliminate the possibility of extraterritorial data claims**. To be effective, customers should be provided with an accessible overview of such measures, at the latest at the moment a contract is concluded. An overview of known and potentially relevant measures is provided in the section below, along with an assessment of their anticipated effectiveness.
- **Deny any requests from third country jurisdictions in relation to EU customer data where such transfer or access would be in conflict with EU or national law**. Requests could be granted in cases where this is supported by mutual legal assistance treaty (MLAT) procedures requiring independent supervision and approval in the EU, by international treaties, or when the third country provides for the same legal safeguards and possibility for judicial redress as it is proposed by EU legislation on international access to electronic evidence[365]. Principally, this implies intervention by a judge, a court, an investigating judge or prosecutor competent in the case concerned; limitations on the nature of cases for which expedited proceedings can be applied; a proportionality check (where investigatory measures can only be addressed to an ICT provider if a measure addressed to the company or the entity is not appropriate, in particular because it might jeopardise the investigation; recognition and due consideration of immunities and privileges; and formal procedures for remedies.

The desired outcome is a policy that objectively applies the same standards to all service providers, irrespective of their place of establishment or the location of their infrastructure. This objective standard is necessary to ensure compliance with GATS obligations.

3.3 Assessment of the policy options

This section presents the assessment of the policy options per domains identified in the previous section with regard to their effectiveness, efficiency and coherence and who will be affected.

This section presents our draft assessment of the impacts of all the options, including the baseline scenario.

The following assessment criteria were agreed on for the assessment of the impacts of the options:

- Effectiveness in achieving the policy objectives:
  - Achievement of general objective;
  - Achievement of specific objectives;
- Efficiency:
  - Costs of the option;
  - Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders;
- Coherence of the option.

Proportionality and legal/political feasibility criteria will be also considered when comparing the policy options.

To the extent possible, the assessment is built on quantitative and qualitative information, including costs and benefits. For this purpose, we took various data sources into account for the assessment of the impacts, including:

[365] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European Production and Preservation Orders for electronic evidence in criminal matters; COM/2018/225 final - 2018/0108 (COD); see https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A225%3AFIN

- Desk research, including a legal analysis;
- Interviews; and
- A workshop.

3.3.1 Business-to-Government data sharing for the public interest

3.3.1.1 Stakeholders Affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 62 – Overview of stakeholders affected by Business-to-Government (B2G) data sharing for the public interest policy options

Who?	How?
Data holders	Data holders would see a reduction of administrative and legal burdens, as well as other additional costs since the B2G data access and reuse activities, currently performed, would be facilitated and centralised through a national structure under policy options 2 and 3. From these options, a harmonised approach to B2G data sharing would be reached, resulting in more homogeneous rules and administrative practices. Data holders would benefit from access to new partnerships with data re-users, allowing them to gain new insights generated by the data analysis, resulting in better business decision making. Access to new partnerships could also mean an expansion of their revenue streams, depending on the use case established by the policy options. More effective and/or efficient policymaking would improve their business ecosystem. And overall, their contributions of making data accessible and helping to tackle societal challenges could bring reputational benefits.
Data (re)users	Data re-users would be able to access important data that could allow them tackling societal challenges, making better policies and improve their decision-making processes. Access to the diverse datasets would be facilitated and costs of this accessibility would be highly reduced (both in economic and time terms), depending on the criteria that the policies establish. A coordinating structure that brings together supply and demand, would also greatly reduce costs and speed up the process, resulting in faster response during emergencies or crises.
Society	Overall, society would benefit from better decision making and better policies and public services. In addition, individual citizens would have greater control and clarity of the data they generate through increased transparency facilitated by the national structure through the public registry. Faster and better public sector responses during emergencies or crises would also benefit the society in general.

3.3.1.2 Policy Option 0: Baseline scenario

3.3.1.2.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

3.3.1.2.1.1 Achievement of specific objectives

In the absence of EU action, current B2G data access/reuse ecosystem will continue to be underdeveloped, with few guidelines that facilitate B2G data access and reuse practices and leaving it to the public/private bodies to define the conditions under which they establish each partnership. The legislative fragmentation will continue to exist with uncoordinated national approaches and uncertainties in terms of contract durations, GDPR, and other legal frameworks governing these practices nationally. Current costs/resources spent throughout the B2G collaboration will continue to be high resulting in reluctance from the private sector to make their data accessible to the public sector. The operational capacity to use and act on the data provided will continue to be limited, constraining the public sector from achieving better results.

As a result, the potential of private sector data to help tackle societal challenges will continue to not be reaped. Unclear procedures and structures for B2G data access and reuse will persist, as well as the disharmonised approach to B2G data access and reuse, resulting in uncertainty and lack of clarity for both the private and public sectors. This lack of a fully systematic, sustainable and responsible approach will continue to **limit the public sector's access to private sector data** that can hold valuable information for the government to achieve public interest purposes such as **better policy-making or delivery of public services in a more flexible manner**.

3.3.1.2.1.2 Achievement of general objectives

Due to a lack of regulatory intervention, **it is uncertain whether unleashing the full potential of private sector data for the benefit of society is possible**. Inaction means **uncertain and disorganised B2G data access and reuse activities**, where companies are approached by different authorities for the same or similar data, making the process lengthy and complicated. Additionally, the **sustainability of access** is not guaranteed, as this access depends on bilateral agreements that may not be signed or renewed by the private sector, making it even more complicated to act according to the public interest purposes.

3.3.1.2.2 Efficiency: costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

3.3.1.2.2.1 Costs of the option

Data holders would continue to bear a number of costs in some Member States. The same costs of establishing partnerships, including the negotiations of agreements to make the data accessible to data re-users will continue to exist and be a burden for both supply and demand. For instance, according to a data re-user interviewed, they spent several months negotiating and coming into an agreement to access retail and credit card data from a private sector company. This represented a cost of at least 20-man days. In the same line, a data holder, estimated that it took them 4 months to come to an agreement where at least 12 people working 20% of their time (2.4 FTEs) were involved in this negotiation.

Other additional costs could include, the costs of pseudonymising and anonymising data prior to making it available. The costs of examining applications for data access. The opportunity costs of not developing new B2G revenue streams in a market that has a high demand for different types of data. For instance, a data holder mentioned that the initial investment of an API was around €1,000,000 and the aim was to provide aggregated and anonymised transactional data that was later used by the government. However, the pilot only lasted 2 years due to high costs and lack of revenue that could not allow them to recover from the initial investment. From several interviews, it was estimated by private-sector stakeholders that the costs of a data infrastructure creation range between 500,000 euros to 2 million euros. One data holder pointed out that the difference of costs are determined by the level of customisation required in the data infrastructure. Therefore, in the current baseline scenario, building an infrastructure is very resource-consuming due to the very different requests and requirements from public sector organisations. For every request there may be from a specific public-sector authority, a new infrastructure may be created to make the data accessible. This results in high costs for data holders.

On the other hand**, data re-users** would continue to bear a number of costs in some Member States, namely:

- Time and resources spent on identifying the data holder holding the desired data;
- Time and resources spent on producing, negotiating and submitting different (and not always successful) applications to access data from different data holders;
- Time and resources spent combining data which is not necessarily interoperable;
- Opportunity costs of not accessing data fast enough, resulting in more elaborated and time-consuming ways to obtain the same data.

3.3.1.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For **data holders** a benefit is that at the moment, they can decide on the prices and conditions of the agreements to establish the partnerships with the data re-users.

For the **data re-users**, there are no particular benefits identified.

3.3.1.2.3 Coherence of the option

No incoherence of this option with existing legislation was identified.

3.3.1.3 Policy Option 1: Recommendations

This section assesses the first policy option for Business-to-Government (B2G) data sharing for the public interest.

3.3.1.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

3.3.1.3.1.1 Achievement of specific objectives

The extent to which this policy option contributes to the specific objectives of the action is contingent on the degree to which Member States and private companies decide to follow the Commission's Recommendations or guidelines.

Regarding the **general objective** of this initiative, in terms of **unleashing the full potential of private sector data for the benefit of society**, this policy option would contribute to the extent that both private and public sectors proactively implement the Commission's recommendations, realizing the needs and the potential benefits of B2G data access and reuse practices.

However, this situation would be on a more case-by-case basis, which could make this **transition to a framework that improves access to and use of (big) data sources held by private companies in a clear and responsible way** a very lengthy process. As a result, the use of data and the benefits this could bring to public sector entities and to society as a whole, will not be reaped to its full potential.

3.3.1.3.1.2 Achievement of general objectives

Problems currently present such as a lack of clear procedures and structures for B2G data access and reuse, would likely continue to exist. This is contingent on the degree to which Member States and private companies decide to follow the Recommendations or guidelines. However, a harmonised approach to B2G data sharing across Member States would likely not be implemented or at least, it would take a lot of time to get there. As a result, companies may continue to be subject to different rules and administrative practices.

3.3.1.3.2 Efficiency: Costs and benefits of the option

The recommendations/guidelines encouraging Member States to set up governance structures, increase their readiness and operational capacity to use and act on data and create and promote the function of a data steward, were discarded from the CBA and subsequent macroeconomic analysis for several reasons.

Several stakeholders expressed doubts as to their overall effectiveness, noting that recommendations and guidelines on data sharing and reuse abound but are not always followed and that the level of ambition of such guidelines or recommendations may not be enough to see an improvement. As a result, under this option, the data sharing model would not scale up to bring substantial benefits.

According to stakeholders interviewed, it can be inferred that this policy option would not bring tangible benefits at the EU level or it may take a long time before some changes are visible. Public administrations interviewed believe that it is unrealistic that all Member States would implement these recommendations nationally, as re-users some of these recommendations for some data holders would be seen as a burden as they would see no direct benefit of implementing them. According to data holders, some elements such as the designation of a data steward within the organisation may be followed, depending on the extent that there would be an increase of requests in comparison to the current baseline scenario. Overall, data holders agree that the extent to which they follow the recommendations highly depends on whether there is an increase of B2G data sharing practices resulting from the recommendations. Some stakeholders interviewed also added that some difficulties that cities may face are regarding technical capacity to use and act on the data as well as budgetary constraints to buy data if that is the case. Such smaller cities would need to have a clear view of the overall objective and benefits of B2G data sharing, as well as some financial support from the national government in order to implement these recommendations.

As a result, it is estimated that policy option 1 would have a limited effectiveness, while any measure of its efficiency would be over reliant on assumptions linked to the content and uptake of such recommendations or guidelines.

3.3.1.3.3 Coherence of the option

No incoherence of this option with existing legislation was identified.

3.3.1.4 Policy Option 2: Low-intensity regulatory intervention

This section assesses the second policy option for Business-to-Government (B2G) data sharing for the public interest.

3.3.1.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the policy objectives.

3.3.1.4.1.1 Achievement of specific objectives

This policy option would contribute to achieving the operational objectives. Designating a national structure to convene public and private sector bodies, as part of the Regulation or Directive defining the types of B2G data sharing, would develop the ecosystem of B2G data access and reuse practices as it addresses the lack of a harmonised approach to B2G data access and reuse. It reduces risk and uncertainty due to the presence of guidelines to assess the public interests, the types of data needed, and the remuneration for providing access, depending on the use case. The national structure would oversee the private-public relationships, increasing transparency and providing a space for decision making and for disputes in case needed.

Tackling the operational objectives, would contribute to the three specific objectives of this intervention. It would improve access to and use of (big) data sources held by private companies, it would allow the development of a framework for responsible access to and use of such data sources and it would clarify and bring certainty to B2G data access and reuse practices. Such structured framework would enable the public sector to reap the benefits of B2G data access and reuse. It would contribute to reaping the potential of private sector data to help tackling societal challenges, to allow better policymaking, delivery of public services in a more flexible manner, and better responses to tackle cross-border challenges in the EU.

3.3.1.4.1.2 Achievement of general objectives

In aiming to achieve these objectives, a **fair balance between protecting the interests of the private sector** (i.e. avoiding an undue administrative burden/ additional costs) **and unleashing the full potential of private sector data for the benefit of society** would be achieved. It would likely **c**ontribute to **modernise public services**, empower citizens to know which data is being used and for which reasons, and **accelerate innovation** as data is more widely used for the **common good**.

3.3.1.4.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

For our assessment of policy options 2 and 3, we identified that both data holders and data re-users would likely incur in some costs should any of these interventions apply. These costs identified below this sub-section, are included in the cost-benefit analysis presented in Annex II. To have a better understanding of what the costs include and how were these calculated, please refer to the sub-section below. With regards to the analysis of the costs for policy option 3, these are further explained in section 3.3.1.5.2.1.

In an effort to delimit our study, we focus on some data holders and data re-users that are currently active in B2G data partnerships. Therefore, our scope for data holders includes commercial banks, mobile, supermarkets, accommodation platforms and ride-hailing companies; and for data re-users we focus on central banks, statistical offices, cities and municipalities, and national ministries.

3.3.1.4.2.1 Costs of the option

This policy option would likely incur the following **costs for data holders**. Costs associated with formalising the partnerships for data access and reuse, including the contractual arrangements (resulting from more of these arrangements). Costs of setting up and developing internal data governance approach to comply with the new legislation. Opportunity cost of not selling the data but rather providing it for free should their type of data fall under the free-of-charge category. Costs related to risk management and mitigation regarding consumer privacy and preventing competitors from accessing high-value datasets.

Designation of the Data Steward function (depending on whether the recommendation is followed by private sector organisations), from the interviews conducted with the stakeholders, it was estimated that this function would be designated if the company would see the need to do so (highly dependent on the number of requests that would result from this policy intervention). From the interviews conducted, we made an assumption for the estimation of the costs of this policy option, that 30% of private sector organisations would follow the recommendation of designating a data steward function. However, this percentage may increase as more B2G data access and reuse practices increase, and the companies see the need to designate this function to cope with the requests. According to private sector stakeholders interviewed, the data steward function would need to be a team of 1 to 5 FTEs, that have the technical and also legal knowledge to deal with the requests and the nature of the data collaboratives. When extrapolating these numbers to the EU level, we made the assumption that within our scope, the biggest companies in terms of market share per MS would likely be impacted by this policy intervention as they cover 60% of the market or more. We also made the assumption that since there would be a recommendation that raises awareness of the potential benefits of B2G data access and reuse practices, then more private and public sector organisations might engage in data collaboratives. We made the assumption that 30% of the companies of our scope would likely implement a data steward function. This would represent an annual cost at the EU level of 20.5 million euros. This cost starts in 2023 since the data steward function may take part of the decision-making process at the national structure level, representing the data holders.

Costs of normalisation (including absence of compensation depending on the use-case), **collection and reporting of meta-data and additional information needed to assess the quality of this data** (e.g. incidents affecting data quality, data gaps, etc.). **Costs of cataloguing and identifying data that can be valuable for public interest purposes and the costs of identifying and documenting aspects of business operation that have a direct impact on data quality**. According to private-sector stakeholders, the costs of identifying the data that can be valuable for public interest purposes, after the public administration requests access to a type of data, and the costs of normalising and making the datasets available for reuse, would cost approximately 4 FTEs. The costs of both activities would amount, at the EU level, to 78.06 million euros annually. We assumed during our cost-benefit analysis that these costs are starting in 2024 due to the scale-up of B2G data access and reuse practices as a result of the policy option taking effect in 2023.

In total, the costs of having a data steward function, identifying the data that can be valuable for the public interest purposes after a request is made by the public administration, and the costs of normalisation of the data sets would amount to 98.5 million euros annually, for data holders at the EU level.

Additionally, there are **costs of making datasets available for reuse.** It was estimated by private-sector stakeholders interviewed, based on their experience of data collaboratives with public authorities, that the costs of creating a data infrastructure that allows the sharing of the data in an anonymised way to comply with GDPR legislation would cost between 500 thousand euros to 2 million euros. This highly depends on the type of data infrastructure that would be required to make data accessible, the format in which data would be delivered to the public sector organisations and the level of customisation needed (where higher customisation would mean higher costs for the data infrastructure creation). Making the assumption that companies may need to create a data infrastructure once (for instance, an API where data re-users can access the data), we see that the one-off costs of such a data infrastructure would amount to 552.5 million euros at the EU level.

This policy option would likely also incur **costs for data re-users.** As with the private sector, we made the assumption that cities above 45000 citizens, ministries, statistical offices, and central banks would likely be impacted by the policy intervention, as during our analysis of current data collaboratives, these actors are currently engaged in B2G data partnerships with the private-sector stakeholders mentioned above. In this sense the costs that data re-users would likely have from this policy intervention are the following.

Costs of audit and verification procedures where the public sector would have to **ensure veracity of results and independence of public sector action**. According to a data re-user, it was estimated that this would be similar to the costs they have now, which are 2 FTEs. On the contrary, another data re-user interviewed estimates it would range from 0.5 FTE to 1 FTE only for the veracity of results per indicator for official statistics. From the interviews with these stakeholders we could see that this element highly depends on the type of data and the amount of data the organisation would handle. We estimated that on average, according to public-sector stakeholders interviewed, that the costs of this activity would amount to 1-2 FTEs. If this is extrapolated to the EU level, the costs are 192.2 million euros annually. We assume that these costs would start in 2024 as more B2G access and reuse practices scale up in this year rather than in 2023, which is when the regulatory intervention would take effect.

As with the private sector, the public sector would also have the recommendation to **designate a data steward function** within the organisation. One governmental institution mentioned that the designation of this function would cost fifty thousand euros a year for one FTE, based on a similar function they have at the moment. However, for bigger-sized public administrations, it was estimated by stakeholders interviewed that it could take up to 20 FTEs for statistical offices from big-sized EU Member States, as they require different types of data from different organisations. From these interviews, it was estimated that the data steward function highly depends on the type of data the public institution may need, and the number of requests they will make. For other public administrations, the data steward function ranges between 1 to 8 FTEs, which is based on the different types of data they manage, or they would manage should a policy intervention be in place. Moreover, the costs of designating a data steward, highly depend on whether this recommendation is followed (similar to private sector). Based on our interviews with stakeholders, the assumption made is that 50% of public-sector organisations would designate this function as the more the B2G practices scale, the more the public administrations may need to have such a function. Extrapolated to the EU level, the costs of a data steward function are 157.4 million annually. This cost, in addition to the cost mentioned above for the data re-users would amount to 349.6 million euros annually. This cost starts in 2023, since we assume that the data steward function may be part of the decision-making process at the national structure to represent the data re-users.

Additionally, there would also likely be costs for having a national structure per Member State. These costs were based on the German Data Forum (RatSWD) which is an advisory council to the federal government with similar tasks as to those the national structure would have, according to the policy options' description. For instance, RatSWD's tasks are representation of interest of data producers and data users, advisory to legislators, event organisation, connection of research data infrastructures on a European and international level.[366] They estimated, that convening public and private actors as decision-making body and assisting in new data access and reuse partnerships would cost approximately 10 FTEs. To oversee the legal and responsible use of data by public sector would be at least 5 FTEs in the beginning.[367] Considering that under this policy option, Member States would be required to designate a national structure, we estimate that this structure would likely cost 21.6 million annually at the EU level, which is likely to increase the more the B2G data collaboratives are. This cost starts in 2023, as we assume the national structure would be the first step taken as a result of a regulatory intervention.

Other costs identified for data re-users include, payments/compensation to data holders for making data available; costs of setting up and maintaining larger data processing and analysis capacity to

[366]See https://www.konsortswd.de/en/ratswd/german-data-forum-ratswd/at-a-glance/

[367] These and other costs related to the national structure highly depend on the amount of data holders and data re-users that would participate, the scope of the structure, and on the willingness of actors to cooperate.

use and act on the data provided; risk of public trust erosion as a result of the use of personal data by the public sector, potentially jeopardizing some of the benefits identified (if there would be a reduction of data access/reuse practices as a result). These costs fall out of our analysis due to the complexities of quantifying these.

The extrapolation of these figures to the EU level can be found in the table below, while the full overview of costs is in Annex II.

Table 63 - Overview of costs for Business-to-Government (B2G) data sharing for the public interest | PO 2

Overview of costs (EUR) – PO 2

		Data holders		**National structure**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures facilitating the use of data held by the private sector	*Direct costs*	EUR 552.5 million	EUR 98.5 million p.a.	-	EUR 21.6 million p.a.	-	EUR 349.6 million p.a.
	Indirect costs	-	-	-	-	-	-

3.3.1.4.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

For reasons also noted by the OECD[368], **quantifying the benefits of B2G data partnerships is "difficult" for a number of reasons** including:

- Lack of data and transparency of efforts about what data is held by companies;
- Legal uncertainties regarding the ability of companies to share data without violating applicable data protection and intellectual property rules;
- The cost and disruption suffered by companies receiving multiple and uncoordinated requests from different arms of government;
- Differences among studies regarding scope of the sectors, the types and openness of the data used, as well as differences in methodologies and metrics. Previous studies significantly differ in terms of scope of the sectors (whether public and/or private sectors were included), the types of data (e.g. personal or public data), the degrees of data openness and the methodologies, including the different level of the impact assessed (whether the effects were assessed at the organisational, sectoral or macroeconomic level).

We have identified that this policy option would likely bring **benefits for data holders**. However, since each data access and reuse activity affects different groups of beneficiaries in different ways it is challenging to construct a complete impact assessment for all data holders and data re-users that would benefit from B2G data access and reuse practices. Due to the nature of the benefits and

[368] OECD (2019), Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. See: https://www.oecd-ilibrary.org/sites/276aaca8-en/index.html?itemId=/content/publication/276aaca8-en

considering that these are more specific to the use-case, we cannot conclude that these would be representative of all data holders that would enter into the B2G data partnerships.

Throughout the interviews conducted, the data holders from different industries pointed out to diverse benefits they have received as a result of their partnerships with the public sector. These benefits identified below are not part of our cost-benefit analysis since these cannot be generalised as they are specific benefits linked to specific use-cases.

- For instance, this policy option could bring benefits for data holders in terms **of better business decision-making resulting from improved business ecosystem**, **better public services delivered, and also access to analyses and insights not available before**. Should the public sector of an EU Member State provide granular insights and thorough analyses, as a result of combining different datasets from the private sector, a positive impact for the latter would result. According to a data holder interviewed, such analyses and insights into how the economy is doing (e.g. exports, imports, consumption levels, etc.) are the main benefits of their collaboration with a public sector organisation. These allow the company to prepare their business strategy and timely adapt to the economic landscape, resulting in costs savings. They estimated that approximately 1% of their annual total revenue would be at risk without these analyses and insights.
- As a form of CSR, collaborating by offering data and data skills can **improve the companies' reputation and enhance community relationships.** From the interviews with private-sector stakeholders, reputation and CSR were acknowledged as potential benefits and also incentives when entering into private-public partnerships. A company interviewed, estimated that within their benefits of providing these services, even though there may be low ROI or even no return, is to gain brand recognition and excellent public relations results, which otherwise they may need to pay for, resulting in high costs (when aiming to reach the same amount of impressions they reached).
- Another data holder interviewed, also pointed out that besides the reputational benefits they have seen from donating their data to public sector authorities, particularly during the COVID-19 crisis, a second benefit has been the **positive purpose-driven work environment for employees** since the latter know that work is having a positive impact on the society. This has also a downstream effect for clients and suppliers because they may be more interested in working with the company because they know the social impact the organisation has. Overall, engaging in these type of activities can have multiple benefits beyond the monetary value.
- Another benefit of B2G data access and reuse practices and, of having a national structure that could bring together private and public sector to share best practices is that private sector may have **access to data domain expertise, analysis methods and models, otherwise not available or at higher cost**. As a result, companies can learn from private and public sectors, for example, on how to maximise their own datasets, which could also improve the value of their own data sets. According to a stakeholder interviewed, one of the biggest benefits of entering into the public-private partnership was the access into an ecosystem with talented data scientists that apply data in novel ways, outside of how data may be used for commercial purposes. There is an innovation value resulting from this.
- Other benefits that were identified for data holders are: **the compensation for making data available and creation of new revenue streams** (depending whether the type of data would fall under the criteria for compensation); **removal of legal and reputational risks** (where an access model by public entities could be agreed and validated**); increased maturity, capacity and discovery of data infrastructure that can be repurposed for other goals** – beyond B2G (learning by doing); **decreased legal uncertainty** due to creation of dispute settlement mechanism; fairer competition, due to relative harmonisation of B2G data accessibility and reuse rules and practices in the EU.

Another benefit identified, would be a **reduced administrative burden and decrease of time and transaction costs to establish data partnerships.** In the exercise below, we attempted to quantify and calculate this administrative burden reduction for data holders, which is possible to do due to the benefit being more generic and the possibility to link it to all types of use cases. However, this exercise is based on several assumptions that are further explained below.

According to the stakeholders that were interviewed, it takes a **large amount of effort to enter into a B2G partnership**, due to the time spent on negotiations. Based on estimates provided by stakeholders we assume that it takes 2-3 FTEs over a period of 4 months (up to a year according to some stakeholders interviewed) to establish an effective partnership for the first time or in an ad-hoc manner. This entails several types of expertise required to assess the legal, operational and technical implications of engaging in B2G sharing for a specific use case.

Assuming that B2G data collaboratives can be expected to increase in the coming years due to the visible benefits these partnerships bring, such requests for either other or similar use cases can be expected for a company that has valuable data for the public interest. One would assume that subsequent requests would be dealt more efficiently based on past experience (since the legal, organisational and technical implications are likely to be similar for the company concerned). Nevertheless, without any coordination among these B2G requests the specifics of each request by various public sector organisations may be such that each time there are still differences that require effort to investigate, either due to the fact that new data elements are requested, the purpose is different or the way in which the data is to be provided is specific to that request. Taking these factors into account, one could assume that the same types of expertise may be required per request but that the assessment of the request can be processed more quickly, e.g. 2-3 FTEs over a period of two weeks. Of course, this could be assumed as an average while it may well be that B2G requests that substantially differ from previous experience may require more detailed assessment and effort.

In the baseline scenario therefore, the cost of dealing with B2G requests would depend on the nature of the request and the extent to which it is similar to previous experiences. The feasibility of first-time requests take more effort to assess for a business than repeated requests of a similar nature.

Take for example one of the existing use cases, whereby larger cities request data from businesses that help them better manage traffic, conduct urban planning or manage tourism. Each of these use cases stem from the same type of public sector organisation and have proven valuable in the past. One might therefore expect that more cities might request this type of data to achieve the same benefits. In Germany for example there are 213 cities of similar size to those that have benefited from such B2G data sharing in the past. This could therefore mean that businesses who have such data could expect similar requests 213 times and these may take place more than once depending on the frequency of those requests (e.g. one-off, once a year, quarterly, monthly, etc.).

For other use cases there may be only one or a few relevant public sector organisations requesting the data, such as in the example of statistical offices acquiring data for the calculation of the consumer price index.

Under the policy options where a national governance mechanism has been established the assumption would be that this would result in the determination of public interest/public good and a definition of use cases based on which request can be made to companies. This would entail that the details of such a request are streamlined (e.g. the type of data elements involved, the way the data is delivered, the terms under which this is provided, etc.). Under such conditions one would assume that each request following that established use case would be far easier to process and would take a few days up to few weeks to process administratively, while companies are already prepared to deliver that data upon request (note that this does not mean that no effort is required to produce the data, but at least establishing the partnership should not require much effort). Herein lies a reduction in the administrative burden stemming from such request. Based on stakeholder estimates, new requests based on existing use cases may take 2-5 days to process administratively. Compared to the baseline scenario, this therefore entails that there could be a saving of 50-80% of efforts to process the new request based on a pre-agreed use case with defined data elements, terms and conditions, etc.

Again, calculating this requires making assumptions as to the types of requests that may arise, the amount of public sector data re-users involved, the frequency of data exchange, etc. Whereby the savings for companies that would expect many requests (e.g. form cities) would stand to gain significantly, while for companies only engaged on relatively few B2G cases the benefits would be proportionally less significant vis-à-vis the requirement to appoint a data steward function (as per PO3).

B2G use case 1: cities	
Assumptions:	in Germany up to 213 larger cities have similar requests for data in the same year
Baseline scenario:	5112 man-days = € 1.175.760,00
PO 2/3:	1022,4 to 2556 man-days = € 235.152,00 to € 587.880,00
	Average cost = € 411.516,00

B2G use case 2: statistical office	
Assumptions:	1 statistical office has a similar request for data in the same year
Baseline scenario:	24 man-days = € 5.520,00
PO 2/3:	4,8 to 12 man-days = € 1.104,00 to € 2.760,00
	Average cost = € 1.932,00

Costs of the data steward function	
Assumptions:	2 to 5 FTEs per year
PO 2/3:	€ 120.000,00 to € 300.000,00
	Average cost: € 210.000,00

The above tables illustrate two examples, one of a German company engaged in B2G data sharing with a multitude of cities in Germany and one of a company engaged in data sharing with one statistical office. It is clear that the amount of requesting public sector organisations determines the costs to process each request. Note that this estimate only includes the FTEs involved in the establishment of the partnership and not the subsequent production of the data itself. While significant savings are made, these should be compared to the costs of the data steward function (obligatory under PO 3 in particular). For the company subject to B2G use case 1, the estimated costs of the data steward function would be within the range of the estimated costs the business would incur based on the volume of requests. While in the business involved in B2G use case 2 the costs of a data steward function be not justified by far based on the same assumptions given the low volume of requests.

These two examples present extremes based on single use cases. One could also assume that the data steward function would be far less involving for the latter company and 2 FTEs on a yearly basis could be too high of an estimate for that type of data holder. This example therefore illustrates the difficulty in making assumptions on the level of aggregated costs and benefits of B2G data sharing as it requires by definition a combination of hypothetical scenarios and assumptions to come to any aggregate-level estimations. The lack of a sound underpinning to make such assumptions therefore makes it methodologically challenging to conduct a CBA; however, based on different assumptions and a hypothetical baseline scenario, we can have an idea of what the impact could be for private and public sectors under the policy interventions.

Therefore, performing the exercise at EU level requires making assumptions regarding:

- The amount of data holders in terms of sectors involved in B2G requests and how they are organised (e.g. organisations such as mobile operators or commercial banks tend to have centralised data operations at national level while others may operate more across borders and centralise data operations at that level (e.g. ride-hailing or accommodation platforms)
- The amount of potential data re-users (e.g. ministries v. cities)
- The amount of relevant use cases and the level of request made to individual data holders

A hypothetical scenario can be constructed, taking into account all private data holders (supermarkets, commercial banks, telcos, accommodation platforms, ride-hailing companies) and public-sector organisations (central banks, cities, national ministries, statistical offices) under our scope.

Looking at actual identified use cases in practice and the literature we identified 30 relevant use cases for this scope.

	Retailers (Supermarket chains) (135)	**Commercial banks (162)**	**Ride-hailing companies (40)**	**Acommodation patforms (10)**	**Mobile operators (95)**
B2G (national)	8	4	1	1	3
B2G (local)		2	5	3	3
Total use cases	8	6	6	6	6

The distinction between more 'national' and 'local' use cases reflects the fact that while some use case are more likely to be relevant for one single public authority in a Member State (one data re-user such as a ministry, statistical office, central bank, etc.) other are more 'localised' such as various large cities having similar data requests leading to many more B2G partnerships to be established for each individual data holder.

Only looking at the costs to establish these partnerships (i.e. essentially the cost of the data steward function) gives a view on how data holders in particular are affected in the baseline scenario or PO 2/3 under similar conditions, i.e. we assume the same amount of use cases and request for comparison.

Making the same assumptions as in the example above gives us the cost estimates per sector. This shows that there are potential benefits for the private sector, stemming from a reduction of administrative burden should public sector re-users ask for access to their data. In the table below we see that the reduction of administrative burden for the private sector would be from roughly EUR 248 million euros to EUR 94 million euros. The difference between the baseline scenario, where the use case are not streamlined and are more ad-hoc with associated time-consuming negotiation processes, and a policy intervention, which aims to facilitate B2G data collaboratives, results in costs savings for the private sector of roughly EUR 155 million ceteris-paribus. This however only takes into account the cost of establishing the partnership, the additional costs calculated in the CBA need to be taken into account as well (CAPEX and OPEX related to retrieving and providing the data), resulting a total cost for data holders of roughly EUR 724 million overall. Hence, the overall outcome is a cost for data holders under PO 3 of EUR 724 million to participate in the B2G data sharing cases in the scope of this example. This cost would have been EUR 155 million higher however without any policy intervention assuming the same level of B2G data sharing.

This calculation does not yet take into account the possible remuneration to data holders for the data sharing under the B2G cases. One could argue that the EUR 78 million represents marginal cost as these are the costs in retrieving and providing the data. Hence, business could be remunerated EUR 78 million annually in total if all B2G use cases are subject to covering marginal cost. Given that the policy options also cover other types of remuneration or none at all, it is difficult to assess the level of remuneration overall and would require further assumptions (as part of the use cases identified as relevant in the scope (30 in total), 3 could be considered linked to emergency situations (e.g. pandemics, refugees, natural disasters (earth quakes))). Hence, it may well be that most use cases would indeed be remunerated at least at marginal cost. Any remuneration above marginal cost (ROI or even market price) would contribute to covering the investments in infrastructure (CAPEX) that are estimated at EUR 552 million.

Table 64- Costs estimations to establish B2G partnerships in 2023

		Baseline costs at EU level		**PO 2/3 costs at EU level**		**PO 2/3 costs savings at EU level**	
		One-off cost[369]	**Repeated costs[370]**	**One-off cost**	**Repeated costs**	**One-off cost**	**Repeated costs**
Mobile operators	National level	€ 9.470.769,23	€ -	€ 1.973.076,92	€ -	€ 7.497.629,31	€ -
	Local level	€ 9.470.769,23	€ 51.782.307,69	€ 1.973.076,92	€ 18.123.907,69	€ 7.497.629,31	€ 33.658.500,00
Commercial banks	National level	€ 21.533.538,46	€ -	€ 4.486.153,85	€ -	€ 17.047.384,62	€ -
	Local level	€ 10.776.769,23	€ 58.868.307,69	€ 1.973.076,92	€ 41.207.815,38	€ 8.793.692,31	€ 17.660.429,31
Supermarkets	National level	€ 35.889.230,77	€ -	€ 7.476.923,08	€ -	€ 28.412.307,69	€ -
	Local level	€ -	€ -	€ -	€ -	€ -	€ -
Accommodation platforms	National level	€ 332.307,69	€ -	€ 69.230,77	€ -	€ 263.076,92	€ -
	Local level	€ 996.923,08	€ 5.450.769,23	€ 207.692,31	€ 1.907.769,23	€ 789.230,77	€ 3.543.000,00
Ride-hailing companies	National level	€ 1.329.230,77	€ -	€ 276.923,08	€ -	€ 1.052.307,69	€ -
	Local level	€ 6.646.153,85	€ 36.338.461,54	€ 1.384.615,38	€ 12.718.461,54	€ 5.261.538,46	€ 23.620.000,00
Total		**€ 96.435.692,31**	**€ 152.439.846,15**	**€ 19.820.769,23**	**€ 73.957.853,85**	**€ 76.614.923,08**	**€ 78.481.992,31**
			Other costs				
			CAPEX	**€ 552.500.000,-**			
			OPEX		**€ 78.063.388,-**		
			Total	**€ 724.342.011,08**			

[369] Costs to establish first partnership per use case

[370] Costs to establish partnership after the first use case partnership was established with another data re-user (i.e 1208 cities across the EU)

We have identified that this policy option would likely bring **benefits for data re-users** as well. However, similar to the benefits for data holders, due to the nature of the benefits and considering that these are more specific to the use-case, we cannot conclude that these would be representative of all data re-users that would enter into the B2G data partnerships.

Throughout the interviews conducted, data re-users from different public organisations pointed out to diverse benefits they have received as a result of their partnerships with the private sector.

- **More (quality) data will be available more easily and timely**, leading to a range of other benefits as are mentioned below.
- **More cost-effective spending and efficiency gains**. Some stakeholders have pointed out that there could be substantial cost-effective spending and efficiency gains due to access to privately-held data. From the data collaborative between LUCA and Highways England, which provided the latter 24/7 **access to mobility insights** based on a specific full year of data, it could be seen that there are valuable insights **for the modelling and infrastructure planning** of England's motorways and major A roads. This collaboration reduced the data-collection time period from 6 months to 7 days, which is a massive savings in labour hours, which resulted in a saving of millions of pounds each year on data-collection costs.[371]
- According to stakeholders interviewed, there is a potential reduction of costs after acquiring data from the private sector. For instance, it was estimated by a public-sector stakeholder that acquiring data for the calculation of their CPI from diverse companies, allowed them to reduce their annual costs by €2.4 million (or the equivalent of 30 FTEs). If we assume that a similar benefit could be achieved by the statistical offices in all EU Member States, there could a potential cost-saving of up to €64.8 million across the EU thanks to the access to privately-held data for the calculation of the CPI.
- This policy option would also allow data re-users with their **enforcement activities**. For instance, one stakeholder interviewed pointed out that one big challenge they face is the illegal short-term rental activity. Due to collaboration with a company they were able to remove illegal listings by 90% in one month.
- Another potential benefit of access to data is for **tourism management purposes**. Cities across Europe see high rates of tourism every year. Barcelona is the fourth most-visited European city. However, Barcelona's tourism industry leads to very serious impacts and conflicts for the local society and the environment.[372] A study on the impact of short-term rental platforms on the housing market points out that these short-term accommodation platforms have increased both rents and prices. For rents, the study suggests that 54 more active listings in a small neighbourhood (about the average level in 2016) increase rents by 1.9%, while transaction and posted prices increase by 5.3% and 3.67%, respectively. However, estimates imply that local impacts can be substantial in the most tourist parts of the city. In particular, 200 listings (the average number of listings in 2016) increase rents by 7% and transaction and posted prices by 19% and 14%, respectively. In this sense, short-term accommodation platforms reduce the supply of residential housing units and thus, reduce the number of resident households in the neighbourhood. As a result, due to the housing affordability problem, locals are forced to move out to more affordable neighbourhoods. This brings societal and cultural problems for the city.[373] Access to privately held data could potentially benefit cities that see these tourism-related issues

[371] https://digital-strategy.ec.europa.eu/en/news/good-practices-b2g-data-sharing-mobility-data-operating-englands-highways-infrastructure

[372] https://stay-grounded.org/barcelona-a-city-exploited-by-tourism/

[373] IEB, 2019. Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona. See https://ieb.ub.edu/wp-content/uploads/2019/07/2019-IEB-WorkingPaper-05-1.pdf

and adapt their policies to have a better tourism management and achieve a better urban equilibrium.

This policy option would also most likely result in a **more data-informed decision making and overall better policies due to increased insights, access to new evidence and access to high-quality data**. Re-using relevant private-sector data would increase the public sector's ability to understand, assess and predict different situations and phenomena affecting the society. It enables logical and fact-based decisions in a timelier and more flexible manner. It also allows to prepare for scenarios like a financial crisis, environmental disasters, urban planning, crimes, to mention some. It creates a solid foundation for decision making and for strategic regulatory initiatives and new policies.

The above examples show that B2G data sharing can have benefits for public administrations when re-using private sector data as part of the public service delivery, statistical offices and highway authorities spend far less on data collection and cities could significantly increase the effectiveness such as the case of enforcement of illegal listings in Barcelona. Such benefits depend on the actual B2G use case and specific circumstances in which the data is being re-used. The CPI use case clearly provides 2.4 million per year in benefits, this is a significant saving. For other use cases no quantifications of such efficiency benefits are available, making it difficult to directly extrapolate potential savings for public sector. However, similar to the cost savings estimate from improved availability and use of Open Data within government[374], it is likely that better use of private sector data for public service delivery is also likely to result in such savings. While the cost savings linked to Open Data include savings linked to the management of public sector data and IT costs, these also include efficiency gains in public service delivery (e.g. fewer manual workflows, less errors, shorter case processing times, improved control (e.g. fraud detection), etc.).[375] Hence, it is reasonable to assume that B2G data sharing will similarly result in efficiency gains across the public sector at national and local levels. Following the logic of the cost calculation presented above in Table 64, if one assumes 459 national public administrations (e.g. ministries, statistical offices, central banks, etc.) and 1208 local administrations (e.g. cities and local authorities) across the EU are involved in a total of 30 B2G use cases, each could incur some saving in terms of efficiency gains. It is no likely that each B2G use case and each data re-user will save EUR 2.4 million a year as per the estimated cost savings of the statistical office in the CPI use case. However, if one would assume average efficiency gains amounting to EUR 50.000 for national authorities and EUR 20.000 for local authorities the potential savings could amount to EUR 337 million across the EU. While actual cost savings will be specific to each B2G use case, it is likely that such benefits will be reaped. With the assumptions taken here, these benefits would account for 90% of the estimated costs for the public sector of policy option 2 (EUR 371.2 million as per Table 63).

Other benefits that were identified for data re-users are: cost savings from datasets that could become free of charge, below market price or donated (depending on the use-case); increased capacity to leverage alternative data sets that can be repurposed across governmental institutions; improved performance management, monitoring, and accountability, likely increasing the trust of organisations and individuals on the public sector; positive impact of increased data sharing and reuse on AI and machine-learning in Europe, with security and strategic autonomy implications; promotion of "secure private computing" and other privacy enhancing technologies in the business

[374] Study to support the review of Directive 2003/98/EC on the re-use of public sector information (SMART 2017/0061).

[375] Good basic data for everyone – a driver for growth and efficiency (Digitaliseringsstyrelsen, 2012). See: https://en.digst.dk/media/18773/good-basic-data-for-everyone-a-driver-for-growth-and-efficiency.pdf

sector; smaller cities no longer disadvantaged compared to large cities, resulting in increased fairness.

This policy option (as well as policy option 3), would also likely result in societal and environmental benefits. While there is no clear estimate about the extent to which B2G could reduce costs, improve efficiency or help tackle certain societal or environmental issues, based on what stakeholders mentioned during our interviews, we assume that 1% for costs reductions, or increased efficiency seems to be a consensus number to calculate the extent to which B2G could actually bring as benefits. Moreover, the societal and environmental benefits identified are examples of use cases of B2G data sharing and benefits obtained from specific data collaboratives. Identified examples include:

- **Environmental impact and urban planning**: According to a study made by INRIX, a location-based data analytics company, in Germany, **the costs of congestion** to the city of the Munich annually due to congestion amounted to 3.1 billion USD (equivalent to approximately 2.5 billion euros) in 2017, while in the city of Berlin, the costs were more than double, amounting to 7.5 billion USD (equivalent to approximately 6.2 billion euros). If we assume that due to access to privately-held data, these cities could have a better understanding of the time and reasons for agglomeration in specific zones inside the cities resulting in better urban planning, there could be a potential benefit of reducing the costs of congestions for the cities. If we assume that 1% could be saved annually as a result of decreasing the congestions, approximately 25 million euros a year could be saved in the city of Munich and 62 million euros annually in Berlin. In addition, according to a French Senate study[376], it was estimated that the annual cost due to air pollution is 101.3 billion euros. Making use of B2G data access and reuse practices, there could be a **faster and more targeted response to such environmental challenges which reduces the costs from inaction or non-targeted responses.** This cost includes health damages from pollution and the impact on building refurbishing, ecosystems and agriculture. In this sense, better urban planning could have a positive impact on not only a reduction of congestions in cities but also a reduction of costs incurred when repairing the damages caused. In another study by INRIX in 2016, the (economic) cost to drivers in Europe (across 18 Member States) of time wasted in congestion across traffic hotspots identified was estimated to around 137 billion EUR by 2025. Assuming that time savings could be generated due to B2G, roughly 14 billion EUR time savings per 1% reduction in congestion could be achieved annually with regard to the main congestion hotspots within the EU.
- **Climate change and resilience**: data is considered essential for achieving the urgent goals of reduction in carbon emissions and increasing the efficiency of natural resources and materials[377]. Data from insurers, for example, on damage to buildings, infrastructure and agriculture can help decision-makers take informed decisions to improve the resilience and adaptation capacity, and inform SMEs on decisions of where to set up business. Early warning systems might benefit from the use of data generated by location and population-based communication technologies and social media. [378] Marine data held by commercial actors if made available could help avoid the need for duplication in data gathering and enable better assessments of the impact of climate change and human activity on marine ecosystems, and to reduce the ecological footprint of economic activity in the ocean.[379]

[376] https://www.industrie-techno.com/article/cout-de-la-pollution-la-start-up-plume-labs-veut-democratiser-le-quantified-environment.39198

[377] Eg European Commission, Forging a climate-resilient Europe - the new EU Strategy on Adaptation to Climate Change, COM/2021/82 final

[378] https://www.eea.europa.eu/publications/climate-change-adaptation-and-disaster/at_download/file

[379] https://www.oecd-ilibrary.org/docserver/a4734a65-en.pdf?expires=1623673818&id=id&accname=guest&checksum=D93D080762A4321B6865D97ED3A36AD6

- **Health emergencies**: during health emergencies such as COVID-19, Ebola, Zika virus and swine flu, the use of, for example, mobility data can bring potential benefits to understand, monitor and control the development of infectious diseases. Global mobile phone penetration rate reached 96% in 2014. In the EU-27 there were, on average, 1 220 mobile phone subscriptions per 1 000 inhabitants in 2018 or 1.2 mobile subscriptions per person.[380] In this sense, the engagement of network operators has resulted in population movement analyses based on call data records (CDRs) that have been particularly promising for improving responses to disasters. During **the swine flu outbreak in 2009,** in an attempt to **understand more about the spread of epidemics through society**, one large telecommunications provider reported using anonymised call detail records to estimate the number of people visiting certain locations and draw inferences as to the link between public health interventions, population mobility and infections number.[381] Seeing the various cases where public and private sector have collaborated for the purpose of facing a health emergency, it can be seen that B2G data collaboratives can prepare and equip the governments with a more targeted and timelier response that may potentially save lives. As seen by the data collaborative between The Data Lab, UNICEF and the Scottish government with owners of data sources such as shopper data, TV adverts, online gaming, and school lunch suppliers, **privately-held data can also be used to tackle child obesity in a country**.[382] According to a study data suggests there are almost 398,00 children aged 6-9 years living with severe obesity, out of around 13.7 million children living in the 21 European countries that participated in the study.[383] Using data to predict, inform and understand the diverse factors impacting obesity in children such as advertising, could allow for better policies that reduce child obesity. If we assume that this collaborative would be implemented in the EU, reducing child obesity by 1%, approximately 3,900 children would have better eating and exercise patterns, that could reduce their health risks in the long term.

These examples show the potential of B2G data sharing linked to specific domains and use cases. The extent to which such benefits would materialise in the EU depends on the actual uptake of specific use cases, level of data sharing and use of the data for improved insights for better decision making. Hence, more generally, it is likely that society would benefit from better policies and better decision-making based on evidence-based policy making and more efficient/flexible public service delivery.

3.3.1.4.2.3 Findings of the Cost-Benefit Analysis

Table 65 - Overview of benefits - Business-to-Government (B2G) data sharing for the public interest | PO 2

Type of action	Description	Amount (EUR)	Stakeholders
Measures facilitating secondary use of data held by the private sector	**Direct benefits**		
	Better business decision-making resulting from improved business ecosystem, better public services delivered, and also access to analyses and	*Not quantifiable due to lack of data* Specific to the use-case, it was estimated 1% of total annual revenue would be at risk without analyses/insights	**Data holders**

[380] https://ec.europa.eu/eurostat/statistics-explained/pdfscache/32183.pdf

[381] https://www.theguardian.com/media-network/media-network-blog/2013/sep/05/combating-epidemics-big-mobile-data

[382] https://www.thedatalab.com/news/unicef-and-the-data-lab-call-for-data-collaboration-to-tackle-child-obesity-in-scotland/

[383] https://www.euro.who.int/__data/assets/pdf_file/0019/400654/COSI-Severe-Obesity-FS-ENG-LowRes.pdf

	insights not available before		
	Reduction of administrative burdens and time/resources saved when establishing data partnerships	*Estimated costs reduction for data holders amounting to 76 million euros (one-off) and 78 million euros (recurrent).*[384]	**Data holders and data re-users**
	Reputational benefits	*Not quantifiable due to lack of data*	**Data holders**
	Access to other datasets, domain expertise, analysis methods and models, which could improve the value of own data	*Not quantifiable due to lack of data*	**Data holders**
	Compensation and new revenue streams	*Not quantifiable due to lack of data*	**Data holders**
	Reduction of legal risks when entering into B2G data partnerships	*Not quantifiable due to lack of data*	**Data holders**
	Increased maturity, capacity and discovery of data infrastructure that can be repurposed for other goals	*Not quantifiable due to lack of data*	**Data holders**
	More (quality) data will be available more easily and timely	*Not quantifiable due to lack of data*	**Data re-users**
	More cost-effective spending and efficiency gains	*Not quantifiable due to lack of data* *Specific to the use-case, reductions by 50% in time/resources normally spent by statistical offices for data collection.*	**Data re-users**
	Data-informed decision making and overall better policies due to increased insights, access to new evidence and access to high-quality data	*Not quantifiable due to lack of data*	**Data re-users**
	Cost savings from datasets that could become free of charge, below market price or donated	*Not quantifiable due to lack of data*	**Data re-users**
	Improved performance management, monitoring, and accountability, likely increasing the trust of organisations and individuals on the public sector	*Not quantifiable due to lack of data*	**Data re-users**

[384] Please refer to table 8 and the exercise on reduction of admin burden for further details.

Increased capacity to leverage alternative data sets that can be repurposed across governmental institutions	*Not quantifiable due to lack of data*	**Data re-users**
Face and prevent current environmental challenges in an efficient and data-informed way. This would also reduce costs from delayed or lack of targeted responses.	*Not quantifiable due to lack of data* *Example: the UK estimated that if geospatial and mapping data, which is vital to a country's economy would be shared to public sector organisations, it could generate economic and social value for the country between 2 billion GBP and 14 billion GBP annually.*	**Society**
Faster and more targeted response to emergencies and to societal challenges	*Not quantifiable due to lack of data*	**Society**
Indirect benefits		
Positive impact of increased data sharing and reuse on AI and machine-learning in Europe, with security and strategic autonomy implications	*Not quantifiable due to lack of data*	**Data holders and data re-users**
Promotion of "secure private computing" and other privacy enhancing technologies	*Not quantifiable due to lack of data*	**Data holders and data re-users**
Increased trust in government as a result of improved public service delivery and evidence-based policy making		**Society**

3.3.1.4.3 Coherence of the option

No incoherence of this option with existing legislation was identified.

3.3.1.5 Policy Option 3: High-intensity regulatory intervention

This section assesses the second policy option for Business-to-Government (B2G) data sharing for the public interest.

3.3.1.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the policy objectives.

3.3.1.5.1.1 Achievement of specific objectives

This policy option would contribute, like the lower-intensity regulatory intervention, to achieving the general and specific objectives. Promoting the implementation of a data steward function in all organisations over a certain size would allow for both the private and public sectors to be able to know which data is needed and how. This would allow both parties to be able to speak the same language when it comes to data access and reuse practices. As mentioned in policy option 2, setting a national structure and establishing a set of B2G data access/reuse practices would bring more certainty by establishing a more harmonised approach and framework for this B2G data access and reuse.

The **three specific objectives** of this intervention, would also most likely be reached. It would most likely **improve access to and use of (big) data sources** held by private companies, which would help tackle societal challenges, and improve policymaking and delivery of public services in a more flexible manner. This policy option would also allow the development of a framework for responsible access to and use of such data sources. This would contribute to **establish clear procedures and structures, and to have a harmonised approach to B2G data access and reuse**, facilitating decisions on use of private sector data by government. As a result, this could strengthen the internal market.

3.3.1.5.1.2 Achievement of general objectives

The general objective of this intervention would likely be reached. **A fair balance between protecting the interests of the private sector and the rights of the data subjects, where personal data is concerned, and unleashing the full potential of private sector data for the benefit of society would likely be achieved**. By making it compulsory for the private sector to make their data accessible for the public good under preferential conditions, this policy option would contribute to reaping the potential of private sector data to help tackle societal challenges in a more efficient and effective way. It would likely **c**ontribute to **modernise public services**, and **accelerate innovation** as data is more widely used for the **common good**.

3.3.1.5.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

3.3.1.5.2.1 Costs of the option

Under this policy option, we identified the same costs as under policy option 2, therefore the costs in table 10 for data holders (one-off) and the costs of a national structure are the same. Similar to Policy option 2, the same scope of data holders and data re-users are considered for our estimations.

There would be additional costs regarding the implementation of the **data steward function for data re-users**, which under this policy option would be an obligation to designate it in all **public and private sector organisations over a certain size**. The same estimations were made as with policy option 2. Our calculations for this cost were based on interviews with stakeholders. The difference between the calculations of this cost are that here there is a 100% compliance due to the compulsory element of designating this function for all public-sector organisations over a certain size. This difference would amount to costs of 314.76 million euros annually at the EU level for the data steward function. The **costs for audit and verification procedures to ensure veracity of results and independence of public-sector action** are the same as policy option 2 of 192.17 million euros. Both costs, therefore, would be of 506.9 million euros annually for data re-users.

For **data holders**, the one-off costs represented in the table are the same as policy option 2, since there is a need to make an initial investment to create a data infrastructure(s) that allows access to data in accordance with GDPR and other regulations. However, the recurrent costs calculated vary form policy option 2 since this policy requires that all private sector organisations over a certain size **designate a data steward function**. This would cost at the EU level 68.3 million euros. We assume that this cost counts, in comparison to policy option 2, as of 2024 since the data holders are not part of the decision-making process at the national structure and thus, the need for a data steward function may come at a later stage when the B2G access and reuse practices scale up.

The costs for identifying the data that would be valuable for the public interest after a public administration's request and the categorization and normalisation of datasets is the same as policy option 2, 78.06 million euros. Both costs, therefore, would be of 146.4 million euros annually for data holders at the EU level.

This policy option would also likely incur opportunity costs of not selling the data but rather providing it free of charge or with only the marginal costs for dissemination covered to data re-users (depending on the use-case).

The extrapolation of these figures to the EU level can be found in the table below, while the full overview of costs is in Annex II.

Table 66 - Overview of costs for Business-to-Government (B2G) data sharing for the public interest | PO 3

Overview of costs (EUR) – PO 3							
		Data holders		**National structure**		**Data re(users)**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures facilitating the use of data held by the private sector	*Direct costs*	552.5 million	146.4 million p.a.	-	21.6 million p.a.	-	506.9 million p.a.
	Indirect costs	-	-	-	-	-	-

3.3.1.5.2.2 Benefits of the option, including cost reductions and other positive impacts

This policy option would likely incur the same benefits as those identified in policy option 2 for **data re-users.** In addition to these, since there would be an obligation for the private sector to share data for the public good under preferential conditions, **this would result in time savings and cost reductions for the public sector institutions**. **There would be a decrease of time and transaction costs when establishing the partnerships, and more opportunities to access quality data which could be free of charge for very high public interest purposes or with only marginal costs for the dissemination of this data for high public interest purposes.**

For instance, a stakeholder interviewed mentioned the differences between 2000 and 2021 regarding their statistical methodology to calculate inflation and consumption patterns. In this sense, due to **access to scanner data it became less costly and more efficient to obtain the data.** In 2000, approximately 7000 shops were visited by a team composed of 29 FTEs in order to obtain more than 2.7 million prices of products in those stores. This resulted in annual costs of 2.32 million euros (it cost 80,000 euros annually per FTE), only to retrieve pricing data. These costs are currently completely eradicated. There are no FTEs required to obtain this data anymore, and access to this data is free of charge due to the current legislation in the country that allows this. In addition, the quality of the data is better and more reliable because the information comes directly from scanner machines whose prices are recorded in databases.

Moreover, according to another public-sector stakeholder, it currently costs them 11-man years (annually) to have access to the data that allows them to calculate inflation. If they would have access to scanner data, it would probably cost them 2 to 3-man years. For household-spending surveys, they believe that credit card data would allow them to save approximately 300,000 to 500,000 euros per survey (which is done every 5 years).

The obligation to create a **data steward function** in all public and private sector organisations over a certain size **could also bring benefits for data holders and data re-users.** These could include **time saved when not having to spend too much time finding the right person to talk to inside the private companies for the creation of data collaboratives.** Moreover, a public-sector stakeholder interviewed believes that the **biggest benefit of such a function within their institution is that this person opens new possible ideas and identifies new ways on how to cooperate with private sector companies**. Another stakeholder interviewed, mentioned that having this data steward function in their organisation and also on other public sector organisations, brings a positive impact in terms of knowledge creation regarding the use of data.

The table below provides a full overview of benefits identified in policy option 3, in addition to those identified in policy option 2.

3.3.1.5.3 Coherence of the option

No incoherence of this option with existing legislation was identified.

3.3.1.5.4 Findings of the Cost-Benefit Analysis

While the quantification of benefits differ on a case-by-case basis and therefore a conclusion in terms of benefits is not possible to fully execute a cost-benefit analysis, as demonstrated during the evaluation of policy options 2 and 3, there are potential societal, environmental and economic benefits for private and public sectors (in terms of costs savings, efficiency gains) derived from a more structured and harmonised approach that incentivises business-to-government data sharing use cases. The table below aims to provide a summary of the possible impacts of the policy options.

3.3.1.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 67 – Summary of impacts for Business-to-Government (B2G) data sharing for the public interest

Type of impact	
Economic impacts	• **Costs for public sector (data re-users)** ○ *Direct/Indirect:* - Audit/verification procedures - National structure investment and operational/maintenance costs - Compensation to private sector - Data management - Set up larger data processing/analysis capacity - Governance structure including data steward function. - Potential impact of using data that is not representative to guide decision-making, which may result in lack of transparency on the methodology used to collect the data. The use of non-representative data could have a significant impact

on the costs and benefits derived from the B2G data sharing partnership.

- **Benefits for public sector (data re-users)**
  - *Direct/Indirect:*
    - More quality data made available more easily and timely
    - Free-of-charge, below-market-price, donated datasets depending on use case
    - Decrease of time and transaction costs to establish partnerships with data holders
    - Increased capacity to leverage alternative datasets
    - Cost-effective spending and efficiency gains
    - Improved performance management
    - Monitoring and accountability
    - Improved public service delivery and policies
    - Prediction accuracy to prevent crises and faster and more targeted crisis response
    - AI and machine learning in Europe
    - Promotion of secure private computing and other privacy enhancing technologies
    - Cost savings and efficiency gains due to data steward function (depending on the policy option and whether under policy options 1 and 2, the organisations would follow the recommendation of having this function).
    - Smaller cities no longer disadvantaged compared to large cities, resulting in increased fairness

- **Costs for private sector (data holders)**
  - *Direct/Indirect:*
    - Costs of normalisation and making datasets available for reuse
    - Costs of formalising partnerships and including contractual arrangements
    - Data steward function (hiring, training and salary)
    - Costs of cataloguing and identifying data that can be made available
    - Internal data governance set up
    - Loss of competitive advantage based on data scarcity.

- **Benefits for private sector (data holders)**
  - Direct/Indirect:
    - Revenue and compensation for making data available
    - Improved value of own datasets and access to research/analysis insights
    - Access to analytical methods and models previously not available
    - Reduced administrative burden due to centralization of requests

	- Public reputational benefits vis-à-vis potential customers and public sector and increased trust of contributing to public interest - Better business decisions resulting from knowledge/insights shared - Removal of legal risks and burdens - Indirect benefits from improved business context as a result of improved decision-making, emergency response, and public service delivery - Increased maturity, capacity and discovery of data infrastructure that can be repurposed for other goals - Decreased legal uncertainty due to creation of dispute settlement mechanism - Fairer competition, due to relative harmonisation of B2G data access rules and practices in the EU - Access to other data sets or domain expertise otherwise not available or at higher cost - Cost savings and efficiency gains due to data steward function
Social and Environmental impacts	• Better policy-making outcomes resulting from better and data-informed decision-making. • Faster and more targeted response to emergencies and to societal challenges. • Faster/better recovery from emergencies/disasters/crises • More efficient/flexible public service delivery

3.3.2 Measures supporting citizen empowerment ('human-centric data economy')

This section presents our draft assessment of the impacts of all the options, including the baseline scenario.

The impact of the proposed policy options is assessed using the following criteria:

- On the effectiveness side of policy option, we will look at how the intervention helps in achieving the general and the specific objectives assumed.
- On the efficiency side, we will try to assess both costs and benefits of the options for the stakeholders.

Coherence of the options will also be considered in the assessment of the policy options' impact as well as the proportionality and legal/political feasibility criteria will be also considered when comparing the policy options.

To the extent possible, the assessment is built on quantitative and qualitative information, including costs and benefits. For this purpose, we took various data sources into account for the assessment of the impacts, including:

- Desk research, including a legal analysis;
- Interviews.

3.3.2.1 Stakeholders affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Table 68 – Overview of stakeholders affected by Measures supporting citizen empowerment ('human-centric data economy') policy options

	Costs	Benefits
Consumer (co-producer)		Greater choices, better services
Data holder	Implementing strong consent management solutions Provision of API solutions Greater competition	Greater array of services provided to users Premium services to re-users of personal data
Competitor of data holder	Implementing strong consent management solutions	Possibility to enter new market
Data re-user	Implementing strong consent management solutions	Possibility to enter new market
PIMS or other intermediaries	Implementing strong consent management solutions	Possibility to enter new market New data sources for analytics and machine learning
Researchers	Implementing strong GDPR compliance mechanisms or consent management solutions	New data sources for analytics and machine learning

For customers will benefit from more and diverse service options, at competitive prices, reducing the "locked-in" effect on specific service providers and device manufacturers. For data holders, the benefits will be reflected through the improvement of customers' trust due to the implementation of strong management solutions, stimulating the innovation process and lower production costs, on the long-term.

The data (re)users will benefit from an improved uptake of data portability as the market competition increase will enhance further innovation developments. The new business models and initiatives for data portability will emerge, contributing to the settlement of complementary markets. Developing and setting up standards for data portability and security will also help improve customers trust and, reduce the transactions' costs, in the long-term.

Overall, society would benefit by the improvement of societal and environmental outcomes, enhancing the potential benefits of citizens' data use for the EU economy and society, within the limits of privacy and data protection regulations.

Further research will generate new insights, leading, in theory, to more effective and efficient decision-making in various domains, such as health, social affairs, transport and the environment.

3.3.2.2 Policy Option 0: baseline scenario

3.3.2.2.1.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

3.3.2.2.1.2 Achievement of specific objectives

Without any intervention, the baseline scenario is developed based on the current provisions: the application of the article 20 of GDPR. In this case, data portability is just a right that it is not yet widely exercised.

3.3.2.2.1.3 Achievement of general objectives

The lack of clarity and lack of provision significantly hamper the uptake of data portability by both users and industries. In this context, reaching any of the objectives proposed might be just a wish without a precise timeframe to be achieved.

3.3.2.2.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with a baseline scenario.

3.3.2.2.2.1 Costs of the option

There are no specific costs linked to this option.

3.3.2.2.2.2 Benefits of the option, including reductions in some of the costs as well as other positive effects on (some of) the stakeholders

There are no specific benefits linked to this option.

3.3.2.2.3 Coherence of the option

This policy option does not entail any piece of legislation which might be incoherent with other policy options.

3.3.2.3 Policy Option 1: Non-regulatory intervention

This section assesses the first policy option for Measures supporting citizen empowerment ('human-centric data economy').

3.3.2.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the policy objectives.

3.3.2.3.1.1 Achievement of specific objectives

This policy intervention will have a rather limited impact in the achievement of the objectives set up. The intervention aims to foster the development of a market for data portability, through a set of non-binding recommendations. The initiative helps addressing several of the operational objectives, such as lack technical interoperability, the uncertainty and lack of trust.

Because of its voluntary nature, it is expected to provide better cost-benefit ratio, since it is likely to be adopted by companies with lower compliance costs and higher expected benefits.

3.3.2.3.1.2 Achievement of general objectives

It will contribute the development of the overall system, but it will not address the fragmentation, due to its sector-based nature. The approach might have positive results at sector level, for example, for smart home appliance, if the degree of adoption is relatively high. Considering the non-binding nature of the measures and the unpredictable level adoption by the market, the results in achieving the specific and general objectives remain limited. However, this initiative can establish a good starting point for further developments, and help setting up the conditions for a broader regulatory intervention.

Underlying these measures, there is the need to reinforce the capacity to monitor and support to data portability. As there is no one size fit all solution, there is the need to develop custom made solution and guidelines for different situation.

3.3.2.3.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 1.

3.3.2.3.2.1 Costs of the option

Several types of cost can be identified, for both data holders and data (re)users. Also, these types of costs affect data holders and (re-)users differently, depending on their preparedness and also the underlying conditions existing in their respective sectors. The most common types of cost encountered are:

- The costs for technical interoperability compliance: low to medium-high impact. Setting up data interoperability standards could be costly if the digital readiness of the systems is relatively low. Data standardisation could end up being an expensive process, depending on the volume, frequency and quality of the data involved. In addition, setting up data security protocols will add another layer of costs that could end up increasing to a high-level impact for compliance for both data holders and re-users.
- The costs of setting up APIs: low to very high impact. Setting up APIs can be a costly process too. The costs can vary significantly, from low to very high, depending on the number of APIs, their scope and complexity. Additional to initial costs of setting up the APIs, maintenance costs (update and upgrade) should also be considered, as complex built APIs have higher annual maintenance costs. Depending on the type of infrastructure and technical characteristics, some recent studies have estimated that the costs of establishing APIs range between 30 000 euro and 2.5 million euro.[385] On average, the cost for setting up an API was estimated around 50.000 euro.[386]
- The costs for developing guidelines and codes of conduct: limited to low impact. These types of costs are often on the lower side of the scale, and their overall impact remain relatively low.

For the two sectors in scope of the proposals - wearables and smart home appliance, the cost impacts might vary significantly between products and service providers. If we were to compare, the wearable data market is much more advanced and standardised than the one of smart home appliances that has higher level of fragmentation in both data standards and interoperability.

The preliminary cost-benefit analysis shows similar level of costs between for the data holders in both sectors in scope. However, significant differences of costs can be seen for data re-users in the two sectors. In the smart home appliances sector the exemption rule applied to micro- and small enterprises (with less than 50 employed persons) impacts significantly the average costs borne by data re-users. In this context, it is assumed that most costs will be borne by larger data re-users rather than small ones (which, however, make up to 99% of the data re-users in this sector).

Table 69 Overview of costs | PO 1 (smart home appliances and fitness trackers)

Overview of costs (EUR) – PO 1 (smart home appliances)				
	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	47.8 million	13.6 million p.a.	2.1 million	0.7 million p.a.

[385] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

[386] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

Indirect costs	-	-	-	-

Overview of costs (EUR) – PO 1 (fitness trackers)

	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	41.8 million	11.9 million p.a.	27 million	9.1 million p.a.
Indirect costs	-	-	-	-

However, since the option proposed is a non-binding one, the above-mentioned costs will incur only to the data holders willing to adopt it. In this sense, estimating the costs of initiative’s adoption remains limited and contains a significant degree of uncertainty.

3.3.2.3.2.2 Benefits of the option, including cost reductions and other positive effects on (some of) the stakeholders

Based on both desk research and results of the interviews, we can say that this policy option could bring some benefits at the EU level, but it may take a long time before some significant changes to appear.

Several types of benefits can be also identified, for each of the sectors in scope, for both data holders and data (re)users.

For fitness trackers, the identified benefits are:

- Increase the customers’ choices and mitigate the “locked-in” effect by easily switching between devices and services. However, the switching costs remain uncertain as adoption of the measure is done voluntarily, and providers are not obliged to do it.
- Develop new or improved services and products for customers.
- For companies and service providers, developing new or additional service by integrating their collected data with data from other sectors will generate new opportunities and help establishing new markets.
- New providers can join the market and increase its competitiveness, enhancing the innovation level for both well-established and new market providers.

For smart home appliances, the identified benefits are:

- Increase the efficiency of home appliances, in both energy consumption and functionalities offered. This will also bring savings benefits to customers, as an efficient energy consumption reduces costs for the users.
- Extend the home appliances’ life, by developing the predictive maintenance options. This contributes also to reduce the environmental impact, by reducing industrial waste.
- Opening up the repair aftermarkets to independent providers, and increasing the competition on the market. At the same time, it offers customers a wider choice of service providers for home appliances’ repairing.
- Better and more diversified products and services, including the ones connected to smart homes that could contribute in lowering the environmental impacts of consumption.
- Emergence of new players and new markets that enhancing innovation developments for all market players.

Table 70 Overview of benefits | PO 1 (smart home appliances)

Description	Amount (EUR)	Stakeholders
Direct benefits		
Additional revenues due to more market opportunities	*12.6 million p.a.*	**Data holders**
	94.3 million p.a.	**Data re-users**
Cost savings for consumers	*0.189 million p.a.*	**Consumers**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

In 2018, the annual market revenues from the smart home appliances were estimated at 2625 million euro.[387] Based on Eurostat data for the number of enterprises in the manufacturing of domestic appliances (C275) sector, we estimated the revenues per company at approximately 855484 euro. We used this value as the base to calculate the potential additional revenues for the companies participating to the voluntary scheme, both data holders and data re-users. However, due to the limited market maturity, the level of this additional benefits is not expected to be very high and it will be dependent on the level of the market adoption. In this context, we estimated around 12.6 million euro for data holders and 94.3 million euro for date re-users in additional benefits, on annual basis.[388]

Consumers will also benefit, especially from energy savings due to more efficient products and services. In 2018 the revenues from energy management in the smart homes in Europe was estimated to 1.35 billion euro.[389] The energy consumption of lighting and appliances accounts for 14% of households' energy consumption.[390] We have assumed that data portability will result into an additional one percent of savings for the customers. However, due to the voluntary nature of the measure, the impact will be weighted by the companies' rate of adoption. In this context, the amount expected was around 189000 euro.

Table 71 Overview of benefits | PO 1 (fitness trackers)

Description	Amount (EUR)	Stakeholders
Direct benefits		

[387] Home Appliance Europe (APPliA), Annual Report "By the Numbers: The Home Appliance Industry in Europe, 2018-2019", 2020.

[388] Only 3% additional benefits for data holders and 4% for data re-users.

[389] Home Appliance Europe (APPliA), Annual Report "By the Numbers: The Home Appliance Industry in Europe, 2018-2019", 2020; and stakeholder interview.

[390] Eurostat data on energy consumption in households.

Additional revenues due to more market opportunities	*116.6 million p.a.*	**Data holders**
	133.6 million p.a.	**Data re-users**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

The fitness (wearables) market is much more mature compared to smart home appliances. In 2021, the revenues from the fitness sector are expected to reach 3310 million euro. [391] Also, the same time, the market is expected to grow at 21% annual growth rate.[392] Considering the market maturity, the additional revenues were estimated around 116.6 million euro for data holders and 133.6 million euro for data re-users.[393]

Seeing the voluntary nature of the proposed option, many of these benefits will remain rather limited, and dependent on the level of adoption by the product and service providers. Also, as some of the initiatives will be sector specific, inter-sectoral data exchange will also be relatively low or limited.

3.3.2.3.2.3 Findings of the cost-benefit analysis

Overall, the option presents relatively moderate costs, while the benefits are mainly concentrated in companies more likely to gain. Both digital readiness and added reciprocity clause play an important role in the participation to this initiative. The addition of the reciprocity clause is both an incentive and a disincentive for participation. It is expected to bring an increase in benefits for participants, but it comes with an increase of compliance costs and the rule of sharing own data.

These conditions determine a selective participation, as ones that will join the initiative will be those data holders and re-users for which the perceived benefits are well above the costs required. Considering the two sectors in scope – smart home appliances and fitness trackers, it is expected that the impact of this option to differ between the two sectors.

The smart home appliances sector is far less digitalised compared to fitness trackers sector. Therefore, for smart home appliances, the participation rate in this voluntary scheme will remain rather limited. The costs for technical and interoperability compliance will be high, as data standardisation and portability in this sector is limited. However, since it’s a voluntary scheme and it will be used mostly by those who are better prepared from technological perspective, the level of costs could be kept on a lower side as to not overpass the benefits levels. Benefits will be higher especially due to the reciprocity clause, and possible exemptions from this clause that might apply for small businesses (with less than 50 employees).

In the case of the fitness trackers, the costs for technical and interoperability compliance will have more of a medium to low impact, as the digitalisation level in this sector is much more developed

[391] https://www.statista.com/outlook/dmo/eservices/fitness/europe

[392] https://www.marketdataforecast.com/market-reports/europe-fitness-trackers-market

[393] The sector benefits of a higher rate of adoption than the one for smart home appliances; 20% additional benefits for data holders and 30% for data re-users.

compared to smart home appliances. As in the previous sector, due the voluntary nature of the initiative, only participants with higher cost-benefits ratio will likely join. With a more mature digital market, the benefits are higher and amplified also by the reciprocity clause for both data holders and data re-users.

3.3.2.3.3 Coherence of the option

Being a non-regulatory initiative, the option does not require any specific legal regulation which might be incoherent with other policy options.

3.3.2.4 Policy Option 2: low-intensity regulatory intervention

This section assesses the second policy option for Measures supporting citizen empowerment ('human-centric data economy').

3.3.2.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the policy objectives.

3.3.2.4.1.1 Achievement of specific objectives

The second policy option gets one step further in fostering the development of a market for data portability. It addresses several of the operational objectives, such as lack technical interoperability, the uncertainty and lack of trust, through soft regulatory measures for product and service providers.

The measures proposed will help improve the development of the overall system and address partially the fragmentation level within certain sectors, such as the home appliances one. Regulatory character of the measures might also help improve the uptake data portability by users and data holders.

3.3.2.4.1.2 Achievement of general objectives

By setting up clear rules for data portability and access, such as the number of data points to be made available, the format and the frequency, as well as the tools to use, the initiative contribute in achieving some of the specific and general objectives. It will enable the development of models and initiative for data portability based on standards, allowing new players to join the market. More players in the market increase competition and enhance innovation that generates more benefits for the end-users.

However, allowing access to only accredited re-users and only for direct service provisions narrows down the full potential that data portability can offer. Additionally, cost provisions can also reduce accessibility as high data costs will reduce the re-use attractiveness to third-party providers. While FRAND is also one of the recommended options, the cost to access the data for re-use is an important factor in making the policy work properly. The limited set of data points and the lack of real-time data also hamper the potential of development of new products and services.

Overall, this low-intensity intervention impact remains rather limited, achieving only partially the objectives.

3.3.2.4.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 2.

On the efficiency side, this policy option has both positive and negative effects. On the positive side, the costs of the measures proposed remain relatively accessible, as these costs will apply only to a limited set of data. This option could significantly reduce the potential impact of costs, depending on the selected data points, and provision formats developed for access and portability.

On the negative side, a limited set of basic data points, with delayed access, offers less opportunities for development of new services and products. In the same context, access to data under strictly regulated conditions can also hamper innovation and limit the development of complementary services and products.

3.3.2.4.2.1 Costs of the option

Similar to the case of the non-regulatory option, there are several types of cost that can be identified, for both data holders and data (re)users. And, their impact on both data holders and (re-)users varies, depending on their preparedness and the overall conditions of the sector's data market. The most common types of cost encountered are:

- The costs for technical interoperability compliance: low to medium impact. While, costs for setting up data interoperability standards (e.g. data standardisation, data security protocols) could be relatively high, considering that this applies only to a limited set of basic data, it can still be kept within reasonable margins. However, depending the number of data points and the complexity, the costs impact might vary significantly. The complete lack of data standards will incur higher costs for compliance, while for the cases where some standards exist, cost will be reflected only by interoperability efforts needed for compliance. These types of costs will also apply to both data holders and re-users and will depend on their level of compatibility with the assumed standards. For example, using an open standard rather than a proprietary one might help limit the cost impacts and keep them on a lower side. Also, open standards stimulate both innovation and competition, while proprietary ones might add some negative impacts on.
- The costs of setting up APIs: low to medium-high impact. They can also vary, from low to very high, depending on the number of APIs, their scope and complexity. Additional to initial costs of setting up the APIs, maintenance costs (update and upgrade) should also be considered, as complex built APIs have higher annual maintenance costs. Depending on the type of infrastructure and technical characteristics, some recent studies have estimated that the costs of establishing APIs range between 30 000 euro and 2.5 million euro.[394] On average, the cost for setting up an API was estimated around 50.000 euro.[395]
- The costs for developing guidelines and codes of conduct: limited to low impact. These types of costs are often on the lower side of the scale, and their overall impact remain relatively low,
- The costs related to the accreditation: low to medium impact. This type of cost will mostly occur for re-users, but data holders can also be affected by it. While it depends on a sector's settings, these costs should not be excessively high. An initial estimate will place this type of cost on low level of impact.

The cost-benefit analysis shows significant differences of costs between the two sectors in scope. In the case of smart home appliances, there are also significant differences of costs within the sector, with the average costs for data holders much higher than the ones for data re-users. The reason is that the exemption rule applied to micro- and small enterprises (with less than 50 employed persons) reduces significantly the average costs for data re-users as 99% of the repair and aftermarket service providers fall in this category. Thus, the number of data re-users (repair and aftermarket service providers) affected is significantly lower than the one of data holders (device manufacturers)

[394] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

[395] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

affected. In the fitness trackers sector, the cost differences are much lower, both data holders and data re-users expecting similar levels of costs.

Table 72 Overview of costs for Measures supporting citizen empowerment ('human-centric data economy') | PO 2 (smart home appliances and fitness trackers)

Overview of costs (EUR) – PO 2 (smart home appliances)				
	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	940 million	168.3 million p.a.	3.9 million	0.87 million p.a.
Indirect costs	-	-	-	-
Overview of costs (EUR) – PO 2 (fitness trackers)				
	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	47.8 million	10.2 million p.a.	35.6 million	7.8 million p.a.
Indirect costs	-	-	-	-

In the case of this low-intensity regulatory option, the overall compliance costs are estimated to go from a low to a medium impact on the data holders and re-users. However, the impact will also be leveraged by the underlying conditions and size of the sector. For example, if we were to compare the effects for wearables and smart home appliances sectors, we will get different impacts. Considering the wearable sector, the costs impact will tend to remain on the low side, as the already existing conditions in the data market are considerably advanced. While some additional improvements might be needed (e.g. common taxonomy used for categories of data) for better interoperability between systems, the costs incurred by these adjustments will remain relatively low. On the other hand, for the smart home appliances sector, the wider diversity of products and the high degree of data fragmentation might result in much higher costs than initially estimated. Discussions with stakeholders from home appliances sector have revealed that the sector remains quite fragmented when it comes to technical interoperability. Different types of appliances have different level of digitalisation, and manufactures often might use different technologies and formats to collect and store data. For example, a vacuum robot might use onboard camera to take pictures of its environment, while others might use laser range finder (LIDAR) technology. While both devices will map their environment using also additional information from different sensors, the results are often not technologically compatible. As long as the systems are not technically interoperable, data portability between these types of devices remains just a theoretical concept at this stage. To increase the technical interoperability, a possible solution will be to set up a common open standard for the sector. However, in this case additional costs will incur to manufacturers to comply with the new standards, and these costs might vary significantly between market players.

3.3.2.4.2.2 Benefits of the option, including cost reductions and other positive effects on (some of) the stakeholders

As in the case of previous policy option proposed, several types of benefits can be identified, for both data holders and data (re)users.

Overall, the identified benefits concern:

- Increase the customers' choices that helps mitigate the "locked-in" effect by easily switching between devices and services at more competitive prices.
- Develop new or improved services and products for customers, increasing the benefits of the end-users.
- At the same time, for companies and service providers, developing new or additional service by integrating their collected data with data from other sectors will generate new opportunities and help establishing new markets.
- New providers can join the market and increase its competitiveness, enhancing the innovation level for both well-established and new market providers.

In particular, for smart home appliances, additional benefits can be included, such as:

- Increased efficiency of home appliances developed, in both energy consumption and functionalities. This will result in savings benefits to end-users due to lower energy consumption of products.
- Extended life for home appliances, by developing the predictive maintenance options. This contributes also to reduce the environmental impact, by reducing industrial waste.
- Opening up the repair aftermarkets to independent providers, and increasing the competition on the market. At the same time, it offers customers a wider choice of service providers for repairing the home appliances, at more competitive prices.
- Better and more diversified products and services, including the ones connected to smart homes that could contribute in lowering the environmental impacts of consumption.
- Emergence of new players and new markets that enhancing innovation developments for all market players.

Table 73 Overview of benefits – Measures supporting citizen empowerment ('human-centric data economy') | PO 2 (smart home appliances)

Description	Amount (EUR)	Stakeholders
Direct benefits		
Additional revenues due to more market opportunities	*62 million p.a.*	**Data holders**
	47.1 million p.a.	**Data re-users**
Cost savings for consumers	*1.89 million p.a.*	**Consumers**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

In this case, due to the compulsory nature of the measure, the values were slightly adjusted downwards. Using the same assumptions from the non-regulatory policy options, we estimate around

121.9 million euro for data holders and 133.6 million euro for date re-users in additional benefits, on annual basis.[396]

The consumers benefits will also increase, as the measure targets all companies and increases the opportunities for energy savings through more efficient products and services. In this case, percent of savings is no longer adjusted by the companies' rate of adoption. Therefore, the amount of expected benefits increases significantly, reaching 1.89 million euro.

Table 74 Overview of benefits – Measures supporting citizen empowerment ('human-centric data economy') | PO 2 (fitness trackers)

Description	Amount (EUR)	Stakeholders
Direct benefits		
Additional revenues due to more market opportunities	*121.9 million p.a.*	**Data holders**
	155.9 million p.a.	**Data re-users**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

In this case, considering the market maturity, and the regulatory nature of the measure, the additional revenues were estimated around 121.9 million euro for data holders and 155.9 million euro for data re-users.[397].

Considering the measures included in this low-intensity policy options, we can say that the overall impact of these benefits will remain rather limited by the data availability. Making available few data sets and adding specific conditions for data access will hamper achieving the full potential of data portability.

3.3.2.4.2.3 Findings of the cost-benefit analysis

This policy option impacts all market participants, as it is no longer limited to only those likely to gain more. Moreover, producers could benefit from additional revenue from both reciprocity and possibility for premium data and services offerings.[398] Overall, this option will translate in high benefits for fitness and wearables re-users, but limited for white appliances re-users due to its limited maturity and demand.

The costs are expected to be high, especially for white appliances that would have to set up from scratch advanced data management solutions for a very embryonic market. At the same time, these

[396] Only 1.5% additional benefits for data holders and 2% for data re-users.
[397] 30% additional benefits for data holders and 50% for data re-users.
[398] At this policy option concerns only to a limited set of data, the data holders have the possibility to widen their offers with either additional data or tailor-made services. Therefore, they can include these options as premiums offerings for data re-users.

costs will no longer be moderated by a voluntary adoption only by those with better cost-benefit ratio as in the previous policy option. Also, the benefits for both producers and data re-users will be high due to the reciprocity clause and the possibility to set up premiums for data offers. However, for data re-users, the possibilities for innovation from large scale machine learning remain limited.

For fitness trackers, the costs could end up on the medium-high side, as often producers are more digitally prepared, but the cost-benefit ratio is tempered by the compulsory nature of the policy option. Benefits remain high for both producers and data re-users, as the reciprocity clause remain in place.

3.3.2.4.3 Coherence of the option

The option could be similar to what has been adopted in car, finance and energy sector. However, there are concerns that planned implementation of the Energy Efficiency directive related to demand response could include provisions on data sharing from smart appliances, hence potentially conflict with this option.

3.3.2.5 Policy Option 3: high-intensity regulatory intervention

This section assesses the second policy option for Measures supporting citizen empowerment (‘human-centric data economy’).

3.3.2.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the policy objectives.

3.3.2.5.1.1 Achievement of specific objectives

The third policy option goes further in fostering the development of a market for data portability. It addresses all of the operational objectives, and provides the necessary conditions for achieving the specific and general objectives. The measures proposed will help improve the development of the overall system and address the fragmentation level within certain sectors, and establish the premises for a wider inter-sectoral data integration.

3.3.2.5.1.2 Achievement of general objectives

Extending the coverage of data sets for portability and including the real-time data provisions expand the applicability area of data portability, and help better achieving the specific and general objectives. It will enable the development of models and initiative for data portability based on standards, allowing new players to join the market. Complemented by lower costs and less access restrictions for re-users, it contributes to the establishment of new and complementary markets. Moreover, more participants on the market increases the competition and enhances innovation, generating more benefits for the end-users.

Overall, this high-intensity intervention impact is significantly higher compared to the previous policy option and help achieving the objectives set up initially.

3.3.2.5.2 Efficiency: Costs and benefits of the option

This subsection presents the costs and benefits associated with policy option 3.

On the efficiency side, the policy option has both positive and negative effects, with different intensities. On the negative side, we have the cost impact. The measures proposed could generate potentially high costs as they are applied to all categories of data collected by the data holders. The costs impact will range from accessible to extremely prohibitive, depending on the digital developments in place for both data holders and re-users. For example, a business with a lower digitalisation level will incur higher costs to comply with the policy option requirements compared to a more advanced one. Also, a large volume of data will generate additional costs, both for setting up the system and for data management maintenance. At the same time, since the reciprocity clause is not active anymore, the equivalent exchange of data between data holders and data re-users is no longer required. Thus, data re-users do not have to provide equivalent sets of data in order to take advantage of the data portability benefits. In this case, the costs burden will fall mainly on the data holders.

On the positive side, we have the increase in benefits. A wider selection of datasets available, with unrestricted access, offers numerous opportunities for development of new services and products. The diversity of data holders and re-users will stimulate competition and innovation, increasing the development of new and complementary services and products, and benefiting both the end-users and service and product providers.

3.3.2.5.2.1 Costs of the option

Similar with the low-intensity option, there are several types of cost that can be identified, for both data holders and data (re)users. And, their impact on both data holders and (re-)users varies, depending on their digital readiness and the overall conditions of the sector's data market. The most common types of cost encountered are:

- The costs for technical interoperability compliance: low to high impact. The volume of data for which the interoperability standards need to be set up increased compared to the previous option. This could in much higher costs for data holders, depending on the volume of data collected. If data standards need to be set up from an empty canvas or if they differ significantly than the ones required, the compliance's cost will get higher. Otherwise, the costs for compliance can be maintained at more reasonable levels. When the standards used correspond to those required, the costs will be minimal. These types of costs could apply to both data holders and re-users and their impact will depend on the level of compatibility with the assumed standards.
- The costs of setting up APIs: low to medium-high impact. Similar reasoning can be applied for setting up APIs. The development and implementation costs can also vary, from low to very high, depending on the number of APIs developed, their scope and their complexity. Additional to initial costs of setting up the APIs, maintenance costs (update and upgrade) should also be considered, as complex built APIs have higher annual maintenance costs. Therefore, depending on the type of infrastructure and technical characteristics, some recent studies have estimated that the costs of establishing APIs range between 30 000 euro and 2.5 million euro.[399] On average, the cost for setting up an API was estimated around 50.000 euro.[400]
- The costs for developing guidelines and codes of conduct: limited to low impact. These types of costs are often on the lower side of the scale, and their overall impact remain relatively low,
- The costs related to the real-time data provision: medium to high impact. These costs are also linked with the technological readiness and data volumes targeted; therefore, it can vary significantly between data holders.

The cost-benefit analysis shows significant differences of the average costs between the two sectors in scope. The estimates accounts for the level of digitalisation of the data holders in the two sectors (on one hand), and the number of affected companies (much higher number for smart home appliances than the one for fitness trackers). Moreover, as the reciprocity clause is no longer in force,

[399] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

[400] Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en

the additional costs for complying with the policy option requirements will mainly affect the data holders rather than the data re-users, as they do not have to share equivalent data to benefit from the data portability effect. In the case, the data re-users are not expected to incur any additional costs for compliance with the policy option.

Table 75 Overview of costs for Measures supporting citizen empowerment ('human-centric data economy') | PO 3 (smart home appliances and fitness trackers)

Overview of costs (EUR) – PO 3 (smart home appliances)

	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	2734.8 million	246.4 million p.a.	-	-
Indirect costs	-	-	-	-

Overview of costs (EUR) – PO 3 (fitness trackers)

	Data holders		**Data re-users**	
	One-off	*Recurrent*	*One-off*	*Recurrent*
Direct costs	166.7 million	15 million p.a.	-	-
Indirect costs	-	-	-	-

For the high-intensity regulatory option, the overall compliance costs are estimated to go from a low to a high impact on the data holders and re-users. However, the impact will also be leveraged by the underlying conditions and size of the sector. For example, comparing the average costs impact for wearables and smart home appliances sectors, we will get different perspective. The wearable sector is far more advanced when it comes to data sharing and portability than smart home appliances. In this context, the costs impact for wearables will remain mostly on the low side, while for the smart home appliances, these could vary from low to very high, depending of the type of products, digitalisation level and the degree of data fragmentation of the market.

3.3.2.5.2.2 Benefits of the option, including cost reductions and other positive effects on (some of) the stakeholders

As in the case of previous policy option proposed, similar types of benefits can be identified, for both data holders and data (re)users. In addition, the effect of these benefits is amplified by the extended measures proposed.

Overall, the identified benefits concern:

- Increase the customers' choices, helping mitigate the "locked-in" effect by easily switching between devices and services, for free or with minimum costs. More choices available will also reduce the dependency on providers.
- Develop new or improved services and products for customers, increasing the end-users' benefits.
- At the same time, companies and service providers will have increased and more variate opportunities for developing new or additional service by integrating their data with data from other sectors.
- New providers, for similar and complementary sectors, will be able to join the market and increase its competitiveness, enhancing the innovation level for all market participants.

- Establish new or complementary markets and generate further benefits for end-users and economy.
- Enhance the inter-sectoral data market, through better technical interoperability, with increased societal and economic benefits.

In particular, for smart home appliances, additional benefits can be included, such as:

- Increased efficiency of home appliances developed, in both energy consumption and functionalities. This will result in savings benefits to end-users due to lower energy consumption of products.
- Extended life for home appliances, by developing the predictive maintenance options. This contributes also to reduce the environmental impact, by reducing industrial waste.
- Opening up the repair aftermarkets to independent providers, and increasing the competition on the market. At the same time, it offers customers a wider choice of service providers for repairing the home appliances.
- Better and more diversified products and services, including the ones connected to smart homes that could contribute in lowering the environmental impacts of consumption.
- Emergence of new players and new markets that enhancing innovation developments for all market players.

Table 76 Overview of benefits – Measures supporting citizen empowerment ('human-centric data economy') | PO 3 (smart home appliances)

Description	Amount (EUR)	Stakeholders
Direct benefits		
Additional revenues due to more market opportunities	-	**Data holders**
	94.3 million p.a.	**Data re-users**
Cost savings for consumers	*1.89 million p.a.*	**Consumers**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

In this case, the data holders are expected to have no additional benefits, as the reciprocity clause no longer apply. On the other hand, data re-users additional benefits are estimated to 94.3 million euro, on annual basis.[401] The consumers benefits will be similar to the previous policy option, accounting for 1.89 million euro due to energy savings.

Table 77 Overview of benefits – Measures supporting citizen empowerment ('human-centric data economy') | PO 3 (fitness trackers)

Description	Amount (EUR)	Stakeholders
Direct benefits		
Additional revenues due to more market opportunities	*60.9 million p.a.*	**Data holders**
	140.4 million p.a.	**Data re-users**
Impact on policymaking and decision-making	*Not quantifiable due to lack of data*	**Data holders, data re-users, Consumers**
Potential new scientific insights with positive outcomes on research and innovation	*Not quantifiable due to lack of data*	**Data (re)users**
Indirect benefits		
Effect on Gross Domestic Product (GDP) Innovation and competitive advancement	*Not quantifiable due to lack of data*	**Data holders, data re-users**
New insights	*Not quantifiable due to lack of data*	**Data holders, data re-users**

In this case, the additional benefits for fitness (wearables) data holders are reduced to half of the ones assumed in the previous policy option, while for data re-users, the decline is much smaller. [402] In this case, the additional revenues were estimated around 60.9 million euro for data holders and 140.4 million euro for data re-users.

The high-intensity policy option has a much better impact overall, but with different result for the two sectors in scope, as reciprocity clause no longer apply which also affect the level of additional benefits for data holders. The most affected in this case are the data holders from the smart home appliances sector for which there were no additional benefits estimated. At the same time, for the data holders in the fitness trackers, the benefits diminish almost by half when compared to the ones estimated for previous policy option.

3.3.2.5.2.3 Findings of the cost-benefit analysis

This policy option brings additional impacts to all market participants, compared to the previous option. Technological interoperability steps up a level by adding the real-time provisions for data and extending to the maximum the list of data categories, while dropping at the same time the reciprocity clause provisions. This will have an important impact on producers, especially the ones of smart home appliances that would bear high costs and have almost no benefits.

Similar to the previous option, for fitness trackers, the costs for setting up and running technological interoperability could end up quite high, as the wider category of data concerned, and the real-time provisions add more costs for companies. Moreover, the benefits for producers will be lower because the reciprocity clause is no longer in place. Costs for the data re-users will be lower, while keeping relatively high benefits due to the possibility of data reuse, cross-selling and added value services.

[401] No additional benefits for data holders and 4% for data re-users.
[402] 15% additional benefits for data holders and 45% for data re-users.

Also, the possibilities for innovation from large scale machine learning improves, as there is wider range of data available and less restrictions to access it.

In the case of the smart home appliances, the costs for setting up and running technological interoperability standards are very due to limited standardisation and data portability currently in the sector. In addition, everyone will have to comply with the requirements for compliance, not only those with better cost-benefit ratio. The benefits will be much lower as the reciprocity clause no longer applies. For data re-users, the costs will be lower, while the benefits will be medium, as there is the possibility of data reuse but there is limited evidence of the value of these data. At the same time, the possibilities for innovation from large scale machine learning also remain rather limited, due to the lower level of maturity of the market.

Overall, this option will translate in higher benefits for re-users, especially in fitness and wearables, but also for white appliances data re-users too, notably for repair shops.

3.3.2.5.3 Coherence of the option

The option could be similar to what has been adopted in car, finance and energy sector. However, there are concerns that planned implementation of the Energy Efficiency directive related to demand response could include provisions on data sharing from smart appliances, hence potentially conflict with this option.

3.3.2.6 Summary of the impacts

The following table summarises the possible impacts of the policy options:

Table 78 – Summary of impacts for Measures supporting citizen empowerment ('human-centric data economy')

Type of impact	Type of impact
Economic impacts	Improved competition and data-driven innovation will support economic growth and create new high value jobs. The development of better and more convenient products and services, customised to individuals' needs, but also new ones fuelled by the behavioural information embedded in the data. Better access to data will also support more efficient processes for businesses, with less burdensome and improved regulatory compliance.
Social impacts	Potential savings for customers and reducing the costs of living Reducing the "lock-in" effect for customers and stimulating switching between producers and service providers Improved health from increased data availability and related health services from fitness trackers
Environmental impacts	Improved products and services that reduces wastes and energy consumptions (energy efficient smart home appliances)
Fundamental rights impacts	

3.3.3 Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

3.3.3.1 Stakeholders Affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Who?	How?
Data holders **(Large machinery manufacturing companies)**	Data holders will mostly face costs linked to implementation, administrative burdens and compliance with this policy intervention. These costs are further linked into elimination of competitive advantages in aftermarkets as well as to reduced innovation capacity in the primary market. In return, they might benefit from time and cost savings, increased business and growth opportunities due to increased trust among the players in the market as well as due to increased and enhanced access into third parties data.
Data co-producers **(Smart machinery / IoT Solution Users)**	Data co-producers could mainly benefit from time and cost savings, increased effectiveness, productivity, growth and innovation capacity associated to increased and enhanced access and use of co-generated data. Additionally, they might benefit from cost savings related to more efficient cost management and lower prices for aftermarket services due to elimination of monopolistic aftermarkets.
Data re-users (independent service providers)	Data re-users might be recipients of indirect benefits of this initiative. This is linked to a) the fact that enhanced clarity and fairness over IoT data access and use rights might increase the volume of B2B data sharing and therefore create business and innovation opportunities for the provision of new data-driven products and services; b) impacts on market competition, as the number of aftermarket players is expected to be significantly increased..
Society	Society might be affected by both positive and negative impacts related to a) market innovation (linked to development of new products and services); b) market competition (linked to employment levels and emerging market players); c) increased digitalization of some industry sectors due to increased trust in the market.

3.3.3.2 Policy Option 0: baseline scenario

3.3.3.2.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of a baseline scenario in achieving the policy objectives.

3.3.3.2.1.1 Achievement of specific objectives

The absence of measures clarifying and potentially further developing rights on co-generated data and B2B data sharing does not contribute to reaching any of the specific policy objectives.

Lack of clarity on determining and disseminating using rights on co-generated data in the economy will persist and B2B data sharing of co-generated data at fair conditions will not be enabled. Therefore, data sets of potentially high importance in terms of value creation will continue to be non-available and accessible by key players (co-producers or re-users).

3.3.3.2.1.2 Achievement of general objectives

The absence of measures clarifying and potentially further developing rights on co-generated data and B2B data sharing does not contribute to reaching any of the general policy objectives.

The identified market inefficiencies will not be addressed. The levels of innovation and development of resilient supply chains will remain limited and monopolistic aftermarkets will persist. Digitalisation-related developments of certain industry sectors will remain at low level due to lack of trust between the market players. Finally, the full potential of data for the EU economy and society will not be maximised.

3.3.3.2.2 Efficiency: Costs and benefits of the option

This subsection examines the costs and benefits associated with a baseline scenario.

3.3.3.2.2.1 Costs of the option

No specific costs are directly associated to the baseline scenario since no action will be taken to change the status quo. Cost assessment of baseline scenario is associated to the market inefficiencies identified by this problem assessment which will remain unaddressed (including limited innovation capacity of IoT solution users in the European market, limited development of resilient supply chains, competition issues in monopolistic aftermarkets and limited digitalization of certain industry sectors due to lack of clarity on access and usage rights on co-generated data in the data economy and limited B2B data sharing).

Additionally, lack of horizontal regulatory framework defining fair conditions for access and usage of co-generated data, might lead to costs related to:

- Lengthy negotiation processes to come to a contractual agreement between the parties;
- Costs related to legal risks (i.e. cost of being entangled to a litigation due to the lack of clarity on rights over co-generated IoT data and non-agreement between the parties), estimated to reach approximately 1M EUR/year by the interviewed stakeholders;
- Switching costs for the users of IoT solutions for having aftermarket services from third parties, estimated to be approximately 100K EUR/year by the interviewed stakeholders;
- Potential persistence of unfair commercial practices with regard to co-generate data access and co-use;
- Costs linked to continuation of wasteful/ inefficient data practices, where most data in Europe is unused;
- Costs linked to limited ability of companies to access the data needed to enter and compete in concentrated markets.

3.3.3.2.2.2 Benefits of the option, including cost reductions and other positive effects on (some of) the stakeholders

No specific benefits are directly associated to the baseline scenario since no action will be taken to change the status quo. However, the absence of horizontal measures regulating data sharing of industrial co-generated data might be linked to positive impacts from avoiding over-regulating and imposing unnecessary burdens and costs to specific industry sectors which are in the phase of growing and do not face particular problems related to rights over co-generated data. Interviewees representing industry sectors with lower productivity rates compared to others were in favour of maintaining the baseline scenario, allowing the possibility of sector specific regulation where respective need exists.

3.3.3.2.2.3 Findings of the cost-benefit analysis

Policy Option 0 entails no directly associated costs and benefits for the stakeholders since no action will be taken to change the status quo. The assessment of both costs and benefits under the baseline scenario is linked to the absence measures clarifying and potentially further developing rights on co-generated data and B2B data sharing.

3.3.3.2.3 Coherence of the option

The absence of action at EU level would not change the status quo, and therefore coherence with the existing EU policy and legal framework is ensured. However, incoherence issues might arise from the fact that the baseline scenario does not strengthen nor promote the following provisions of the European Strategy for Data: *a) data can flow within the EU and across sectors; b) the rules for access and use of data are fair, practical and clear*. Furthermore, identified problems linked to B2B data-sharing and usage rights on co-generated industrial data (IoT data created in industrial settings) will not be addressed.

3.3.3.3 Policy Option 1: Industry-driven self-regulatory framework for co-generated data

This section assesses the first policy option for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.

3.3.3.3.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 1 in achieving the specific and general policy objectives.

3.3.3.3.1.1 Achievement of specific policy objectives

Policy option 1, which entails the establishment of an industry-driven self-regulatory framework aiming to promote measures clarifying and potentially further developing rights on co-generated data and B2B data sharing could contribute in achieving the specific policy objectives.

In particular, through the discussions among the industry stakeholders in the frame of the expert group, an agreement could be reached on measures clarifying and disseminating access and using rights on co-generated data in the economy. Additionally, potential agreement on best practices could enable and promote B2B data sharing of co-generated data at fair conditions, making available and accessible to key players (co-producers or re-users) data sets of potentially high importance in terms of value creation. The standardisation exercise could contribute in creating new standards for data sharing or opening-up existing ones. This could further lead in the reduction of other technical barriers of B2B sharing linked to the lack of interoperability. However, this option can only be partially effective due to its non-binding nature.

3.3.3.3.1.2 Achievement of general policy objectives

Policy option 1 could only partially contribute in achieving the general policy objectives. An industry driven self-regulatory framework for co-generated data could increase trust among the market players and accelerate digitalization of some industry sectors which might lag behind due to trust issues over access and usage rights, as well as enable the development of more resilient supply chains due to better access and use of data.

It could also partially contribute in addressing other market failures, linked to limited competition and innovation in monopolistic aftermarkets. However, at some extent, those will most likely persist. The reason is that this kind of market failures are not directly or exclusively linked to the lack of clarity on rights over co-generated data. On the one side, they are associated to rights over internal machine operation data over which certain large smart machinery manufacturers have *de facto* control. Such an issue could not be easily addressed by a non-regulatory policy option. On the other side, these market failures are linked to technical aspects arising from the lack of interoperability and common standards in some industry sectors. As far as the development of common standards is concerned, such an industry-driven and tailored policy option could be significantly effective. Another factor however that could negatively affect the effectiveness assessment of this policy option is linked to its non-binding nature, which leaves questions and gaps in terms of implementation of best practices. Therefore, the potential of data for the EU economy and society would be maximised but would still not reach its full levels.

3.3.3.3.2 Efficiency: Costs and benefits of the option

This subsection examines the costs and benefits associated with policy option 1.

3.3.3.3.2.1 Costs of the option

Under the non-regulatory policy option, data holders (i.e. large companies that are IoT solution providers, smart machinery manufacturers) will face one-off and recurrent implementation costs linked to technical developments. These refer to:

- The development of data management agreements and relevant administrative/overhead cost. This could be linked to potential costs for development document management systems that allow a) the creation of such agreements; b) the interaction among the different parties involved, as well as c) tracking of the agreements. The estimated amount of such costs varies between big companies and SMEs. The interviewed stakeholders estimated the amount of this cost to reach approximately up to 1M EUR/year for the data holders.
- Cost of facilitating data exchanging and portability (i.e. cost of transferring data in the cloud, establishing and maintain the API). The estimated amount of this cost is approximately 30K EUR/year for the data holders.
- Additional potential costs related to modifying internal technical architectures and back-end procedures, which might arise for some stakeholders, depending on a) what methodologies companies already use and how much of that should be changed; b) the specific standards and terms that will be agreed among the industry stakeholders; c) specific data sets under "regulation"; d) the size of the company and volume of data handling. The estimated amount of this cost under this policy option is approximately up to 10-50M EUR/year for the data holders.

However, an important aspect to be mentioned is that, since this is a non-regulatory (and therefore non-binding) option, it remains up to the choice of each stakeholder, whether they comply and face the costs or not (optional costs).

Finally, the establishment of an industry-driven, self-regulatory framework for co-generated IoT data entails costs for both data holders (large companies that are IoT solution providers) and data co-producers (IoT solution user) related to:

- one-off resources and time (working days) spent for the participation to the relevant expert groups or stakeholder forums;
- recurrent legal risk costs associated to the absence of relevant legislation (i.e. cost of being entangled to a litigation, which is possible if there is no legal clarity with regard to terms of access and usage of co-generated IoT data and non-agreement between the parties). The estimated amount of this cost under this policy option ranges between 750K-1M EUR/year

3.3.3.3.2.2 Benefits of the option

The overall benefit of a policy option aiming to clarify rights on co-generated IoT data would be the increased legal certainty, further accompanied by transparency and trust among the various market players. . This could be then associated to several direct and indirect benefits (such as efficiency gains, cost savings, increased business opportunities etc.), depending on the type of stakeholder, machine-generated data set affected. Additionally, studies have shown that a big number of industry players consider that lack of clarity on access and usage rights on data is a major factor preventing them from sharing data and limiting B2B data sharing. In particular, the Federation of German Industries (BDI) has recently conducted a study on the biggest existing barriers to data use for companies. According to the results, 84,2% of 500 participating companies think that "legal uncertainty regarding data usage rights" is holding them back from data sharing, being considered as the third most important obstacle.[403] Similarly, according to another survey conducted in the frame of a study on data sharing between companies in Europe, legal uncertainty about rights over data constitutes the second most important obstacle to data sharing, within 54% of the respondents confirming this statement.[404] Therefore, a policy initiative contributing in providing clarity on rights over data, could significantly increase B2B data sharing and its associated benefits. The benefits assessment for each stakeholder category under this policy option is presented in detail below.

Benefits for data co-producers (IoT solution users)

- Cost savings from reduction of switching costs for aftermarket services (due to reduction of monopolistic aftermarkets and provision of services at lower prices by independent service providers)[405]. Under this policy option, interviewed stakeholders estimated a maximum of 10% of cost savings from reduction of switching costs for aftermarket services.
- Increased and enhanced usage of data (e.g. agronomic data or data produced on a construction site), could increase effectiveness, productivity and innovation for the development of new products and services. Recent statistics from the oil and gas industry sector showed that oil and gas companies can improve their production by 6% to 8% with proper utilisation of IoT data.[406] Under this policy option, interviewed stakeholders confirmed this and estimated 5%-8% increased effectiveness and productivity due to enhanced data access and use. These effectiveness gain can further lead to several other indirect benefits, such as:
  - Cost savings with regard to more efficient cost management (i.e. waste management)
  - Increased growth of business players in terms of revenue and workforce
  - Development of more resilient supply chains due to enhanced usage of data for the prediction of supply and demand issues
- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard legal risk costs associated to the absence of relevant legislation (i.e. cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed stakeholders estimated a maximum of 5% of cost savings related to legal risk cost reduction.

[403] BDI,Datenwirtschaft in Deutschland, Wo stehen die Unternehmen in der Datennutzungund was sind ihre größten Hemmnisse?, https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/

[404] European Commission, Study on data sharing between companies in Europe, 2018, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

[405] The reduction of monopolistic aftermarkets would result at the same time to economic losses of smart machinery manufacturers who currently hold dominant positions.

[406] Joshi N., (2019), Refining the Oil and Gas Industry with IoT, Forbes, https://www.forbes.com/sites/cognitiveworld/2019/09/17/refining-the-oil-and-gas-industry-with-iot/?sh=2e3ca56779f9

Benefits for data holders (large companies that are IoT solution providers, smart object manufacturers)

- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard legal risk costs associated to the absence of relevant legislation (i.e. cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed stakeholders estimated a maximum of 5% of cost savings related to legal risk cost reduction.
- Increased business and growth opportunities (in terms of client base, revenues, employees) due to:
    - Enhanced access to competitors' machines data could provide new business and innovation opportunities for the provision of new and enhanced aftermarket services (referring to the case where the IoT solution user would allow portability with other manufacturers than the initial machine manufacturer)
    - Increased trust in the market that would eliminate the margin of lost business opportunities due to hesitancy and non-agreement on best practices.

Benefits for potential data (re-)users (independent service providers)

- Increased business and opportunities for independent service providers (including better product design of spare parts and components, development of new data-driven services). Under this policy option, interviewed stakeholders estimated a potential 5%-10% increase of innovation and business opportunities.
- Increased aftermarket competition. The number of market players providing data-driven services is expected to be significantly increased, as an indirect benefit resulting from the increased B2B data sharing. Under this policy option, interviewed stakeholders estimated a potential 5%-10% increase of market competition.

In terms of indirect impacts, this initiative, which promotes more efficient use of data and fairer allocation data value among the market players can be also associated to environmental benefits, contributing, therefore, to European green deal objectives. In particular, more efficient use of data enables innovation that can improve energy efficiency and reduce greenhouse gas emissions. Additionally, it provides enhanced and increased reparability and optimisation opportunities, due to better data access in the context of predictive maintenance services, carried out by independent repairers, which should translate into a longer usage time for smart machines or devices.

It should be noted, that this policy option appears to be the preferred policy option of several interviewed stakeholders, due to the flexibility it offers, allowing each industry sector to be regulated according to its problems, needs and specificities as well as due to its non-binding nature. The various industry sectors present different levels of digitalisation and maturity in terms of B2B data sharing (for example some sectors have already developed standards between the OEMs, while others not), so each industry sector has different problems and needs in that regard. Several interviewees from different sectors further explained that this is a growing field, and stakeholders -regardless their size or position in the value chain- are currently examining ways of creating value from the various types of data sets. Therefore, a non-regulatory intervention would be the only option allowing flexibility (in terms of both its non-binding and industry-tailored nature) and not limiting the potential of value creation in the future.

3.3.3.3.2.3 Findings of the cost-benefit analysis

Under this policy option, the exact estimation of costs and benefits is challenging due to the non-binding nature of the intervention. In the case of high-level of compliance by industry stakeholders, benefits would be expected to outweigh the costs.

3.3.3.3.3 Coherence of the option

This policy option promotes the objectives of the single market for data under the European Strategy for Data and in particular the following provisions: a) data can flow within the EU and across sectors; b) the rules for access and use of data are fair, practical and clear. Furthermore, it contributes to address identified issues linked to B2B data-sharing and usage rights on co-generated industrial data (IoT data created in industrial settings). Therefore, the policy option remains coherent with the current EU legal and policy framework.

3.3.3.4 Policy Option 2: Adoption of a legal instrument aiming to bring legal certainty and promote contractual fairness for accessing and (co-)using IoT co-generated data

This section assesses the second policy option for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.

3.3.3.4.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 2 in achieving the specific and general policy objectives.

3.3.3.4.1.1 Achievement of specific policy objectives

Policy option 2, which entails the adoption of a legal instrument establishing certain access and usage rights and a fairness control mechanism would significantly contribute in achieving the specific policy objectives.

The establishment of certain access and usage rights as well as of a fairness control mechanism by the legal instrument as described in section 3.3.3.4 would provide legal certainty and clarity on access and usage rights over co-generated data in the economy. Furthermore, this policy measure would promote B2B data sharing of co-generated data at fair conditions, making available and accessible to key players (co-producers or re-users) data sets of high importance in terms of value creation. However, it should be mentioned that other technical barriers of B2B data sharing linked to lack of interoperability and common standards might persist. The exact effectiveness level depends on the content of the legislation in terms of the extent that it contributes in creating and/or opening-up common standards.

3.3.3.4.1.2 Achievement of general policy objectives

Policy option 2 could partially contribute in achieving the general policy objectives. The adoption of a legal instrument described in section 3.3.3.4 could enhance trust and accelerate digitalization of some industry sectors, which might lag behind due to trust issues over access and usage rights, as well as enable the development of more resilient supply chains due to better access and use of data.

It could also contribute in addressing other market failures, linked to limited competition and innovation in monopolistic aftermarkets. However, at some extent, these might persist depending on the exact formulation and provisions of the legal instrument. The reason is that competition issues in the aftermarkets are not directly or exclusively linked to the lack of clarity on rights over co-generated data. As mentioned above, on the one hand this kind of market failures are associated either to rights over "internal machine operation data" over which certain large smart machinery manufacturers have *de facto* control. On the other hand, they are linked to technical problems arising from the lack of interoperability and common standards. Therefore, Policy Option 2 could meet all general policy objectives, only in the case where the formulation of the "standard contractual clauses" includes specific provisions for addressing these issues. In this case, the potential of data for the EU economy and society would be maximised.

3.3.3.4.2 Efficiency: Costs and benefits of the option

This subsection examines the costs and benefits associated with policy option 2.

3.3.3.4.2.1 Costs of the option

Similar to the previous policy option, under PO2 - the low intensity regulatory intervention, data holders (i.e. large companies, which are IoT solution providers, smart machinery manufacturers) will face one-off and recurrent implementation costs linked to technical developments. These are expected to be at higher levels than PO1, due to the nature and intensity of a regulatory intervention, vis-à-vis a non-regulatory one, and they mainly refer to:

- The development of data management agreements, in compliance with the legislation and relevant administrative/overhead cost. This could be linked to potential costs for development of document management systems that allow a) the creation of such agreements; b) the interaction among the different parties involved, as well as c) tracking of the agreements. The interviewed stakeholders estimated the amount of this cost to reach approximately 1M EUR/year for the data holders.[407]
- Cost of facilitating data exchanging and portability (i.e. cost of transferring data in the cloud, establishing and maintain the API). The estimated amount of this cost, under this policy option, is approximately 50K EUR/year for the data holders.

Additionally, the main difference with the previous policy option is that a binding regulatory intervention imposes these costs for all the market players. This entails also costs related to increased need for legal workforce to follow up and ensure compliance with the legislation. Additionally, this might further affect:

a) the innovation capacity (investment capacity in the development of new products and services) on the data holders' side, which is expected to be reduced due to those high amounts spent in implementation costs of the regulatory measure;

b) competition in the primary markets which is also expected to be reduced, as certain marker players will be able to deal with the implementation costs, while others might not be able to afford the costs in combination with technological development and digitalization and drop out of the market.

Furthermore, mandatory access rights potentially created by the legislation are also likely to reduce incentives to collect data, as well as to upset the current business models of smart machinery manufacturers in certain industry sectors. The interviewees pointed out that given the specificities and the different levels of digitalisation and maturity in terms of B2B data sharing (for example some sectors have already developed standards between the OEMs, while others not), each industry sector has different needs in that regard. A horizontal regulatory policy measure could therefore impose unnecessary administrative and compliance burdens to some industry sectors and limit productivity. Several interviewees from different sectors further explained that this is a growing field, and

[407] These cost assessment figures are based on the current state of infrastructure and data management experience at the time of reporting (i.e. 2020), whereas the Data Act would likely not be applicable before 2024. Those 4 years of additional experience and evolving technologies might have an impact in terms of diminishing these costs. Additionally, the adoption of a policy measure might create legal and technical safeguards that would automatize and facilitate the implementation and monitoring of the data management agreements, leading to cost savings.

stakeholders -regardless their size or position in the value chain- are currently examining ways of creating value from the various types of data sets. Therefore, any kind of regulatory intervention could significantly block the potential of value creation in the future.

Finally, the interviewees made reference to three other "cost estimation scenarios", which might be relevant or not, depending on the exact formulation of the regulatory policy options:

a) potential costs related to modifying internal technical architectures and back-end procedures, which might arise for some stakeholders from data portability related provisions, depending on a) what methodologies companies already use and how much of that should be changed; b) the specific provisions and data sets under regulation, laid down by the policy measure; c) the size of the company and volume of data handling. The estimated amount of this cost under this policy option is approximately 10-50M EUR/year for the data holders.

b) in the case where the regulatory measure entails provisions for portability of specific data sets which are currently non-reproducible in a lab, then the need of building additional labs arises, the cost of which can be estimated at 100M EUR for a big OEM;

c) in the case where specific data sets of the "user" should not be visible at all by the manufacturer, then a need arises of having a separate external company in order to handle and store the data. Taking into account all the operational costs of such a solution, the amount of this cost is estimated at 10-100M EUR/year depending on the exact size of the manufacturing company and amount of data handled.

d) Provisions aiming to significantly change and enable B2B data sharing in order to boost aftermarket competition, come along with the risk to allow new industries (e.g. big-tech companies) to enter into the existing industries, limiting the role of European OEMs into only providing machines and vehicles to a larger system that they have no control of.

Table 79 - Overview of costs for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing | PO 2

Overview of costs (EUR) – PO 2

		Data holders		**Data co-producers**		**Data re-users**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures clarifying and potentially further developing rights on co-generated data and B2B data sharing	*Direct costs*	371 M EUR	6 500 M EUR p.a.	-	-	-	-
	Indirect costs	-	-	-	-	-	-

3.3.3.4.2.2 Benefits of the option

Under the low intensity regulatory intervention, similar types of benefits are observed, as in the PO1 – i.e. non-regulatory policy intervention. However, it is estimated that the ranges of some particular benefits (i.e. efficiency gains, time and cost savings linked to contract negotiations and legal risk costs, etc) might be differentiated, due to the different intensity and effectiveness levels between a binding and non-binding policy measure.

As described above, the overall benefit of the intervention aiming to clarify rights on co-generated IoT data would be the increased legal certainty, further accompanied by transparency and trust among the various market players. This could be then associated to several direct and indirect benefits (such as efficiency gains, cost savings, increased business opportunities etc.), depending on the type of stakeholder and machine-generated data set affected. Additionally, studies have shown that a big number of industry players consider that lack of clarity on access and usage rights on data is a major factor preventing them from sharing data and limiting B2B data sharing. In particular, the Federation of German Industries (BDI) has recently conducted a study on the biggest existing barriers to data use for companies. According to the results, 84,2% of 500 participating companies think that "legal uncertainty regarding data usage rights" is holding them back from data sharing, being considered as the third most important obstacle.[408] Similarly, according to another survey conducted in the frame of a study on data sharing between companies in Europe, legal uncertainty about rights over data constitutes the second most important obstacle to data sharing, within 54% of the respondents confirming this statement.[409] Therefore, a policy initiative contributing in providing clarity on rights over data, could significantly increase B2B data sharing and its associated benefits. The benefits assessment for each stakeholder category under this policy option is presented in detail below.

<u>Benefits for data co-producers (IoT solution users)</u>

- Cost savings from reduction of switching costs for aftermarket services (due to reduction of monopolistic aftermarkets and provision of services at lower prices by independent service providers)[410]. Under this policy option, interviewed stakeholders estimated 15% cost savings from reduction of switching costs for aftermarket services.
- Increased and enhanced usage of data (e.g. agronomic data or data produced in the construction site), could increase effectiveness, productivity and innovation for the development of new products and services. Under this policy option, interviewed stakeholders estimated 15% increased effectiveness and productivity due to enhanced data access and use. This effectiveness gain can further lead to several other indirect benefits, such as:
  - Cost savings with regard to more efficient cost management (i.e. waste management);
  - Increased growth of business players in terms of revenue and workforce;
  - Development of more resilient supply chains due to enhanced usage of data for the prediction of supply and demand issues.
- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard legal risk costs associated to the absence of relevant legislation (i.e. cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed

[408] BDI,Datenwirtschaft in Deutschland, Wo stehen die Unternehmen in der Datennutzungund was sind ihre größten Hemmnisse?, https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/
[409] European Commission, Study on data sharing between companies in Europe, 2018, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en
[410] The reduction of monopolistic aftermarkets would result at the same time to economic losses of smart machinery manufacturers who currently hold dominant positions.

stakeholders estimated up to 10% cost savings related to legal risk cost reduction. These savings are also linked to the legal and technical safeguards, expected to be introduced by the regulatory measure that will automatise and facilitate the implementation and monitoring of data management agreements.

Benefits for data holders (large companies that are IoT solution providers – smart object manufacturers)

- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard to legal risk costs associated to the absence of relevant legislation (i.e. cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed stakeholders estimated up to 10% cost savings related to legal risk cost reduction. These savings are also linked to the legal and technical safeguards, expected to be introduced by the regulatory measure that will automatise and facilitate the implementation and monitoring of data management agreements.
- 1% increased business and growth opportunities (in terms of client base, revenues, employees) due to:
    - Enhanced access to competitors' machines data could provide new business and innovation opportunities for the provision of new and enhanced aftermarket services (referring to the case where the IoT solution user would allow portability with other manufacturers than the initial machine manufacturer);
    - Increased trust in the market that would eliminate the margin of lost business opportunities due to hesitancy and non-agreement on best practices.

Benefits for potential data (re-)users (independent service providers)

- Increased business and opportunities for independent service providers (including better product design of spare parts and components, development of new data-driven services). Under this policy option, interviewed stakeholders estimated 500% increase in agricultural sector and 10% increase in other industry sectors, which are more advanced in B2B data sharing.
- Increased aftermarket competition. The number of market players providing data-driven services is expected to be significantly increased, as an indirect benefit resulting from the increased B2B data sharing. Under this policy option, interviewed stakeholders estimated 500% increase in agricultural sector and 10% increase in other industry sectors, which are more advanced in B2B data sharing.

In terms of indirect impacts, this initiative, which promotes more efficient use of data and fairer allocation data value among the market players can be also associated to environmental benefits, contributing, therefore, to European green deal objectives. In particular, more efficient use of data enables innovation that can improve energy efficiency and reduce greenhouse gas emissions. Additionally, it provides enhanced and increased reparability and optimisation opportunities, due to better data access in the context of predictive maintenance services, carried out by independent repairers, which should translate into a longer usage time for smart machines or devices.

Table 80 Overview of benefits – Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing | PO 2

Description	Amount (EUR)	Stakeholders
Direct benefits		
Increased effectiveness and productivity	196 727 M EUR p.a.	Data co-producers
Cost savings from reduction of switching costs for aftermarket services	68 130 M EUR p.a.	Data co-producers
Cost savings related to legal risk cost reduction	Non-quantifiable due to the lack of data	Data co-producers
Increased business opportunities	176 M EUR p.a.	Data holders
Cost savings related to legal risk cost reduction	Non-quantifiable due to the lack of data	Data holders
Indirect benefits		
Increased business opportunities	Non-quantifiable due to the lack of data	Data re-users
Increased market competition	Non-quantifiable due to the lack of data	Data re-users
Increased Innovation (development of new or better product and services)	Non-quantifiable due to the lack of data	Data re-users

3.3.3.4.2.3 Findings of the cost-benefit analysis

Taking into consideration that the benefits apply to a bigger and broader range of stakeholders compared to the costs, the benefits under this policy option would be expected to outweigh the costs, presenting the best balance between costs and benefits, compared to the other options.

3.3.3.4.3 Coherence of the option

This policy option promotes the objectives of the single market for data under the European Strategy for Data and in particular the following provisions: *a) data can flow within the EU and across sectors; b) the rules for access and use of data are fair, practical and clear.* Furthermore, it contributes to address identified issues linked to B2B data-sharing and usage rights on co-generated industrial data (IoT data created in industrial settings). Therefore, the policy option remains coherent with the current EU legal and policy framework. However, incoherence issues might arise, linked to intellectual property rights and trade secret protection legislation which currently protects *"internal machine operations data",* in the case where rights over this type of data are being affected by the new regulatory measure.

3.3.3.5 Policy Option 3: Legal Instrument clarifying access and using rights over co-generated data

This section assesses the third policy option for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.

3.3.3.5.1 Effectiveness in achieving the policy objectives

This subsection examines the effectiveness of policy option 3 in achieving the specific and general policy objectives.

3.3.3.5.1.1 Achievement of specific policy objectives

Policy option 3, which entails the adoption of a legal instrument clarifying access and usage rights over co-generated data could significantly contribute in the achievement of specific policy objectives. In particular, this legal instrument would provide clarity on access and usage rights on co-generated data in the economy and promote B2B data sharing of co-generated data at fair conditions, making available and accessible to key players (co-producers or re-users) data sets of high importance in terms of value creation. However, it should be mentioned that other technical barriers of B2B data sharing linked to lack of interoperability and common standards might persist. The exact effectiveness level depends on the content of the legislation in terms of the extent that it contributes in creating and/or opening-up common standards.

3.3.3.5.1.2 Achievement of general policy objectives

Policy option 3 could significantly contribute in achieving the general policy objectives. The identified market inefficiencies could be addressed by this intervention to the degree that these are connected to the lack of clarity on access and usage rights over co-generated data. Increased clarity and fair conditions for B2B data sharing are expected to enable innovation in the European market, the development of resilient supply chains, as well as to accelerate digitalisation of sectors, which might lag behind due to trust issues regarding access and usage rights. This intervention could also provide increased choices for IoT solution users with regard to aftermarket services by independent service providers and enable fair competition by eliminating competitive advantages and dominant positions of certain large IoT smart object manufacturers. As such, the policy option could also significantly help maximising the potential of data for the EU economy and society, in line with fundamental values.

3.3.3.5.2 Efficiency: Costs and benefits of the option

This subsection examines the costs and benefits associated with policy option 3.

3.3.3.5.2.1 Costs of the option

Similar to the previous policy options, under PO3-the high intensity regulatory intervention, data holders (i.e. large companies, which are IoT solution providers, smart machinery manufacturers) will face one-off and recurrent implementation costs linked to technical developments. These are expected to be at higher levels than PO1 and PO2 due to expected increased complexity in functionality and in operations, increased need for legal compliance, and more rigid roll out of systems to support the gathering and maintenance of contracts. In particular, the types and amount of cost estimation under this policy option refer to:

- The development of data management agreements, in compliance with the legislation and relevant administrative/overhead cost. This could be linked to potential costs for development of document management systems that allow a) the creation of such agreements; b) the interaction among the different parties involved, as well as c) tracking of the agreements. The interviewed stakeholders estimated the amount of this cost to reach approximately 2M EUR/year for the data holders.[411]

[411] These cost assessment figures are based on the current state of infrastructure and data management experience at the time of reporting (i.e. 2020), whereas the Data Act would likely not be applicable before 2024. Those 4 years of additional experience and evolving technologies might have an impact in terms of diminishing these costs. Additionally, the adoption of a policy measure might create legal and technical safeguards that would

- Cost of facilitating data exchanging and portability (i.e. cost of transferring data in the cloud, establishing and maintain the API). The estimated amount of this cost, under this policy option, is approximately 70-100K EUR/year for the data holders.
- Additional potential costs related to modifying internal technical architectures and back-end procedures, which might arise for some stakeholders from data portability related provisions, depending on a) what methodologies companies already use and how much of that should be changed; b) the specific provisions and data sets under regulation, laid down by the policy measure; c) the size of the company and volume of data handling. The estimated amount of this cost under this policy option is approximately 10-50M EUR/year for the data holders.

Similar to PO2, a binding regulatory intervention entails also costs related to increased need for legal workforce to follow up and ensure compliance with the legislation. Additionally, it might further affect

a) the innovation capacity (investment capacity in the development of new products and services) on the data holders' side, which is expected to be reduced due to those high amounts spent in implementation costs of the regulatory measure;

b) competition in the primary markets which is also expected to be reduced, as certain marker players will be able to deal with the implementation costs, while others might not be able to afford the costs in combination with technological development and digitalization and drop out of the market.

Furthermore, mandatory access rights potentially created by the legislation are also likely to reduce incentives to collect data, as well as to upset the current business models of smart machinery manufacturers in certain industry sectors. The interviewees also pointed out that given the specificities and the different levels of digitalisation and maturity in terms of B2B data sharing (for example some sectors have already developed standards between the OEMs, while others not), each industry sector has different needs in that regard. A horizontal regulatory policy measure could therefore impose unnecessary administrative and compliance burdens to some industry sectors, and limit productivity. Several interviewees from different sectors further explained that this is a growing field, and stakeholders -regardless their size or position in the value chain- are currently examining ways of creating value from the various types of data sets. Therefore, any kind of regulatory intervention could significantly block the potential of value creation in the future. Finally, the interviewees made reference to three other "cost estimation scenarios", which might be relevant or not, depending on the exact formulation of the regulatory policy options:

a) Potential cost of a dispute resolution mechanism. The average mediation cost for the parties involved is estimated to be approximately 20K-30K EUR, while the average arbitration cost approximately 50K EUR per case.

b) in the case where the regulatory measure entails provisions for portability of specific data sets which are currently non-reproducible in a lab, then the need of building additional labs arises, the cost of which can be estimated at 100M EUR for a big OEM;

c) in the case where specific data sets of the "user" should not be visible at all by the manufacturer, then a need arises of having a separate external company in order to handle and store the data. Taking into account all the operational costs of such a solution, the

automatize and facilitate the implementation and monitoring of the data management agreements, leading to cost savings.

amount of this cost is estimated at 10-100M EUR/year depending on the exact size of the manufacturing company and amount of data handled.

d) Provisions aiming to significantly change and enable B2B data sharing in order to boost aftermarket competition, come along with the risk to allow new industries (e.g. big-tech companies) to enter into the existing industries, limiting the role of European OEMs into only providing machines and vehicles to a larger system that they have no control of.

Table 81 - Overview of costs for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing | PO 3

Overview of costs (EUR) – PO 3

		Data holders		**Data co-producers**		**Data re-users**	
		One-off	*Recurrent*	*One-off*	*Recurrent*	*One-off*	*Recurrent*
Measures clarifying and potentially further developing rights on co-generated data and B2B data sharing	*Direct costs*	520 M EUR	12 813 M EUR p.a.	-	-	-	-
	Indirect costs	-	-	-	-	-	-

3.3.3.5.2.2 Benefits of the option

Under the high intensity regulatory intervention, similar types of benefits are observed, as in the first two policy options. However, it is estimated that the ranges of some particular benefits (i.e. efficiency gains, time and cost savings linked to contract negotiations and legal risk costs, etc) might be differentiated, due to the different intensity and effectiveness levels between non-regulatory, low-intensity and high-intensity policy measures.

As described above, the overall benefit of the intervention aiming to clarify rights on co-generated IoT data would be the increased legal certainty, further accompanied by transparency and trust among the various market players. This could be then associated to several direct and indirect benefits (such as efficiency gains, cost savings, increased business opportunities etc.), depending on the type of stakeholder and machine-generated data set affected. Additionally, studies have shown that a big number of industry players consider that lack of clarity on access and usage rights on data is a major factor preventing them from sharing data and limiting B2B data sharing. In particular, the Federation of German Industries (BDI) has recently conducted a study on the biggest existing barriers to data use for companies. According to the results, 84,2% of 500 participating companies think that "legal uncertainty regarding data usage rights" is holding them back from data sharing, being considered as the third most important obstacle.[412] Similarly, according to another survey conducted in the frame of a study on data sharing between companies in Europe, legal uncertainty about rights over data constitutes the second most important obstacle to data sharing, within 54% of the respondents confirming this statement.[413] Therefore, a policy initiative contributing in providing clarity on rights over data, could significantly increase B2B data sharing and its associated

[412] BDI,Datenwirtschaft in Deutschland, Wo stehen die Unternehmen in der Datennutzungund was sind ihre größten Hemmnisse?, https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/

[413] European Commission, Study on data sharing between companies in Europe, 2018, https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

benefits. The benefits assessment for each stakeholder category under this policy option is presented in detail below.

Benefits for data co-producers (IoT solution users)

- Cost savings from reduction of switching costs for aftermarket services (due to reduction of monopolistic aftermarkets and provision of services at lower prices by independent service providers)[414]. Under this policy option, interviewed stakeholders estimated up to 20% cost savings from reduction of switching costs for aftermarket services. These savings are also linked to the legal and technical safeguards, expected to be introduced by the regulatory measure that will automatise and facilitate the implementation and monitoring of data management agreements.
- Increased and enhanced usage of data (e.g. agronomic data or data produced in the construction site), could increase effectiveness, productivity and innovation for the development of new products and services. Under this policy option, interviewed stakeholders estimated 10% increased effectiveness and productivity due to enhanced data access and use. These effectiveness gain can further lead to several other indirect benefits, such as:
  - Cost savings with regard to more efficient cost management (i.e. waste management)
  - Increased growth of business players in terms of revenue and workforce
  - Development of more resilient supply chains due to enhanced usage of data for the prediction of supply and demand issues
- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard legal risk costs associated to the absence of relevant legislation (i.e. cost of dispute resolution settlement, or cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed stakeholders estimated up to 25% cost savings related to legal risk cost reduction

Benefits for data holders (IoT solution providers – smart object manufacturers)

- Time and costs savings with regard to reduction of resources and working time needed for contract negotiations as well as with regard legal risk costs associated to the absence of relevant legislation (i.e. cost of dispute resolution settlement, or cost of being entangled to a litigation due to the lack of clarity on co-generated IoT data and non-agreement between the parties). Under this policy option, interviewed stakeholders estimated up to 25% cost savings related to legal risk cost reduction. These savings are also linked to the legal and technical safeguards, expected to be introduced by the regulatory measure that will automatise and facilitate the implementation and monitoring of data management agreements.
- 1% increased business and growth opportunities (in terms of client base, revenues, employees) due to:
  - Enhanced access to competitors' machines data could provide new business and innovation opportunities for the provision of new and enhanced aftermarket services (referring to the case where the IoT solution user would allow portability with other manufacturers than the initial machine manufacturer)
  - Increased trust in the market that would eliminate the margin of lost business opportunities due to hesitancy and non-agreement on best practices.

[414] The reduction of monopolistic aftermarkets would result at the same time to economic losses of smart machinery manufacturers who currently hold dominant positions.

Benefits for potential data (re-)users (independent service providers)

- Increased business and opportunities for independent service providers (including better product design of spare parts and components, development of new data-driven services). Under this policy option, interviewed stakeholders estimated 500% increase in agricultural sector and 10% increase in other industry sectors, which are more advanced in B2B data sharing.
- Increased aftermarket competition. The number of market players providing data-driven services is expected to be significantly increased, as an indirect benefit resulting from the increased B2B data sharing. Under this policy option, interviewed stakeholders estimated 500% increase in agricultural sector and 10% increase in other industry sectors, which are more advanced in B2B data sharing.

In terms of indirect impacts, this initiative, which promotes more efficient use of data and fairer allocation data value among the market players can be also associated to environmental benefits, contributing, therefore, to European green deal objectives. In particular, more efficient use of data enables innovation that can improve energy efficiency and reduce greenhouse gas emissions. Additionally, it provides enhanced and increased reparability and optimisation opportunities, due to better data access in the context of predictive maintenance services, carried out by independent repairers, which should translate into a longer usage time for smart machines or devices.

Table 82 Overview of benefits – Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing | PO 3

Description	Amount (EUR)	Stakeholders
Direct benefits		
Increased effectiveness and productivity	131 151 M EUR p.a.	Data co-producers
Cost savings from reduction of switching costs for aftermarket services	90 840 M EUR p.a.	Data co-producers
Cost savings related to legal risk cost reduction	Non-quantifiable due to the lack of data	Data co-producers
Increased business opportunities	176 M EUR p.a.	Data holders
Cost savings related to legal risk cost reduction	Non-quantifiable due to the lack of data	Data holders
Indirect benefits		
Increased business opportunities	Non-quantifiable due to the lack of data	Data re-users
Increased market competition	Non-quantifiable due to the lack of data	Data re-users
Increased Innovation (development of new or better product and services)	Non-quantifiable due to the lack of data	Data re-users

Finding of the Cost-Benefit Analysis for Policy Option 3

The high intensity of legislation under this policy option is expected to significantly increase the implementation costs, burdens and decrease some of the benefits due to limited flexibility and innovation. Policy Option 3, therefore, presents a less ideal balance between costs and benefits, even though benefits are still likely to outweigh the costs.

3.3.3.5.3 Coherence of the option

This policy option promotes the objectives of the single market for data under the European Strategy for Data and in particular the following provisions: *a) data can flow within the EU and across sectors; b) the rules for access and use of data are fair, practical and clear.* Furthermore, it contributes to address identified issues linked to B2B data-sharing and usage rights on co-generated industrial data (IoT data created in industrial settings). Therefore, the policy option remains coherent with the current EU legal and policy framework. For the avoidance of doubt, coherence is only assured if clarifications are included on the relationship between any new legislation under this policy option and existing legislation, notably in relation to the Database Directive and the Trade Secrets Directive. These clarifications should ensure that the allocation of any rights or exceptions granted by the new legislation acknowledge and reference their relationship with the general framework created by these existing laws.

3.3.3.5.4 Summary of Impacts

The following table summarises the possible impacts of the policy options:

Type of Impact	**Description of Impact**
Economic Impact	Various economic impacts will follow this policy intervention, depending on the policy option and the type of machine-generated data regulated. Data holders (large companies, which are IoT solution providers) will mostly face costs linked to implementation, administrative burdens and compliance with this policy intervention. These costs are further linked into elimination of competitive advantages in aftermarkets as well as to reduced innovation capacity in the primary market. In return, they might benefit from time and cost savings, increased business and growth opportunities due to increased trust among the players in the market as well as due to increased and enhanced access to third parties' data. Data co-producers (IoT solution users) could mainly benefit from time and cost savings, increased effectiveness, productivity, growth and innovation capacity associated to increased and enhanced access and use of co-generated data. Additionally, they might benefit from cost savings related to more efficient cost management and lower prices for aftermarket services due to elimination of monopolistic aftermarkets. Data re-users and B2B data intermediaries might also be recipients of indirect benefits of this initiative. This is linked to a) the fact that enhanced clarity and fairness over IoT data access and use rights might increase the volume of B2B data sharing and therefore create business and innovation opportunities for the provision of new data-driven products and services; b) impacts on market

	competition, as the number of aftermarket players is expected to be significantly increased
Social Impact	Society might be affected by both positive and negative impacts related to a) market innovation (linked to development of new products and services); b) market competition (linked to employment levels and emerging market players); c) increased digitalization of some industry sectors due to increased trust in the market.
Environmental Impact	Environment might be affected in both ways. Better use of existing industrial data might reduce environmental impacts linked to new data collection and processing. Additionally, more efficient use of data enables innovation that can improve energy efficiency and reduce greenhouse gas emissions. It also provides enhanced and increased reparability and optimisation opportunities, due to better data access in the context of predictive maintenance services, carried out by independent repairers, which should translate into a longer usage time for smart machines or devices.. On the other side, increased B2B data sharing and use might lead to increased energy consumption of data processing facilities and technologies.
Impacts on fundamental rights	There might be an impact on the fundamental right to property, especially for smart object manufacturers, in the sense that they may have (rightly or wrongly) have come to see the data collected or created through their devices as somehow constituting part of their 'property' – i.e. they may perceive that they should have some form of exclusive rights on that data, merely as a practical result of their ability to establish data creation, capture and transfer parameters of the devices that they manufacture. This should not be a blocking point for future policy intervention however, since any such 'ownership' perceptions lack a clear basis in law. Moreover, existing instruments governing rights to data (including the database directive and trade secrets directive) apply a more balanced consideration of interests than merely granting unfettered exclusivity rights .

3.3.4 Measures supporting companies in cases of conflict of laws at international level

3.3.4.1.1 Stakeholders Affected

The following table provides an overview of the key stakeholders affected by the possible policy options and how:

Who?	How?
Data holders	Customers (including citizens, business and public administrations) would be more adequately protected against unlawful third country data access requests. This comes at a cost since additional protection measures will need to be passed on to the customers in some form.

(ICT service providers and their customers)	For ICT service providers: see Intermediaries below.
Data co-producers **(Complementary service providers – data intelligence and analytics)**	Data co-producers could benefit from the measures, depending on their role in the value chain. Co-producers whose role is to provide measures to protect data against unlawful access requests would benefit significantly, since their services are implicitly promoted. Co-producers who are themselves cloud providers however would face the same costs as ICT service providers in general: see Intermediaries below.
Data re-users (law enforcement, national security)	Data re-users in this context are public sector bodies from third countries seeking data access to European data. They could face additional costs and burdens in the short term, since the core objective of the policy options is to ensure that data access is only possible when equivalent safeguards are applied to those foreseen under EU law. However, assuming that such re-users are willing to align to EU requirements, in the medium and longer term their efforts and costs may actually decrease, since the common understanding of EU requirements and the applicability of a homogeneous framework to personal and non-personal data will allow procedures to be streamlined, reducing fragmentation. As a result, practical efforts and administrative costs (including in relation to legal proceedings) should decrease.
Intermediaries (ICT service providers, often but not exclusively cloud based)	ICT providers would face additional costs, since they are required at a minimum to invest in transparency measures (including notification duties under some of the policy options), and to invest in operational, technical and legal measures that decrease the risk of unlawful data access requests. These investments are not exclusively linked to the policy options though, since most providers would at any rate need to (continue to) invest in effective security measures to keep pace with the advancing state of the art. Moreover, the measures would also support other current and future policy initiatives, e.g. by establishing conditions under which in general governments are able to access data, including under business to government data sharing initiatives. None the less, certainly in the short term, burdens and expenses are likely to be imposed on targeted ICT providers. They could be partially offset by a potential reduction in costs related to disputes (when the lawfulness of such access requests is disputed), but the offsetting effect of relatively rare proceedings is unlikely to compensate the costs of investment.
Society	Society should be affected positively, since data sovereignty of citizens, businesses and administrations increases. Data will be more effectively and homogeneously protected, irrespective of its qualification as personal and non-personal data. While this comes at a cost, this cost is principally borne by the customers of ICT providers (see Data holders above), and it does not distribute evenly across all of society.

3.3.4.1.2 Efficiency: Costs and benefits of the options

This section presents the methodological approach in relation to the assessment of impacts of the policy options, including anticipated costs and benefits (cost and benefit analysis – CBA).

CBAs typically face methodological challenges, especially regarding quantitative data collection. That applies to an even greater extent to the issue of cross border conflicts of law, not just because of the complexity of calculating the costs of some of the measures, but also and more importantly the uncertainty in relation to benefits.

Benefits to the reduction in international conflicts of law are largely ephemeral: a significant part of the benefit is increased data sovereignty over one's own data and a reduction in confidentiality breaches, both of which are benefits that are significant, but also complex or even impossible to quantify. The analogy to data protection law could be made: while there are quantifiable benefits to harmonising data protection rules (as a result of improved internal market practices), the principal objective of safeguarding a fundamental right cannot and should not be economically valued: a fundamental right inherently deserves protection. This reasoning can be extended to data sovereignty outside the context of personal data as well: if one accepts that the societal importance and value of some data is significant enough for it to warrant protection (even outside the context of personal data), the economic benefits of avoiding or mitigating conflicts of law should not be decisive.

None the less, economic benefits do exist. Beyond the nonquantifiable interest in data sovereignty, a secondary benefit is the ability to select from a broader range of service providers, since more service providers would be capable of providing a service offering that's in line with reasonable expectations of data sovereignty; but this benefit would assume that those service providers are currently not chosen at a substantial rate (which seems to be belied by the market share of potentially affected service providers); and that it would be beneficial to the development of the European internal market to impose requirements that make it easier to acquire services from service providers subject to foreign jurisdictions. Admittedly, avoiding legal proceedings in relation to unlawful data claims is a clear benefit, but such proceedings are relatively rare, and such avoided costs therefore do not weigh heavily.

Additionally, the policy options comprise a broad range of potential measures that could be taken, each of which has a separate **cost** implication. This too is a complex issue, not only because of the wide divergence in the costs of various measures, which could be chose as alternatives or applied cumulatively, but also because of the fact that most conceivable measures are already taken up in at least some instances (i.e. none of the potential measures is fictitious in the sense that there exists no service provider offering it). The goal of the policy intervention, when requiring measures to be taken, is therefore not to introduce measures that do not currently exist at all (inventing a new bar for security), but to incentivise increased adoption (raising the bar for security).

For that reason, in order to assess the feasibility, effectiveness and efficiency of the policy options, we first have to examine the state of the market: which measures exist already, and to what extent are they taken up? In the sections below we will describe some of the measures that could be taken in the implementation of the policy options, along with an assessment of their effectiveness in reducing the problem. Next, we will examine to what extent these measures are already offered or available in the market today, since measures that are already in wide circulation generally will not trigger equally significant costs when they are made a part of a policy option. In effect, measures that are already broadly adopted are a part of the baseline scenario, and thus do not need to be quantified. Finally, a cost/benefit analysis will be done for the policy options, taking into account the available data on costs of measures.

Given the uncertainties in relation to exact numbers, this study will quantify as many benefits and indirect costs as possible, but will need to heavily rely on qualitative analysis as a complement.

3.3.4.1.3 Description of potential measures to mitigate international conflicts of law in the context of the policy options

The policy options as described above have an impact on multiple stakeholders on the value chain, most significantly for cloud providers targeted by the policy option, and for the European Commission itself. Very briefly and conceptually summarised, the following overview of measures can be provided:

Policy option	Stakeholder	Measure to be taken
Policy Option 1: soft non-regulatory options focusing on transparency and/or operational changes	European Commission	• Creation and maintenance of a data sovereignty knowledge centre • Collection and promotion of best practices – model terms and technical measures – nonbinding recommendations
	Cloud providers	No mandatory measure – optional cooperation with nonbinding recommendations.
Policy Option 2: soft regulatory option focusing on transparency	European Commission	Same as policy option 1
	Cloud providers	• Notification duty towards customers • Notification duty towards the platform
Policy Option 3: high impact regulatory intervention - focus on operational change	European Commission	Same as policy option 1
	Cloud providers	• Obligation to implement reasonable legal, technical and organisational measures that reduce or eliminate the possibility of extraterritorial data claims • Obligation to assess the legal validity of non-European data claims
Policy Option 4: high impact regulatory intervention – international alignment	European Commission	• Efforts in international alignment
	Cloud providers	• Alignment to international consensus if it can be achieved

Out of all of the measures to be taken, the most difficult to assess is the obligation for cloud providers under policy option 3 to implement reasonable legal, technical and organisational measures that reduce or eliminate the possibility of extraterritorial data claims. This is a generic criterion, which will need to be assessed and implemented on a case by case basis by cloud providers, making it difficult to assess feasibility, cost and impact generically. In order to determine what the implications for cloud providers would be, it is useful to establish an overview and description of currently available mitigation measures, based on an assessment of best practices and guidance from regulatory bodies[415].

The following overview of relevant measures can be provided, comprising legal, technical and organisational measures, along with an assessment of their effectiveness in reducing cross border conflicts of law:

Legal measures	**Summary description and anticipated impact**
Transparency – notification duty	**What?** A contractual duty to notify cloud customers when a data access request is made by a third country authority that targets their data (or that includes their data) (comparable to the measure required in relation to personal data by article 28(3)(a) of the GDPR). **Effective?** Marginally effective – does not avoid the conflict, but potentially increases its visibility. Effectiveness is likely to be undercut by gag orders (i.e. legally binding instructions from the authority to the cloud customer not to disclose the request to the customers), or by limitations on notification options included in the relevant foreign law itself[416] (not just in the gag order).
Transparency – warrant canary clauses	**What?** A contractual duty to notify cloud customers periodically (e.g. on a weekly basis) that no data access request was made by a third country authority that targets their data (or that includes their data) in the period preceding the request. If no such notification is received, this implicitly means that the data was targeted. **Effective?** Somewhat effective – does not avoid the conflict, but potentially increases its visibility. Effectiveness could still be undercut by gag orders if they are phrased in a way that covers the canary clause, but this would require the authority to be aware of this possibility; or by limitations on notification options included in the relevant foreign law itself. Likely to create false positives when no notice is sent in error

[415] Notably, the EDPB *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data* were taken as a starting point. See https://edpb.europa.eu/sites/edpb/files/consultation/edpb_recommendations_202001_supplementarymeasures transferstools_en.pdf. These were subsequently enriched and nuanced with feedback provided to the consultation in relation to these measures; see https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/recommendations-012020-measures-supplement-transfer_en

[416] By way of example, the USA FREEDOM Act of 2015 requires service providers targeted by FISA to delay any reporting by 6 months and report in bands of 500. Major providers adhere to this requirement - see e.g. https://www.apple.com/legal/transparency/us.html, https://www.microsoft.com/en-us/corporate-responsibility/us-national-security-orders-report?activetab=pivot_1%3aprimaryr2, and https://d1.awsstatic.com/certifications/Information_Request_Report_June_2020.pdf for the reports from Apple, Microsoft and Amazon respectively.

	(which could be construed as an access notification, even though none occurred).
Legal review duty for foreign data claims	**What?** A contractual duty to for the cloud provider to actively and substantively review the legal validity of any data access request from a third country authority. **Effective?** Marginally effective, since it implies that the cloud provider essentially must represent the legal interests of the cloud customer. If the request is valid, the review duty will however have no benefit, even if the request is contrary to EU law.
Legal litigation/protection duty for foreign data claims	**What?** A contractual duty to for the cloud provider to actively and substantively review the legal validity of any data access request from a third country authority, and to refer the requesting authority to any existing international cooperation mechanisms (such as MLATs), or to legally contest the validity of the request using any legal means at its disposal. **Effective?** Somewhat effective, since it implies that the cloud provider must represent the legal interests of the cloud customer, including in legal proceedings. If the request is valid, the legal protection duty will however have no benefit, even if the request is contrary to EU law.
Audit rights for third party	**What?** A contractual right for the cloud customer to ensure that the cloud provider's access to the data in the context of third country authority requests (e.g. instances where it hands over data to a law enforcement body at the request of that body) is audited by a neutral third party. **Effective?** Somewhat effective – does not avoid the conflict, but potentially increases its visibility. Assumes that logs are available for audit that are not tampered with, and that access to these logs is not prohibited by the third country authority.
Audit rights for customer	**What?** A contractual right for the cloud customer to audit the cloud provider's access to the data (including at the behest of a third country authority). **Effective?** Somewhat effective – does not avoid the conflict, but potentially increases its visibility. Assumes that logs are available for audit that are not tampered with, and that access to these logs is not prohibited by the third country authority.

Globally, legal measures are a useful component in mitigating measures, but unlikely to be sufficiently effective in isolation (i.e. in the absence of further organisational or technical measures), since they either ensure only the transparency of conflicts but don't avoid them, or can be overruled and offset by third country legislation and/or orders from third country authorities.

Technical measures	Summary description and anticipated impact
Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider	**What?** Application of effective encryption algorithms that obfuscate stored data to any third parties, and where the encryption keys are held in stewardship by the cloud provider. **Effective?** Somewhat effective. Such encryption disables direct data access by third country authorities, and is therefore an effective security measure; but if the authority can coerce the cloud provider to cooperate by decrypting data, the effectiveness is lost. Furthermore, encryption at rest is less effective if data in transit is decrypted; and unencrypted metadata can still be revealing in its own right.
Encryption for data at rest (including encrypted cloud VMs) using external key management	**What?** Application of effective encryption algorithms that obfuscate stored data to any third parties, and where the encryption keys are held by a third party who is independent from the cloud provider. **Effective?** Highly effective. Such encryption disables data access by third country authorities and the cloud provider. It is effective, unless the third party can also be coerced to cooperate by decrypting data (which is inapplicable if the customer holds the keys under its sole control itself). The model is technically complex and occasionally impossible to apply if the cloud service requires the cloud provider to be able to access data in the clear (e.g. for data maintenance, analytics or support), except through the application of uncommon techniques such as homomorphic encryption. Furthermore, encryption at rest is less effective if data in transit is decrypted; and unencrypted metadata can still be revealing in its own right.
Encryption for data in transit using PKI systems controlled by the cloud provider	**What?** Application of effective encryption algorithms that obfuscate data being transferred between the cloud provider and any other party (including the customer), and where the encryption keys are held in stewardship by the cloud provider. **Effective?** Marginally effective. Useful and common as a security measure that complicates data capture during transit, but does not resolve the more fundamental challenge of third country authorities targeting the cloud provider directly to access data at rest.
Encryption for data in transit using external key	**What?** Application of effective encryption algorithms that obfuscate data being transferred between the cloud provider and any other party (including the customer), where the encryption

management (including zero trust networks)	keys are held by a third party who is independent from the cloud provider, or (in the case of zero trust networks) dynamically generated. **Effective?** Somewhat effective. Useful as a security measure that complicates data capture during transit, but does not resolve the more fundamental challenge of third country authorities targeting the cloud provider directly to access data at rest.
Anonymization or pseudonymization	**What?** Application of data minimisation or transformation techniques that reduce the linkability of the data to a given subject. In this context, the notion therefore refers not to anonymisation or pseudonymisation of personal data exclusively, but rather to the broader field of using techniques to reduce linkability between data sets, or between data and the subjects of that data (which could be natural persons, companies, physical objects, software agents, immaterial goods or rights, etc). **Effective?** Somewhat effective, provided that anonymisation/pseudonymisation happens prior to entrusting the data to the cloud provider, or at least using models that ensure that raw source data is deleted by the cloud provider after it anonymises/pseudonymises data. However, the measure is generally unsuitable in most cloud use cases, since it requires the customers data to be edited, often to the point of becoming unusable in practice.
Split processing (including cryptography based on secret sharing / secret splitting)	**What?** Application of a data processing architecture where no single cloud provider holds all relevant data, and collusion between the provider and at least one other company is required to access the data in an unencrypted form. **Effective?** Highly effective, provided that the additional providers are not subject to the same jurisdiction as the principal cloud provider. However, the model is technically complex and occasionally impossible to apply if the cloud service requires the cloud provider to be able to access data in the clear (e.g. for data maintenance, analytics or support), except through the application of uncommon techniques such as homomorphic encryption.
Multi-party processing by independent providers (including stack splitting)	**What?** Application of a data processing architecture where the cloud provider provides most of the software and assumes responsibility for business arrangements, but relies on an independent third party to store and otherwise process the data, so that the cloud provider cannot enable access to the data by third

	country authorities without collusion from the independent third party. **Effective?** Highly effective, provided that the additional provider is not subject to the same jurisdiction as the principal cloud provider. Account must be taken of the possibility of backdoors in the technology provider by the original cloud provider. Innovation may be slowed due to the need for the independent party to deploy the relevant technology responsibly.
Multi-party processing involving independent verification by a logging service provider	**What?** Application of a data processing architecture where the cloud provider operates the entire service, but relies on an independent third party to maintain log files, so that the cloud provider cannot enable access to the data by third country authorities without detection by the independent third party. **Effective?** Marginally effective, provided that the log provider is not subject to the same jurisdiction as the principal cloud provider. The measure at any rate does not avoid the conflict, but potentially increases its visibility. Assumes that logging by the independent party cannot be impeded or tampered with by the cloud provider.

Globally, technical measures can be effective in mitigating access to data by third country authorities, but they are technologically complex, and not universally applicable to all cloud models (notably being nearly impossible to apply to SaaS cloud models where the cloud provider must be able to access customer data itself).

Organisational measures	**Summary description and anticipated impact**
Periodic certification against a reputable standard	**What?** Commitment to undergo periodic third party (re-)certification of the cloud service against a disclosed and reputable standard. **Effective?** Somewhat effective. While the certification process improves governance and accountability over the cloud provider in general, there is no assurance that the certification (or the certification scheme) actually addresses the accessibility of the data by third country authorities.
Intra-group policies within the provider aiming to support minimisation of data sharing	**What?** The implementation of a policy creating a separation of duties within the cloud provider, wherein a separate team within the cloud provider ensures that no more data is provided in response to an access request from a third country authority than is strictly required.

	Effective? Somewhat effective. The process improves governance and accountability within the cloud provider in general, but provides no assurance that request is substantively assessed before being granted.
Intra-group policies within the provider aiming to support accountability	**What?** The implementation of a policy creating a separation of duties within the cloud provider, wherein a separate team within the cloud provider verifies the legal validity of any access request from a third country authority, and assumes responsibility for this assessment. **Effective?** Somewhat effective. The process improves governance and accountability within the cloud provider in general, but provides no structural solution to potential conflicts of law.
Customer control over data access and logs via dashboards	**What?** The implementation of specific interfaces that allow the customer to monitor access and usage logs in relation to their data. **Effective?** Somewhat effective. While interfaces certainly offer substantial security benefits, there is no inherent assurance that they cover all access cases. If access by a third country authority (including via the intermediation of the provider) is not shown, the measure is ineffective in mitigating international conflicts. In addition, the measure does not avoid the conflict, but potentially increases its visibility.
Customer control via rights-based access control techniques (RBAC) or other technical compliance tools	**What?** The implementation of specific interfaces that allow the customer to control access and usage rights in relation to their data. **Effective?** Somewhat effective. While interfaces certainly offer substantial security benefits, there is no inherent assurance that they cover all access cases. If access and usage rights can be overridden by administrative accounts within the cloud provider (including at the request of a third country authority, the measure is ineffective in mitigating international conflicts.
E-evidence and data freeze services, in the context of legal proceedings that require electronic evidence to be isolated from further manipulation in	**What?** The implementation of specific interfaces that allow the customer to limit access and usage rights in relation to their data in the context of specific judicial proceedings.

order to preserve its integrity for the proceedings.	**Effective?** Somewhat effective. The interfaces are not intended to isolate data from access by public authorities, but techniques that generally aim to protect data against external manipulation could provide useful elements to protect data more broadly against third party access or interference. While interfaces certainly offer substantial security benefits, there is no inherent assurance that they cover all access cases. If the process allows data to be frozen but without limiting its accessibility to third parties, the measure is ineffective in mitigating international conflicts.

Globally, organisational measures are somewhat effective in mitigating access to data by third country authorities, although no organisational measure is effective in situations where the cloud provider has control over the implemented measures and where it can be compelled to cooperate with third country authorities.

3.3.4.1.3.1 Prevalence of these measures in the market

The cost of measures is determined to a large extent by their prevalence and availability in the market. More specifically, measures that are already widely adopted will have only a limited costs, since most cloud providers will no longer need to invest to implement them. In effect, they are largely a part of the Baseline scenario rather than of any policy option. Inversely, measures that are unavailable under economically feasible terms likely must be discounted to some extent as well, since they will never be implemented under an obligation that imposes *reasonable* legal, technical and organisational measures on cloud providers.

For that reason, in the sections below we will examine to what extent the listed measures are available in the market today. Two different segments of the cloud industry are provided:

- Firstly, **leading** (in terms of capitalisation, turnover or market share) **cloud service providers** are examined, covering both SaaS and IaaS providers. They are relevant since the scope of their activities and the scale of their budgets makes it more plausible that they have already adopted measures that are ‘reasonable’ in terms of effectiveness and economic viability. In effect, they are used as a yardstick on the adoption of the state of the art.
- Secondly, a range of **specialised innovators** is examined, each specialised in a specific measure that’s not yet routinely taken up by the market. They are relevant since they focus on niche measures that are designed specifically to support sovereignty, and therefor act as leading indicators of likely developments of the market, irrespective of whether their services are taken up at a wide scale currently.

In the sections below, a sample of both categories of service providers will be examined.

3.3.4.1.4 Uptake by leading cloud service providers

By way of an assessment of the state of the art, in the sections below we will briefly examine which of the measures listed above are already adopted (either by default or offered as a free or paid option) in six leading cloud services. These include:

- Three leading **SaaS providers**. Microsoft, Salesforce, Adobe, SAP, and Oracle account for half of the total enterprise SaaS market share worldwide (Statista 2021). Microsoft, Salesforce and Adobe are included in the sample below.
- Three leading **IaaS providers** - Amazon Web Services (AWS), Microsoft and Google comprise the 2021 Gardner Magic Quadrant leaders, and are therefore included in the analysis below.

More in detail, the following service offering is assessed:

SaaS providers	**Summary description of the cloud service**
Microsoft 365 (previously Office 365; assessed at the Microsoft 365 Business Premium level)	Microsoft 365 comprises a broad range of subscription based SaaS productivity services (https://www.office.com/) including the Microsoft Office software suite, and various cloud-based software-as-a-service products for business environments, such as hosted Exchange Server and SharePoint.
Salesforce	Salesforces provides a SaaS customer relationship management (CRM) service, coupled with a suite of enterprise applications focused on customer service, marketing automation, analytics, and application development.
Adobe Document Cloud (selected due to its relevance in terms of information security)	Document Cloud is a SaaS based document editing, managing, signing and collaboration solution.

IaaS providers	**Summary description of the cloud service**
Amazon Web Services (AWS)	AWS is a portfolio of services providing on-demand cloud computing platforms and APIs. These cloud computing web services provide a variety of basic abstract technical infrastructure and distributed computing building blocks and tools. The services include Amazon Elastic Compute Cloud (EC2), S3 scalable storage, and a broad range of security related services
Microsoft Azure	Azure comprises a broad range of cloud services (https://azure.microsoft.com/en-gb/services/) that comprises IaaS, PaaS and SaaS components, geared towards building, testing, deploying, and managing applications and services through Microsoft-managed data centres. This particular assessment focuses on its IaaS offering, which is available both in public and hybrid deployment models.

Google Compute Engine	Google Compute Engine is Google's principal IaaS offering, allowing users to create and run virtual machines on Google's infrastructure (https://cloud.google.com/compute). Recently expanded with Google Cloud Confidential Computing, offering fully encrypted VMs also for data at rest.

For each of these services, a commented heatmap is presented in the sections below, containing each measure described above, and indicating:

- Whether the measure is supported by default (marked in green)
- Whether the measure is available as an option or with certain limitations (marked in yellow)
- Whether the measures are unavailable (or there is no indication of its availability) (marked in red)

The assessment was made based on publicly available information, including service agreements, published policies, user interfaces, compliance statements, FAQs, and similar information made available by the service provider. Where relevant, comments are added in italics in order to clarify why a specific assessment was made.

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider	Periodic certification against a reputable standard
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management *Supported via Customer Lockbox (Customer Key, available for Exchange Online, SharePoint Online, and OneDrive for Business; not available at the Microsoft 365 Business Premium level)*	Intra-group policies within the provider aiming to support minimization of data sharing
Legal review duty for foreign data claims	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability

[417] See https://docs.microsoft.com/en-us/office365/servicedescriptions/office-365-platform-service-description/office-365-securitycompliance-center

Legal litigation/protection duty for foreign data claims ***Not contractually guaranteed, but public case law shows active litigation***	Encryption for data in transit using external key management (including zero trust networks) *Only for authentication services, building on Azure AD*	Customer control via dashboards *Supported via dedicated management tools, including log analytics*
Audit rights for third party	Anonymization or pseudonymization *Log pseudonymization is available.*	Customer control via RBAC or other technical compliance tools *Supported via dedicated management tools, including Security & Compliance Center and Privileged Access Management*
Audit rights for customer ***Basic audit logging is always included, but this is a log search function; not an audit right***	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services *eDiscovery services are available, but only at a basic level (Advanced eDiscovery exists, but not at the Microsoft 365 Business Premium level)*
	Multi-party processing by independent providers	

	(including stack splitting) Multi-party processing involving independent verification by a logging service provider	

Salesforce - Table of supported / available measures[418]

Legal measures	Technical measures	Organisational measures
Transparency – notification ***Covered by Salesforce's Principles for Government Requests for Customer Data***	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider *Available as a paid option - Shield Platform Encryption*	Periodic certification against a reputable standard
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management *Available as a paid option - Shield Platform Encryption. Keys can be stored elsewhere and fetched dynamically when needed*	Intra-group policies within the provider aiming to support minimization of data sharing *Explicitly covered by its internal Code of Conduct*

[418] See https://compliance.salesforce.com/en, https://security.salesforce.com/, https://www.salesforce.com/company/legal/ and https://trust.salesforce.com/en/trust-and-compliance-documentation/

Legal review duty for foreign data claims ***Covered by Salesforce's Principles for Government Requests for Customer Data***	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability *Explicitly covered by its internal Code of Conduct*
Legal litigation/protection duty for foreign data claims ***Covered by Salesforce's Principles for Government Requests for Customer Data (review and challenge commitment)***	Encryption for data in transit using external key management (including zero trust networks)	Customer control via dashboards *Supported via dedicated management tools, including log analytics and a Real Time Event Monitoring Service*
Audit rights for third party ***Extensive audits/certifications are maintained; see also the Salesforce Processor BCR***	Anonymization or pseudonymization	Customer control via RBAC or other technical compliance tools *Role management and allocation are possible*
Audit rights for customer ***Exceptionally permissible under the Salesforce Processor BCR***	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services
	Multi-party processing by independent	

	providers (including stack splitting) Multi-party processing involving independent verification by a logging service provider	

Adobe Document Cloud - Table of supported / available measures[419]

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider *Basic encryption under Adobe's control is activated by default. Users can opt to add an extra layer of encryption, using a key generated (and controlled) by Adobe*	Periodic certification against a reputable standard *Large number of certifications published and maintained online*
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management	Intra-group policies within the provider aiming to support minimization of data sharing

[419] See https://helpx.adobe.com/enterprise/admin-guide.html/enterprise/using/content-logs.ug.html, https://acrobat.adobe.com/content/dam/doc-cloud/en/pdfs/Document-Cloud-Security-Overview.pdf and https://www.adobe.com/legal/lawenforcementrequests.html

Legal review duty for foreign data claims ***Adobe applies specific Guidelines for law enforcement seeking customer data, under which it structurally assesses compliance with Irish law (for non-US customers)***	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability *Adobe applies a Security Culture policy with internal Champions and Security Specialists*
Legal litigation/protection duty for foreign data claims	Encryption for data in transit using external key management (including zero trust networks) *Supported via Adobe's Zero-Trust Enterprise Network Platform*	Customer control via dashboards *Access to logs is available*
Audit rights for third party	Anonymization or pseudonymization	Customer control via RBAC or other technical compliance tools *Authorisation rights can be managed*
Audit rights for customer	Split processing (including cryptography based on secret sharing / secret splitting) Multi-party processing by	E-evidence and data freeze services

	independent providers (including stack splitting) Multi-party processing involving independent verification by a logging service provider	

Amazon AWS - Table of supported / available measures

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider *Available via AWS CloudHSM and AWS Key Management Services*	Periodic certification against a reputable standard *Large number of certifications published and maintained online*
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management *AWS Key Management Service can use AWS CloudHSM as a custom key store. The system is still internal, however.*	Intra-group policies within the provider aiming to support minimization of data sharing

Legal review duty for foreign data claims	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability
Legal litigation/protection duty for foreign data claims	Encryption for data in transit using external key management (including zero trust networks)	Customer control via dashboards *Supported via dedicated management tools, including AWS Audit Manager and AWS Security Hub*
Audit rights for third party	Anonymization or pseudonymization *Could be implemented by customer. Log pseudonymization is available*	Customer control via RBAC or other technical compliance tools *Supported via dedicated management tools, including log analytics*
Audit rights for customer ***Supported via dedicated management tools, including AWS Audit Manager***	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services
	Multi-party processing by independent providers (including stack splitting)	

Support for hybrid models, including via Amazon Outpost
Multi-party processing involving independent verification by a logging service provider

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider *Supported via dedicated key vault*	Periodic certification against a reputable standard *Large number of certifications published and maintained online*
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management	Intra-group policies within the provider aiming to support minimization of data sharing
Legal review duty for foreign data claims	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability

		Adobe applies a Security Culture policy with internal Champions and Security Specialists
Legal litigation/protection duty for foreign data claims ***Not contractually guaranteed, but public case law shows active litigation***	Encryption for data in transit using external key management (including zero trust networks) *Only for authentication services*	Customer control via dashboards *Supported via dedicated management tools, including log analytics*
Audit rights for third party	Anonymization or pseudonymization *Could be implemented by customer. Log pseudonymization is available.*	Customer control via RBAC or other technical compliance tools *Supported via dedicated management tools, including log analytics*
Audit rights for customer	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services *Not applicable for IaaS*
	Multi-party processing by independent providers (including stack splitting) *Support for hybrid models*	

	Multi-party processing involving independent verification by a logging service provider	

Google Compute Engine - Table of supported / available measures[420]

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider *Supported via Google Cloud Confidential Computing and dedicated key management services*	Periodic certification against a reputable standard *Large number of certifications published and maintained online*
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management *Compute Engine allows user to generate and manage their own keys, which are not stored by Google*	Intra-group policies within the provider aiming to support minimization of data sharing

[420] See https://cloud.google.com/compute/confidential-vm/docs/about-cvm, https://cloud.google.com/security/encryption-at-rest, https://transparencyreport.google.com/, https://cloud.google.com/security/infrastructure?hl=hr, https://cloud.google.com/files/gcp-trust-whitepaper.pdf and https://cloud.google.com/terms/data-processing-terms

Legal review duty for foreign data claims	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability
Legal litigation/protection duty for foreign data claims	Encryption for data in transit using external key management (including zero trust networks) *Available through the BeyondCorp services in Google*	Customer control via dashboards *Supported via dedicated management tools, including log analytics*
Audit rights for third party	Anonymization or pseudonymization *Available via Cloud Data Loss Prevention services as a paid option.*	Customer control via RBAC or other technical compliance tools *Supported via dedicated management tools, including log analytics*
Audit rights for customer	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services *Not applicable for IaaS. Google Vault is available for the (SaaS) Google Workspace services.*
	Multi-party processing by independent	

	providers (including stack splitting) *Support for hybrid models*	
	Multi-party processing involving independent verification by a logging service provider	

3.3.4.1.5 Conclusions in relation to leading cloud service providers

As the overview above shows, all examined leading providers already implement multiple mitigating measures. Aggregating all responses into a global heatmap to determine general trends and patterns, the following overview could be provided, where:

- Green measures are offered by 5 or 6 providers
- Yellow measures are offered by 3 or 4 providers
- Red measures are provided by 1 or 2 providers

Legal measures	Technical measures	Organisational measures
Transparency – notification	Encryption for data at rest (including encrypted cloud VMs) using key management services by the cloud provider	Periodic certification against a reputable standard
Transparency – warrant canary clauses	Encryption for data at rest (including encrypted cloud VMs) using external key management	Intra-group policies within the provider aiming to support minimization of data sharing

🟥	🟨	🟥
Legal review duty for foreign data claims	Encryption for data in transit using PKI systems controlled by the cloud provider	Intra-group policies within the provider aiming to support collaboration and accountability
Legal litigation/protection duty for foreign data claims	Encryption for data in transit using external key management (including zero trust networks)	Customer control via dashboards
Audit rights for third party	Anonymization or pseudonymization	Customer control via RBAC or other technical compliance tools
Audit rights for customer	Split processing (including cryptography based on secret sharing / secret splitting)	E-evidence and data freeze services
	Multi-party processing by independent providers (including stack splitting)	
	Multi-party processing involving independent verification by a logging service provider	

Thus, out of the examined measures, 7 already occur (nearly) systematically with the examined large providers:

- Transparency – notification of the customer
- Encryption for data at rest using key management services by the cloud provider
- Periodic certification against a reputable standard
- Encryption for data in transit using PKI systems controlled by the cloud provider
- Customer control via dashboards
- Customer control via RBAC or other technical compliance tools
- Audit rights for third party

Out of these, the notification duties, certification, encryption using internal systems and role-based access controls are available literally universally (although the scoping of these measures of course differs very substantially in practice). They can therefore be considered a common element in the state of the art (and therefore would be a part of the baseline scenario that either already is common among most providers, or would likely become increasingly common even without policy intervention).

Inversely, warrant canary clauses are never encountered, nor are split processing or independent verification by an external logging service providers. The latter are likely perceived as having too little added value for canary clauses, or to be too niche to be effective in the current state of play.

3.3.4.1.6 Service offering from specialised innovators

Of course, additional measures can be built onto cloud service offerings by relying on independent innovators. These can offer a broad range of services that directly translate in increased protections against third country authority interference. A part of their value lies precisely in the fact that they are independent from the principal cloud service provider, meaning that, in order to hide access to data by a third country authority, it would no longer be sufficient to compel the cloud provider to cooperate; the secondary service provider would need to be implicated as well. This can be significantly more complex, especially when that provider is established in a different jurisdiction from the principal cloud service providers.

While an exhaustive overview of such innovative complementary services is of course out of scope, hereunder we will examine a short selection of relevant companies, along with a description of the scope and relevance of their services. The main objective is to get initial insights on the potential role of such service providers.

Case 1 – Boxcryptor	**https://www.boxcryptor.com/**
Service description	Boxcryptor encrypts selected files and folders in supported SaaS storage services. Functionality of such services is retained, but encryption is applied before sending the data to the cloud providers. Boxcryptor applies the zero-knowledge paradigm, and is itself incapable of decrypting files.
Application	Can be deployed on supported SaaS storage services, including Dropbox, Google Drive, Apple iDrive, Microsoft OneDrive, and the Microsoft Teams collaborative spaces.
Effectiveness	Highly effective (assuming reliable encryption). Such encryption disables data access by third country authorities and the cloud provider. It is effective, since Boxcryptor is incapable of decrypting data even if targeted by a foreign authority. However, the model is only useful for cloud storage models. Any advanced cloud service that requires the cloud provider to be able to access data in the clear (e.g. for data maintenance, analytics or support) would not qualify as a use case.

Case 2 - KinectIQ - https://www.knectiq.com/

Service description	KnectIQ is a zero-knowledge networking systems integrator. The software can be integrated to any connected location, providing end to end communication in any of the applications below. Provides dockerized containers in Linux Servers and PostgreSQL DB; supported endpoints are generic. Windows, Linux, Mac PCs and Android/iOS devices.
Application	Can be deployed in generic IaaS containers (including most common VMs); or integrated into SaaS services
Effectiveness	Somewhat effective. Useful as a security measure that complicates data capture during transit, but does not resolve the more fundamental challenge of third country authorities targeting the cloud provider directly to access data at rest

Case 3 - TrustArc - https://trustarc.com/

Service description	Provider of a suite of privacy management software (comprising most functions: record keeping, risk assessment, impact assessment, cookie consent management, etc), and operator of the TRUSTe certification and assurance scheme.
Application	Can be applied to any service; not cloud specific
Effectiveness	Somewhat effective. While the certification process and use of a standardised software suite improves governance and accountability over the cloud provider in general, there is no assurance that the certification (or the certification scheme) actually addresses the accessibility of the data by third country authorities.

Case 4 - Darktrace - https://www.darktrace.com/

Service description	AI enabled and cloud-based threat detection system that can detect and block potential intrusions and other attacks. The 'Enterprise Immune System' learns normal 'patterns of life' to discover unpredictable cyber-threats and block their effectiveness.
Application	Can be applied to many cloud services (IaaS or SaaS, including Microsoft 365, Azure or Salesforce) and to their connections to endpoints and the corporate network.
Effectiveness	Somewhat effective, since it allows third country attacks to be detected, logged and potentially deflected. It would not be capable to detect intrusions directly in the cloud providers infrastructure that don't involve external attacks, which undercuts some of the impact.

Case 5 - AWS Outposts - https://aws.amazon.com/outposts/

Service description	AWS Outposts is a fully managed service that offers the same AWS infrastructure, AWS services, APIs, and tools to virtually any datacentre, co-location space, or on-premises facility for a hybrid experience.
Application	AWS compute, storage, database, and other services run locally on Outposts. Interactions with AWS services are possible too.

	Obviously, interactions are limited to AWS instances (although the model conceptually could be viable for other cloud services too).
Effectiveness	Somewhat effective since the data resides on a separate system. However, the stack is provided by Amazon, so that trust in the original cloud service provider is still required. Additionally, some use cases will require interactions with other AWS services outside of the Outpost, which compromises the benefits to some extent.

As the summary overview shows, some of the examined measures can thus also be built on top of standardised cloud service packages. Even in cases where there appears to be a clear overlap with existing pre-deployed measures (e.g. encrypting data before sending it to a cloud provider who will re-encrypt the data), this can increase security since it stops the cloud provider from disabling the effectiveness of any given measure, unless the secondary provider can also be compelled to cooperate. It is however also clear that no measure is universally suitable or viable: a case by case assessment of feasibility and suitability is always required.

3.3.4.1.7 Observations on potential measures included within the policy options and on CBA data

Prior to examining costs and benefits, some preliminary observations can already be made on the relevance and effectiveness of potential measures.

Firstly, it should be recognised that not all measures are conceptually possible for all types of service providers. Especially for technical measures, their applicability and feasibility depend on the service and provisioning model, and it is not possible to standardise these. By way of example, effective encryption of data at rest, where data is encrypted prior to sending it to a cloud provider, is not possible in SaaS models where the cloud provider can only offer its services if the data are accessible to the cloud provider in a decrypted form[421]. In contract, most legal and operational measures are more broadly feasible, although they may not be appropriate for all services. Given this environment, standardising a set of measures that must be implemented by all providers is challenging, if not impossible.

Secondly, it is important to note that the examined measures are currently offered in three principal ways, as is also reflected in the analysis above:

- Some measures are simply **built into a service offering by default** – no choices or additional services must be selected by customers, and **no additional cost applies**. In practical terms, these costs can be considered part of the state of the art. This is the case for notification duties, certification, encryption using internal systems and role-based access controls, which the analysis shows to be available (nearly)-universally without added costs.
- Other measures are **optional but standardised**, and provided by cloud providers or by third parties to the end users at an **additional cost borne by the customer,** reflected in an **increased subscription fee**. By way of examples:
  - A basic Microsoft Azure server is available as of €0,0044/hour. Opting for a higher performance server with additional security and confidentiality guarantees can cost up to €1,3342/hour[422]. Adding Extensible Key Management services will add €0,026/10.000 transactions. Costs can thus increase substantially.
  - Similarly, a basic Office 365 subscription costs for business use costs 4,20 € per user per month. Adding Advanced Threat Protection raises the price to 16,90 € per user per month[423].

[421] Outside of more niche technologies such as homomorphic encryption, which are not commonly found in the market today.
[422] https://azure.microsoft.com/nl-nl/pricing/details/virtual-machines/linux/
[423] https://www.microsoft.com/nl-be/microsoft-365/business/compare-all-microsoft-365-business-products

  - In the Salesforce pricing model, adding the Shield Platform Encryption measure increases the net price of the subscription fee by 30% (assuming full option protection), calculated on how much the customer spends on other applicable, technically compatible Salesforce products[424].
  - Adding external services on a subscription basis can similarly double subscription prices: adding e.g. Boxcryptor encryption to a SaaS storage solution adds 10 € per user per month [425]; more tailored firewall and protection solutions can easily imply one-time set-up costs between 5.000 and 100.000 EUR € even for SME customers; and an additional monthly cost of around 10 to 20 € per user per month.
- Finally, other measures are custom and must be integrated in a tailored fashion into specific applications and infrastructures. They are not generally purchased by customers, but rather by ICT service providers themselves. Such services commonly have a very high costs, with set-up and integration costs between 100.000 EUR and 5 million EUR; and annual recurring costs that are linked to the use case. By way of example, Amazon Outpost carries a cost of $126,104.75 to $935,233.39 per system unit to be deployed[426] (depending on the sophistication of the unit. Cost relates to EMEA rates, assuming full upfront payment). This refers only to the cost of the Outpost units, not to any deployment or integration costs to be assumed by the buyer.

Thus, technical investments can have a significant impact on costs, generally ranging from a 30% to 400% price increase compared to a baseline price for subscription based services; but as the overview above showed, they are also generally the most effective in terms of diminishing the risk of unlawful third country access requests. Legal and organisational measures on the other hand tend require mainly the engagement of appropriate HR profiles to assess data requests, to take legal action if needed, and to communicate towards customers. While such costs can also be substantial, the available data suggests that most cloud providers already have appropriate resources in place to a large extent. Notifications to cloud customers are already very common as a contractual guarantee (wherever permitted in the light of potential gag orders), and many providers already formulate publicly accessible policies for vetting the prima facie lawfulness of data access orders before responding to them. Even the use of transparency portals towards the customers is no longer uncommon with large cloud providers, as is the definition of specific internal responsibility allocations ensuring that data access requests are handled in an independent and accountable manner. Thus, while some cloud providers would undoubtedly need to invest further, this investment is principally necessary to establish parity with the status quo in the market.

3.4 Comparison of the policy options

The aim of this section is to compare of the policy options in order to identify the preferred policy option for each of the domains. The following MCA has been performed in line with the European Commission's *Better Regulation Guidelines*[427] and its toolbox[428], most importantly tool 63[429]. The assessment builds on the prior analysis of each individual option.

It has been concluded in the previous section that for none of the area under investigation the baseline will be able to achieve the desired results and resolved identify problems. The assessment concludes that a policy intervention is needed. It remains to be seen the type (regulatory vs. non regulatory) and the intensity (low vs. high) of intervention. The MCA will assess which of the three policy options under each area is the most adequate:

- B2G data sharing for the public interest

[424] https://www.salesforce.com/editions-pricing/platform/shield/
[425] https://www.boxcryptor.com/en/pricing/for-teams/
[426] https://aws.amazon.com/outposts/pricing/
[427] http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm
[428] http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm
[429] https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-63_en_0.pdf

  - **PO 1:** Soft measures and recommendations encouraging B2G data access and reuse practices to unleash the potential benefits of B2G data collaboratives.
  - **PO 2:** Low-intensity regulatory intervention to achieve a fair balance between protecting the legitimate interests of the private sector and their incentives to invest in data value generation, and unleashing the full potential of private sector data for the benefit of society.
  - **PO 3:** High-intensity regulatory intervention to achieve a fair balance between protecting the legitimate interests of the private sector and their incentives to invest in data value generation, and unleashing the full potential of private sector data for the benefit of society.
- Measures supporting citizen empowerment
  - **PO 1:** Voluntary white button scheme plus reciprocity
  - **PO 2:** Compulsory but limited data sharing with reciprocity and possibility for charging for data
  - **PO 3:** Compulsory data sharing without reciprocity and no possibility for charging for data
- Measures on rights on co-generated data and B2B data sharing
  - **PO 1:** Non regulatory option focusing on the establishment of an industry-driven self-regulation framework for co-generated data
  - **PO 2:** Low intensity regulatory option focusing on the adoption of a legal instrument aiming to bring legal certainty and promote contractual fairness for accessing and (co-)using IoT co-generated data
  - **PO 3:** High intensity regulatory option focusing on the adoption of a legal instrument clarifying access and usage rights of co-generated data.
- Measures supporting companies in cases of conflict of laws at international level
  - **PO 1:** Soft non-regulatory options focusing on transparency and/or operational changes
  - **PO 2:** Soft regulatory option focusing on transparency
  - **PO 3:** High impact regulatory intervention - focus on operational change

The MCA was carried out in the following three distinct steps:

- ***Step 1***: Establish indicators or assessment criteria against which the policy options are assessed and compared. This includes establishing the performance of a policy option (i.e. the magnitude of its impact), the weight of the criteria in relation to each other, as well as the direction of the impact (negative/positive). The indicators are established in an analytical grid;
  - Effectiveness
  - Efficiency
  - Coherence
  - Feasibility (legal and political)
- ***Step 2***: Build an outranking matrix in which the scores for all policy options and criteria are provided in order to summarise how the policy options compare with each other in relation to established criteria; and
- ***Step 3***: Prepare a permutation matrix that enables the selection of a final ranking of all the possible policy options against each other for each domain. This means that it is possible not only to select a preferred policy option but also a ranking of all other options against each other.

3.4.1 Assessment criteria and indicators

The following assessment criteria were agreed with the European Commission for the assessment of the impacts of the options. A weight has been defined for each criterion. The direction of the change desired are all positive. The proportionality assessment criteria is considered as an exclusion criteria, and is therefore not included in the MCA.

Table 83 – Weight, direction and performance value allocated to the assessment criteria

Assessment criterion	Weight	Direction	Performance value

Effectiveness			Qualitative +/-3 scale
Efficiency			Qualitative +/-3 scale
Coherence			Qualitative +/-3 scale
Legal and political feasibility			Qualitative +/-3 scale
Proportionality	This exclusion criteria will not be assessed as part of the MCA	N/A	N/A

Based on the results of the Cost-Benefit analysis and the qualitative assessment of each individual options, we have drafted an **input grid** for each domain in which the scores for all policy options are collected and compared in relation to each criterion towards each other. With regard to Measures supporting citizen empowerment ('human-centric data economy'), the assessment has been performed individually for the sectors "Smart home appliances" and "Fitness trackers".

Table 84 – Input Matrix

Input matrix			Business-to-Government data sharing for the public interest					
			PO1 - Non regulatory		PO2 - Low intensity		PO3 - High intensity	
Criteria	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness	0,3	1	1	0,3	2	0,6	2,5	0,9
Efficiency	0,3	1	0,5	0,15	2,5	0,75	1,75	0,525
Coherence of the policy options	0,25	1	1,5	0,375	2	0,5	2	0,5
Legal and political feasibility	0,15	1	2,5	0,375	2	0,3	1,5	0,225
			Measures supporting citizen empowerment ('human-centric data economy') - Smart home appliances					
			PO1 - Non regulatory		PO2 - Low intensity		PO3 - High intensity	
Criteria	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness	0,3	1	1,75	0,525	1,5	0,45	2,25	0,675
Efficiency	0,3	1	2	0,6	1,75	0,525	1	0,3
Coherence of the policy options	0,25	1	0	0	-1	-0,25	-1	-0,25
Legal and political feasibility	0,15	1	0	0	2	0,3	1	0,15
			Measures supporting citizen empowerment ('human-centric data economy') - Fitness trackers					
			PO1 - Non regulatory		PO2 - Low intensity		PO3 - High intensity	
Criteria	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness	0,3	1	1,75	0,525	2,25	0,675	2,5	0,75
Efficiency	0,3	1	2	0,6	2,15	0,645	1,75	0,525
Coherence of the policy options	0,25	1	0	0	-1	-0,25	-1	-0,25
Legal and political feasibility	0,15	1	0	0	2	0,3	1	0,15
			Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing					
			PO1 - Non regulatory		PO2 - Low intensity		PO3 - High intensity	
Criteria	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness	0,3	1	1	0,3	1,5	0,45	2	0,6
Efficiency	0,3	1	1	0,3	1,5	0,45	0,5	0,15
Coherence of the policy options	0,25	1	1	0,25	1	0,25	0,5	0,125
Legal and political feasibility	0,15	1	1	0,15	1	0,15	1	0,15
			Measures supporting companies in cases of conflict of laws at international level					
			PO1 - Non regulatory		PO2 - Low intensity		PO3 - High intensity	
Criteria	Weight	Direction	Performance	Weighted performance	Performance	Weighted performance	Performance	Weighted performance
Effectiveness	0,3	1	0,5	0,15	1	0,3	2	0,6
Efficiency	0,3	1	1,5	0,45	1,5	0,45	1	0,3
Coherence of the policy options	0,25	1	1,5	0,375	1,75	0,44	2	0,5
Legal and political feasibility	0,15	1	2,5	0,375	2	0,3	1,5	0,225

3.4.1.1 Business-to-Government data sharing for the public interest

The analysis of the different policy options allows to draw a preliminary comparison of the different options, based on the current available evidence.

Regarding the soft measures/recommendations (policy option 1), it remains unclear whether these would result in a larger number of Member States following all the recommendations and setting up structures to enhance the access and reuse of data for the public interest. When asked their view, **participants to the 30 March 2021 workshop expressed PO1 as likely to not increase or bring a minor increase of B2G data access/reuse practices** due to the non-binding nature of the intervention. From the interviews and the workshop, it can be inferred that **the policy options that seem to have been most favoured are the other two options, PO2 and PO3**.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 85 – Summary comparison table between low/high intensity policy interventions for Business-to-Government (B2G) data sharing for the public interest

	Low-intensity policy intervention (PO2)	**High-intensity policy intervention (PO3)**
Efficiency	This option would bring costs for each Member State. However, it would also result in very significant time/resource savings, and other multiple benefits as described in section 3.3.1., that overall, the benefits would likely outweigh the costs.	This option would create higher costs than PO2 (linked to the obligations to create a data steward function, having a decision-making body composed by public parties only within the national structure and obligations to share data with marginal costs for dissemination covered resulting in opportunity costs for the private sector, particularly those that currently sell data to the public sector). In terms of benefits, this policy option may bring additional ones to those identified under policy option 2. However, it is not very clear the extent to which the benefits would increase or differ from those identified under policy option 2, especially for the private sector.
Effectiveness	Contributes to achieving the specific and general objectives.	Contributes, to a certain degree more than PO2, to achieving the specific and general objectives, as it also includes an obligation to designate a data steward function in all private and public organisations over a certain size.
Coherence	Coherent with EU law and with national laws (where those exist), in view of the existing tradition of balancing public and private interest in determining data access and use rights.	Coherent with EU law and with national laws (where those exist), in view of the existing tradition of balancing public and private interest in determining data access and use rights.
Legal/political feasibility	This option appears to be legally and politically feasible, in particular due to the greater influence of national laws and the affected companies.	This option appears to be legally and politically feasible, but arguably may face more political opposition due to the greater EU harmonisation of permissible remuneration and due to the mandatory appointment of data stewards based on EU level criteria, leaving less margin for national appreciation.
Proportionality	This option appears to be proportionate.	This option does not seem to be proportionate to the same extent as PO2, particularly due to the designation of the data steward function based

	on the size of the company rather than on the type of data or amount of requests a company may receive annually from public authorities.

The higher-intensity policy intervention (Policy Option 3) is the option likely to achieve best the policy objectives, as it is also the most ambitious. This policy option encompasses the designation of a national structure assisting with and overseeing response to B2G data access and reuse practices, obligations for public sector to ensure veracity of results and independence of public sector action, and obligations for the private sector to identify the data that can be valuable for the public interest purposes identified after a request from a public authority is submitted (Policy Option 2), with additional elements aimed at enhancing B2G data access and reuse practices for the public interest.

In terms of coherence, both policy option 2 and 3 are coherent with existing EU laws and with national laws in relation to B2G data sharing (which are scarce, but do exist in some countries, as described above). While the policy options do affect to some extent existing legal protection frameworks that aim to create preconditions for data sharing (such as copyright, database rights, trade secrets, and data protection law), none of these frameworks is absolute. All of these already contain provisions that establish specific exceptions where this would serve certain public interest goals. It would therefore be possible to adopt a new instrument that clarifies the relationship with these existing frameworks explicitly (without necessarily requiring that these frameworks themselves are changed), e.g. by creating a legal basis for responding to B2G data sharing requests. Policy options 2 and 3 would essentially create a new framework in a B2G context, which is coherent with the existing practice of balancing the public and private interest in establishing access and usage rights to data. Similarly, the designation of national authorities to oversee B2G data sharing (and in the case of policy option 3, the introduction of data stewards) is coherent with comparable policy choices in other data oriented legal frameworks.

In terms of legal and political feasibility, both policy options 2 and 3 appear to be feasible, although arguably policy option 2 is the politically more acceptable option. The principal difference between both policy options is the greater margin for national considerations under policy option 2, which comprises greater involvement of the private sector in identifying eligible data and in agreeing on appropriate conditions for access and use. In contrast, policy option 3 is more prescriptive, particularly in terms of setting a pre-fixed and limited remuneration for companies whose data might be shared, and in requiring a data steward to be appointed based on EU level criteria. This may make policy option 3 less politically palatable, also taking into account that policy option 3 allows less possibility for Member States to define public interest justifications for which they would require B2G data sharing. Since policy option 2 grants Member States a broader margin for national policies on that point, it indirectly also allows Member States to more easily include or exclude certain industries from the data sharing obligation, which may be seen as a politically favourable possibility by some Member States.

In terms of proportionality, PO2 seems to be proportionate. PO3 does not seem to be proportionate to the same extent as PO2, particularly due to the designation of the data steward function based on the size of the company rather than on the type of data or amount of requests a company may receive annually from public authorities. By comparing the annual costs a company would likely incur in order to comply with the data steward function, we see that these are only proportionate to the number of requests a company may receive annually. We calculated, based on the costs and number of FTEs that several stakeholders mentioned it would take to comply with PO3, that the break-even point is from 21 requests for smaller companies where 2 FTEs would be designated as data stewards, to 54 requests for big-sized companies where 5 FTEs may be necessary to address all requests. However, if we assume that for a specific type of data, only a single national public authority may be interested in, resulting in fewer requests, it would be more costly to have the data steward function than to address the request without one. Therefore, for specific cases and for certain types of data or companies, the data steward function may bring more costs than benefits. Due to this, the size of a company may not be as beneficial as to focusing to the type of data or the number of requests the specific company will actually have.

After assessing the elements of efficiency, effectiveness, coherence, legal and political feasibility, and proportionality, it can be concluded that policy option 2 seems to be the preferred policy option as it provides a set of rules that are likely to increase the B2G data access and reuse practices across the EU, benefitting both data holders and data re-users and the society in general.

3.4.1.2 Measures supporting citizen empowerment

The following table includes short description of how the three policy options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 86 – Summary comparison table between low/high intensity policy options for measures supporting citizen empowerment ('human-centric data economy')

	Non-regulatory intervention (PO1)	**Regulatory intervention with low intensity (PO2)**	**Regulatory intervention with high intensity (PO3)**
Efficiency	This option would bring medium costs to data holders and re-users, due to voluntary nature of the option and that the ones joining the scheme will be those with better cost-benefit ratios. The option will bring in significant benefits for the participants to the voluntary scheme, as the ones with better cost-benefits ratios are more likely to join it. Moreover, the reciprocity clause will enhance the benefits for both producers and re-users.	This option would bring substantial costs to data holders and re-users. The costs are expected to be high, especially for smart home appliances that would have to set up from scratch advanced data management solutions for a very embryonic market. The option will bring in high benefits for fitness and wearables re-users, but limited for smart home appliances re-users due to its limited maturity and demand. The reciprocity clause will enhance the benefits for producers in addition to the possibility for premium data offering.	This option would bring substantial costs to data holders and re-users. The costs are expected to be high, especially for smart home appliances that would have to set up from scratch advanced data management solutions for a very embryonic market. This option will translate in high benefits for re-users, especially in fitness and wearables. Some benefits for smart home appliances data re-user will also be there, notably for repair shop. Producers of smart home appliances would have almost no benefits but bear high costs.
Effectiveness	The option will a rather limited overall impact in achieving the general and specific objectives. It will contribute to the development of the overall system, it will address only partially the market	The option will have a rather limited impact, achieving only partially some of the specific and general objectives. It will help improve the development of the overall system and address partially the fragmentation level within	The option will address all of the operational objectives, and will provide the necessary conditions for achieving the specific and general objectives. The measures proposed will help improve the overall system and address the

	fragmentation, due to its sector-based nature. Also, being a voluntary scheme its market adoption level by the companies remains unpredictable.	certain sectors (e.g. the home appliances one). It will enable the development of models and initiative for data portability based on standards, allowing new players to join the market and contribute to increase customers choice. Also, the development of new products and services, and the improvement of innovation remain rather limited.	fragmentation level within certain sectors, and establish the premises for a wider inter-sectoral data integration.
Coherence	This option does not require any specific legal regulation and it will not be incoherent with the other policy options.	Coherent with EU law as it is similar with the regulation adopted in car, finance and energy sector. However, some concerns of potential conflict with provisions on data sharing from smart appliances planned for the implementation of the Energy Efficiency might arise.	Coherent with EU law as it is similar with the regulation adopted in car, finance and energy sector. However, some concerns of potential conflict with provisions on data sharing from smart appliances planned for the implementation of the Energy Efficiency might arise.
Legal/political feasibility	-	This option appears to be feasible.	This option appears to be feasible.
Proportionality	-	This option appears to be proportionate.	This option appears to be proportionate.

PO1: Voluntary white button scheme plus reciprocity

This option will translate in relatively high benefits for participants as the ones using the voluntary scheme will be those with better cost-benefit ratio. Costs will also be moderated by the voluntary nature of the initiative. Overall, this option will have moderate costs, with the benefits concentrated in the companies more likely to gain. The companies where the cost-benefit ratio remain relatively low (or even negative) will have less reasons to join the scheme than the ones with high cost-benefit ratio, which are expecting high benefits with low costs.

Data holders (OEM)	Fitness	Costs for setting up and running standardised API will not be too high since most already have API and this voluntary scheme will only be used by those with better cost-benefit ratio. Benefits are high because of the reciprocity clause; they are similar to those of the re-user.

Data holders (OEM)	Smart home appliances	Costs for setting up and running standardized API are high since there is limited standardisation and data portability at the moment. Yet, they will be moderated by the fact that this voluntary scheme will mostly appeal to those with better cost-benefit ratio. Benefits are high because of the reciprocity clause, they are similar to those of the re-users, although many re-users will be small business and thereby exempt from the costs burden. However, the level of benefits will be lower compared to the fitness sector, due to a lower level of maturity of the market.
Data re-users (aftermarket)	Fitness	Costs and benefit similar to data holders
Data re-users (aftermarket)	Smart home appliances	Costs and benefit similar to data holders

PO2: Compulsory but limited data sharing with reciprocity and possibility for charging for data

This option will translate in high benefits for fitness and wearables re-users, but limited for white appliances re-users due to limited demand and maturity of the market. Producers could benefit from additional revenue from both reciprocity and possibility for premium data offering.

The costs are expected to be high, especially for producers of smart home appliances that would have to set up from scratch advanced data management solutions for a very embryonic market.

Data holders (OEM)	Fitness	Costs for setting up and running standardised API could end up quite high even if most already have APIs especially due to a worse cost-benefit ratio because compulsory nature of the option and it will no longer be limited to only those likely to gain more. Benefits are high because of the reciprocity clause; they are similar to those of the re-users.
Data holders (OEM)	Smart home appliances	Costs for setting up and running standardised API are very high since there is limited standardisation and data portability at the moment. Additionally, the costs will no longer be moderated by the voluntary nature of the scheme as for the previous option when it applied only to those with better cost-benefit ratio. Benefits are quite high because of the reciprocity clause, and they are similar to those of the re-users. However, they remain lower compared to those for fitness as the market is less mature.

Data re-users (aftermarket)	Fitness	Costs and benefit are similar to those for data holders. Innovation from large scale machine learning remains rather limited.
Data re-users (aftermarket)	Smart home appliances	Costs and benefit are similar to those for data holders. Innovation from large scale machine learning remains rather limited.

PO3: Compulsory data sharing without reciprocity and no possibility for charging for data

This option will translate in high benefits for re-users, especially in fitness and wearables. Some benefits for smart home appliances data re-user will also be there, notably for repair shops. Producers of smart home appliances would have almost no benefits but bear high costs.

The costs are expected to be high, especially for producers of smart home appliances that would have to set up from scratch advanced data management solutions for a very embryonic market.

Data holders (OEM)	Fitness	Costs for setting up and running standardised API could end up quite high even if most already have APIs, especially due to a worse cost-benefit ratio as the option applies to everyone and it will no longer be limited to only those likely to gain more. Moreover, the wider category of data concerned, and the real-time provisions might result in additional costs for companies. Benefits are lower as the reciprocity clause is no longer in place and they are similar to those of the re-users.
Data holders (OEM)	Smart home appliances	Costs for setting up and running standardised API are very high since there is limited standardisation and data portability at the moment. Moreover, the costs will no longer be moderated by the voluntary character of the scheme as everyone will have to comply with the requirements not only those with better cost-benefits ratio. Benefits are low because the reciprocity clause no longer applies, and they are similar to those of the re-users.
Data re-users (aftermarket)	Fitness	Costs remain relatively low, while benefits are high from the possibility of data reuse, the possibility for cross-selling and added value services. Innovation from large scale machine learning improves significantly.
Data re-users (aftermarket)	Smart home appliances	Low cost, medium benefits from possibility of data reuse since there is limited evidence of value of these data. Most of the value will lie in repair shops. Innovation from large scale machine learning remains rather limited as the market is less mature and there are fewer opportunities available.

3.4.1.3 Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

The analysis of the different policy options allows to draw a preliminary comparison of the different options, based on the current available evidence.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 87 – Summary comparison table between low/high intensity policy options for measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	This option might be followed by different types of implementation costs for IoT solution providers, as well as by several benefits for different market players (IoT solution users, third party re-users). Given that the benefits apply to a bigger and broader range of stakeholders compared to the costs, the benefits are likely to outweigh the costs, presenting therefore the best balance between costs and benefits, compared to the other options.	Similarly to PO2 this option is associated to several types and ranges of costs and benefits for the different market stakeholders. However, the high intensity of regulation is expected to significantly increase the implementation costs, which are then likely to outweigh the benefits
Effectiveness	Significant contribution in achieving policy objectives in terms of providing legal certainty and clarity on access and usage rights over co-generated data; however the exact effectiveness level depends on the content of the legislation in terms of the extent that it contributes in creating and/or opening-up common standards.	Potentially higher effectiveness levels compared to PO2 in achieving policy objectives in terms of providing clarity on access and usage rights over co-generated data; however, similarly to PO2, the exact effectiveness level depends on the content of the legislation in terms of the extent that it contributes in creating and/or opening-up common standards.
Coherence	Coherent with EU laws and policies	Coherent with EU laws and policies, although relationship with other legal frameworks (GDPR, Database Directive and Trade Secrets Directive) should be addressed
Legal/political feasibility	This policy option appears to be feasible.	This policy option appears to be feasible.
Proportionality	This policy option appears to be proportionate as its (limited) intensity matches the identified problems and objectives of this study.	This policy option might be disproportionate with the problems and objectives of this study, given the evolving nature of the topic and the different maturity levels between various industry sectors with regard to co-generated B2B data sharing.

In terms of effectiveness, all policy options (apart from Policy Option O – baseline scenario) could contribute in achieving the policy objectives set by this domain, and therefore could be characterized as effective. The non-regulatory policy option (PO1) might be partially effective due to its non-binding nature and not significantly change the status quo. However, the active involvement of the industry stakeholders in this policy option could be a factor increasing its effectiveness in terms of creating common standards. The two regulatory policy options (PO2 and PO3) present higher levels of effectiveness, with PO3 being potentially the most effective one due to the high intensity of legislation in terms of providing legal certainty and clarity on access and usage rights over co-generated data. However, the exact effectiveness level of both of them depends on the content of the legislation with regard to the extent that it contributes in creating and/or opening-up common standards, apart from providing legal clarity.

In terms of efficiency, all policy options present different types of implementation costs for IoT solution providers, as well as several types of benefits for the various stakeholders of the value chain (mainly for IoT solution users and third-party re-users). In Policy Option 1, the exact estimation of costs and benefits is difficult due to the non-binding nature of the intervention. In the case of high-level of compliance by industry stakeholders, benefits would be expected to outweigh the costs. In Policy Option 2, taking into consideration that the benefits apply to a bigger and broader range of stakeholders[430] compared to the costs, the benefits would be expected to outweigh the costs, presenting, therefore, the best balance between costs and benefits, compared to the other options. Finally, in Policy Option 3, the high intensity of legislation is expected to significantly increase the implementation costs and burdens and decrease flexibility and innovation- mostly on the side of data holders-, without necessarily increasing the associated benefits accordingly.

All policy options are coherent with EU laws and policies, since they aim to support the objectives of the single market for data under the European Strategy for Data, in particular the following provisions: a) data can flow within the EU and across sectors; b) the rules for access and use of data are fair, practical and clear. Furthermore, they contribute to address identified issues linked to B2B data-sharing and usage rights on co-generated industrial data (IoT data created in industrial settings). However, policy option 3 in particular aims to regulate access and usage rights to data in a manner that could also concern personal data, raising the need to ensure compliance with the GDPR. Moreover, there is some ambiguity in terms of the applicability of the Database Directive and the Trade Secrets Directive to co-generated data; since policy option 3 aims to clarify access and usage rights, coherence with these frameworks is more complex.

With respect to legal and political feasibility, all policy options appear viable. Policy option 3 requires a stronger market intervention by harmonising specific usage rights on co-generated data and promoting B2B data sharing of co-generated data under fair conditions, and as a result may be more politically challenging to adopt than policy option 2. The high-level focus on fairness and transparency however implies that both options should be both legally and politically feasible.

In terms of proportionality, Policy Option 1 and Policy Option 2 appear to be proportionate to the problem assessment, as the non-binding nature of Policy Option 1 and the (limited) intensity of Policy Option 2 match the volume of the identified problems and objectives of this study. Policy Option 3 could be characterized as disproportionate to the problem assessment, given the evolving nature of the topic and the different maturity levels between various industry sectors with regard to co-generated B2B data sharing.

Based on this multi-criteria analysis, assessing the effectiveness, efficiency, coherence, feasibility and proportionality of the different policy options, the low-intensity regulatory intervention (PO2) aiming to adopt

[430] Including data co-generators, data re-users and data holders, affecting in total more than 4 billion European enterprises.

a legal instrument aiming to bring legal certainty and promote contractual fairness for accessing and (co-)using IoT co-generated data appears to be the preferred policy option.

3.4.1.4 Measures supporting companies in cases of conflict of laws at international level

As the analysis indicated, no single policy option is entirely effective in eliminating the problem. None the less, a comparative assessment of the policy options is possible, taking into account the available evidence and indicators of likely impacts.

The non-regulatory option (policy option 1) is efficient, but also the least effective: while it can improve transparency in relation to potential conflicts of laws and awareness of mitigating measures, there is no assurance that any operational improvement would occur that exceeds the natural increase of security measures in the status quo scenario where no policy measure is taken. In other words: while security maturity can reasonably be expected to occur in the absence of any policy intervention, and such increased security will naturally also result in some degree of protection against conflicts of laws, the process is slow and unlikely to yield significant and effective results. For that reason, non-regulatory intervention (policy option 1) is not seen as an appropriate measure that would significantly contribute to achieving greater certainty for cloud users and cloud providers, nor would it have any beneficial impact on data sovereignty.

The two other policy options are both regulatory in nature, with policy option 2 being a low-intensity intervention that principally focuses on improving transparency by compelling cloud providers to invest in proactive and reactive notifications of extraterritorial data access requests; and policy option 3 being a high-intensity intervention that additionally requires investments from cloud providers to mitigate the likelihood of successful extraterritorial data access attempts that do not satisfy European legal requirements.

The following table includes short description of how the low/high intensity options compare in terms of efficiency, effectiveness, coherence, legal/political feasibility and proportionality.

Table 88 – Summary comparison table between low/high intensity policy options for measures supporting companies in cases of conflict of laws at international level

	Regulatory intervention with low intensity (PO2)	Regulatory intervention with high intensity (PO3)
Efficiency	Costs are relatively limited for both cloud providers (who largely already commit to more limited notification duties) and for the EU (who principally needs to invest in a notifications platform).	Costs are significant for most cloud providers, due to the required investments in technical, legal and organizational security measures. Some of these would naturally occur as the state-of-the-art progresses, but additional required investments in compliance would none the less be substantial.
Effectiveness	The effectiveness is limited, since the principal impact is on transparency and awareness raising, not on prevention of cross border conflicts of law.	Effectiveness is higher than under policy option 2. While no perfectly effective silver bullet measures are available, some cross-border conflicts of law are likely to be prevented.
Coherence	Entirely coherent with existing EU laws and policies.	Entirely coherent with existing EU laws and policies. Analogous

		development with the elevated bar of security and trustworthiness set under the GDPR, notably as a result of the Schrems II ruling.
Legal/political feasibility	This option appears to be feasible.	This option appears to be feasible. Challenges may arise in relation to limitations on free trade and GATT compliance, but based on international trends in data sovereignty discussions and strategic autonomy in critical sectors, the option appears feasible.
Proportionality	High. Intervention is limited and certainly proportionate to the problem at hand.	High. While intervention is stronger, the increased effectiveness of the measure ensures its proportionality.

The analysis strongly suggests that, while policy option 2 carries a significantly lower cost both for targeted companies and for the EU, that policy option also has only a limited effectiveness. This is due to the option's focus on transparency and best practices: it increases awareness of the potential risks of international conflicts of law and of best practices to mitigate and manage these risks, but there is no assurance whatsoever of any operational changes, or of avoiding any actual incidents. Policy option 3 on the other hand aims to elevate the bar for technical, legal and organisational security measures to be implemented by cloud providers, requiring them to invest in solutions that allow them to identify and mitigate these risks. Policy option 3 therefore implies a much more significant cost of business for cloud providers, but is also significantly more likely to actually eliminate at least some of the risks.

Gains of both policy options are difficult to measure, since the main anticipated benefit is increased security and sovereignty over data, which is an intangible and non-quantifiable benefit. On that point too, however, policy option 3 is certain to be more effective than policy option 2. Policy option 3 also has indirect gains, since the measures taken would generally elevate security levels in cloud services, and therefore generally improve the quality of services and reduce costs linked to security breaches, also outside of the context of the specific problem of international conflicts of law. Additionally, the analysis shows that some of the likely measures to be taken already see increasing market adoption, so that some (but likely not all) of the costs would eventually have been incurred by cloud providers at some point, as the state of the art progresses. Policy option 3 mainly imposes accelerated investment. It should also be recognised however that policy option 3 is not optimally effective either: as the overview above showed, the effectiveness of measures appears to be inversely proportionate to their general applicability and feasibility. In simpler terms, legal measures are nearly universally feasible, but have limited effectiveness in stopping data access requests that do not comply with EU requirements. Technical and operational measures generally are more effective (and occasionally perfectly effective) in solving the problem, but are often inapplicable to the cloud use cases that are most widely taken up in the market.

In terms of coherence, both policy options 2 and 3 are entirely coherent with EU legislation and policy. The increased focus on transparency (under policy option 2) and on elevating the bar for security (policy option 3) are entirely in line with other EU policy areas. Notably, policy option 3 is entirely in line with EU policies and jurisprudence regarding transfers of personal data to third countries under the GDPR (and notably the guidance on complementary measures to safeguard such transfers), and the focus on enhanced and effective security is perfectly in line with the approach of the Cybersecurity Act, and the ongoing work surrounding cloud security certification schemes undertaken by ENISA, as well as broader cloud initiatives such as GAIA-X, which also include cloud security requirements.

In terms of legal and political feasibility, policy option 2 appears to be entirely unproblematic. Policy option 3 may face legal doubts in terms of free trade impacts and GATT compliance, but this issue is mitigated to a significant extent through the fact that it is essentially an extension of existing EU policies stressing the importance of MLATs (and more generally international cooperation) in ensuring the lawfulness of data exchanges. Furthermore, under policy option 3, the EU aims to apply an equal bar for EU data sovereignty towards all cloud providers, irrespective of their place of establishment or the location of their infrastructure, which ultimately could facilitate and enable international trade. Politically, policy option 3 seems viable in particular due to its recognition of the increasing political importance of data sovereignty and strategic autonomy in critical sectors, of which the data economy is a particularly relevant example. Moreover, as the analysis shows, the EU would not be unique in safeguarding its data assets by setting a high bar of protection against third country interference. On the basis of this consideration, policy options 2 and 3 can also be assessed as being proportionate to the desired goal.

The preferred policy option seems to be Policy Option 3, aiming to effect operational changes by imposing an elevated standard of security on cloud providers, and by requiring greater legal diligence from such providers prior to responding to any third country data access requests. Even though the level of costs is likely to be significantly higher for policy option 3, the effectiveness of the option is also significantly higher, and the investments to be made have considerable ancillary benefits in terms of improving security of data in the cloud and therefore strengthening the trust that European users can place in cloud providers, irrespective of their place of establishment or the location of their infrastructure.

3.4.2 Comparison of the policy options

In relation to Step 2, the following table provides an **outranking matrix** in which all the weights indicated in the table under step 1 are totalled for the criteria in relation to which a policy option is favoured over another policy option (abbreviated e.g. as "P1/P2") as indicated by the weighted performance of each criterion.

This means that the outranking matrix provides an overview of the overall scores of the policy options compared to each other (i.e. the differences between them).

Table 89 – Outranking matrix

Outranking matrix			
Business-to-Government data sharing for the public interest	PO 1	PO 2	PO 3
PO 1	**0**	**0,15**	**0,15**
PO 2	**0,85**	**0**	**0,45**
PO 3	**0,85**	**0,3**	**0**
Measures supporting citizen empowerment (‘human-centric data economy’) - Smart home appliances	PO 1	PO 2	PO 3
PO 1	**0**	**0,85**	**0,55**
PO 2	**0,15**	**0**	**0,45**
PO 3	**0,45**	**0,3**	**0**
Measures supporting citizen empowerment (‘human-centric data economy’) - Fitness trackers	PO 1	PO 2	PO 3
PO 1	**0**	**0,25**	**0,55**
PO 2	**0,75**	**0**	**0,45**
PO 3	**0,45**	**0,3**	**0**
Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing	PO 1	PO 2	PO 3
PO 1	**0**	**0**	**0,55**
PO 2	**0,6**	**0**	**0,55**
PO 3	**0,3**	**0,3**	**0**
Measures supporting companies in cases of conflict of laws at international level	PO 1	PO 2	PO 3
PO 1	**0**	**0,15**	**0,45**
PO 2	**0,55**	**0**	**0,45**
PO 3	**0,55**	**0,3**	**0**

Naturally, the elements/combinations in the diagonal of each square matrix received a score of 0 as it does not make sense to compare these (each option is compared with itself). In essence, the table shows that the impacts of the policy options outrank those of the baseline scenario and that policy options with a higher score outrank those with a lower score.

The differences between the overall rankings of each policy option between each other as presented above are derived from the sum of the individual scores per policy option and assessment criterion in the analytical grid.

The table below present the six different combination of policy options for the four areas under investigation (with an individual assessment of the each of the two sectors under Measures supporting citizen empowerment (‘human-centric data economy’)).

Table 90 – Policy ranking permutation

Policy ranking permutation	Policy pairings	Coefficients of policy pairings	Final score
Business-to-Government data sharing for the public interest			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,15 + 0,15 + 0,45	0,75
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,15 + 0,15 + 0,3	0,6
PO2/PO1/PO3	PO2/PO1 + PO1/PO3 + PO2/PO3	0,85 + 0,45 + 0,15	1,45
PO2/PO3/PO1	*PO2/PO3 + PO3/PO1 + PO3/PO2*	*0,45 + 0,85 + 0,85*	*2,15*
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,85 + 0,3 + 0,15	1,3
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,85 + 0,85	2
Measures supporting citizen empowerment ('human-centric data economy') - Smart home appliances			
PO1/PO2/PO3	*PO1/PO2 + PO1/PO3 + PO2/PO3*	*0,85 + 0,55 + 0,45*	*1,85*
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,55 + 0,85 + 0,3	1,7
PO2/PO1/PO3	PO2/PO1 + PO1/PO3 + PO2/PO3	0,15 + 0,45 + 0,55	1,15
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,45 + 0,15 + 0,45	1,05
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,45 + 0,3 + 0,85	1,6
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,45 + 0,15	0,9
Measures supporting citizen empowerment ('human-centric data economy') - Fitness trackers			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,25 + 0,55 + 0,45	1,25
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,55 + 0,25 + 0,3	1,1
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,75 + 0,45 + 0,55*	*1,75*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,45 + 0,75 + 0,45	1,65
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,45 + 0,3 + 0,25	1
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,45 + 0,75	1,5
Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0 + 0,55 + 0,55	1,1
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,55 + 0 + 0,3	0,85
PO2/PO1/PO3	*PO2/PO1 + PO1/PO3 + PO2/PO3*	*0,6 + 0,55 + 0,55*	*1,7*
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,55 + 0,6 + 0,3	1,45
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,3 + 0,3 + 0	0,6
PO3/PO2/PO1	PO3/PO2 + PO3/PO1 + PO2/PO1	0,3 + 0,3 + 0,6	1,2
Measures supporting companies in cases of conflict of laws at international level			
PO1/PO2/PO3	PO1/PO2 + PO1/PO3 + PO2/PO3	0,15 + 0,45 + 0,45	1,05
PO1/PO3/PO2	PO1/PO3 + PO3/PO2 + PO1/PO2	0,45 + 0,15 + 0,55	1,15
PO2/PO1/PO3	PO2/PO1 + PO1/PO3 + PO2/PO3	0,55 + 0,45 + 0,45	1,45
PO2/PO3/PO1	PO2/PO3 + PO3/PO1 + PO3/PO2	0,45 + 0,55 + 0,55	1,55
PO3/PO1/PO2	PO3/PO1 + PO1/PO2 + PO3/PO2	0,55 + 0,55 + 0,15	1,25
PO3/PO2/PO1	*PO3/PO2 + PO3/PO1 + PO2/PO1*	*0,55 + 0,55 + 0,55*	*1,65*

This means the following:

- For Business-to-Government (B2G) data sharing for the public interest, **policy option PO2 – Low-intensity regulatory intervention** is the preferred option as it provides the most combination of effectiveness, efficiency and coherence.
- For Measures supporting citizen empowerment ('human-centric data economy') – Smart home appliances, **policy option PO1 – Voluntary white button scheme plus reciprocity** is the preferred option;
- For Measures supporting citizen empowerment ('human-centric data economy') – Fitness trackers, **policy option PO2 – Compulsory but limited data sharing with reciprocity and possibility for charging for data** is the preferred option;

- For Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing., **policy option PO2 – Low intensity regulatory option focusing on the adoption of a legal instrument aiming to bring legal certainty and promote contractual fairness for accessing and (co-)using IoT co-generated data** is the preferred option;
- For Measures supporting companies in cases of conflict of laws at international level, **policy option PO3 – High intensity regulatory option** focusing on transparency is the preferred option.

3.5 Assessment of macro-economic impacts

This section presents the expected macro-economic impacts of the policy options on the overall economy and society compared to the baseline scenario.

3.5.1 Methodological approach

This section provides a brief explanation about the methodological approach for the macroeconomic analysis.

For the analysis of the economic impact a bottom-up analysis is conducted. The bottom-up approach is based on the micro-analysis of estimated impacts conducted for each of the domains under consideration. Within the CBA, certain benefits (e.g. additional revenues, profits, productivity gains) and costs (e.g. implementation, infrastructure, compliance costs) are assessed. As far as possible, the impact on GDP is estimated based on the CBA results and/or case studies. The results and estimations of the micro-analyses are extrapolated and scaled in this regard. The bottom-up approach is described in more detail in the following sections.

3.5.1.1 Calculation of the baseline

The baseline (scenario) reflects the no policy change option for all the domains for which problems could be identified. Considering that with regard to Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, a wide range of economic sectors is potentially affected, the overall/total GDP for the EU27 is selected as the baseline and the projections.[431] In the year 2020, the outbreak of Covid-19 massively affected the European economy. Expected figures for 2020 have been corrected to take into account the impact of this crisis.

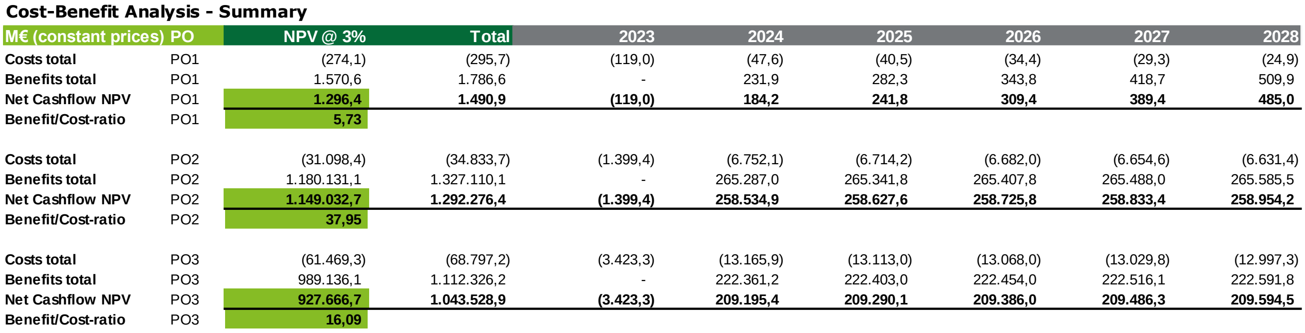
3.5.1.2 Bottom-up analysis

A bottom-up assessment of economic impacts has been performed based on the results of the cost-benefit analysis for the Measures supporting citizen empowerment ('human-centric data economy') and Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, whereby the first one consists of two parts; a) for the 'Smart home appliances' and b) for the 'Fitness trackers'.

The cost-benefit calculations and assumptions used are explained in section 3.5.3., the details and results are presented in the tables in Annex II. As a starting point for the bottom-up analysis, the net benefits for the domains under consideration are summed up per policy option (e.g. the net benefits for Measures supporting citizen empowerment ('human-centric data economy') a) and b), and Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, under policy option 1 sum up to 184 million EUR in 2024). The results (sum of the net benefits under each policy option) are presented in the table below.

[431] The baseline for the GDP has been projected based on the GDP forecast used in the European Data Market Monitoring Tool, see: http://datalandscape.eu/european-data-market-monitoring-tool-2018 and beyond 2025 based on GDP growth rate forecasts of the OECD (1.5%-1.6% p.a.).

Table 91 – Cost-Benefit Analysis summary table



Cost-Benefit Analysis - Summary

M€ (constant prices)	PO	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Costs total	PO1	(274,1)	(295,7)	(119,0)	(47,6)	(40,5)	(34,4)	(29,3)	(24,9)
Benefits total	PO1	1.570,6	1.786,6	-	231,9	282,3	343,8	418,7	509,9
Net Cashflow NPV	PO1	**1.296,4**	**1.490,9**	**(119,0)**	**184,2**	**241,8**	**309,4**	**389,4**	**485,0**
Benefit/Cost-ratio	PO1	**5,73**							
Costs total	PO2	(31.098,4)	(34.833,7)	(1.399,4)	(6.752,1)	(6.714,2)	(6.682,0)	(6.654,6)	(6.631,4)
Benefits total	PO2	1.180.131,1	1.327.110,1	-	265.287,0	265.341,8	265.407,8	265.488,0	265.585,5
Net Cashflow NPV	PO2	**1.149.032,7**	**1.292.276,4**	**(1.399,4)**	**258.534,9**	**258.627,6**	**258.725,8**	**258.833,4**	**258.954,2**
Benefit/Cost-ratio	PO2	**37,95**							
Costs total	PO3	(61.469,3)	(68.797,2)	(3.423,3)	(13.165,9)	(13.113,0)	(13.068,0)	(13.029,8)	(12.997,3)
Benefits total	PO3	989.136,1	1.112.326,2	-	222.361,2	222.403,0	222.454,0	222.516,1	222.591,8
Net Cashflow NPV	PO3	**927.666,7**	**1.043.528,9**	**(3.423,3)**	**209.195,4**	**209.290,1**	**209.386,0**	**209.486,3**	**209.594,5**
Benefit/Cost-ratio	PO3	**16,09**							

Within the cost-benefit analysis, substantial benefits in terms of producer surplus have been estimated already. Those benefits refer to the producer surplus e.g. for data holders or data re-users/co-producers. The economic value/benefit was calculated either estimating profits or revenues minus costs. Thus, these benefits already represent economic contributions to GDP/GVA.[432] In a next step, a minor correction is made regarding benefits, that do not represent economic impacts which contribute to GDP/GVA.[433] Those impacts have been excluded/subtracted from the net benefits presented in the table above. Consequently, the bottom-up results presented in the following table differ slightly from the aggregated results of the cost-benefit calculations presented in the table before.

With regard to the Measures supporting citizen empowerment ('human-centric data economy') market growth rates have been considered for the market segments in focus. Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, on the other hand is expected to have an impact on a broad range of sectors of the economy in general. Consequently, for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing it is expected, that the measures will have an immediate impact on the economic undertakings after becoming effective. In this regard, in line with the CBA, an adoption in 2023 is assumed and impacts are considered from 2024 onwards.[434]

[432] Contrarily, benefits e.g. in terms of consumer surplus or societal benefits have not been included in the estimation of economic impacts.

[433] This refers e.g. to estimated consumer benefits under Measures supporting citizen empowerment ('human-centric data economy') a); however, the difference is insignificant.

[434] Negative numbers are shown in parentheses/brackets in the following tables.

Table 92 – Direct Economic impacts based on the results of the CBA per policy option

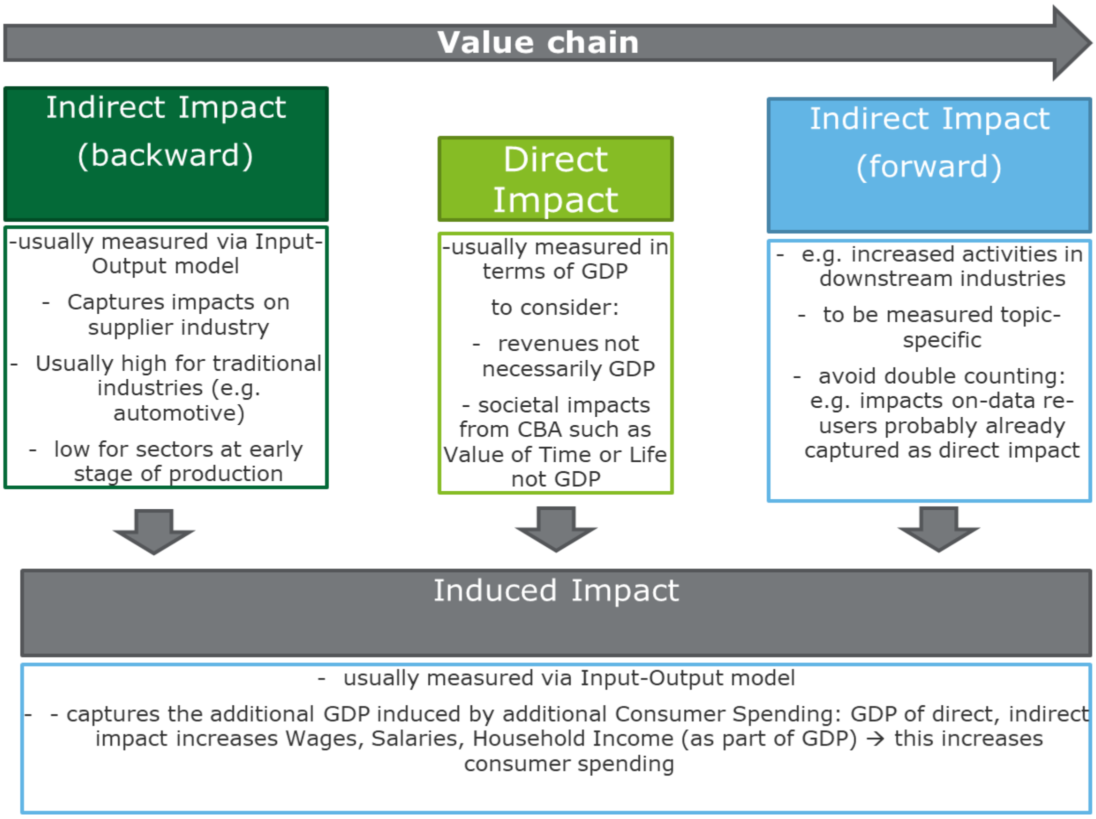
Economic Impact │ direct impact CBA

M€	2024	2025	2026	2027	2028
Policy impact - bottom up (based on CBA results)					
Policy Option 1 - direct					
2.2a	48.2	66.7	88.5	114.3	145.3
2.2b	135.9	174.9	220.7	274.9	339.5
2.3	-	-	-	-	-
Policy Option 2 - direct					
2.2a	(158.6)	(108.3)	(60.0)	(11.9)	37.8
2.2b	158.0	200.3	250.3	309.7	380.8
2.3	258 533.6	258 533.6	258 533.6	258 533.6	258 533.6
Policy Option 3 - direct					
2.2a	(272.7)	(209.1)	(149.7)	(92.8)	(36.3)
2.2b	111.9	142.9	179.5	222.8	274.6
2.3	209 354.3	209 354.3	209 354.3	209 354.3	209 354.3
Policy Option 1	184	242	309	389	485
Policy Option 2	258 533	258 626	258 724	258 831	258 952
Policy Option 3	209 193	209 288	209 384	209 484	209 593

In order to fully reflect on the reality of the impact, indirect impacts have been added to the estimates based on the CBA results. As the estimated effects already refer largely to producers of goods and services in final demand, only backward indirect effects are added in order to avoid an overestimation of impacts. Those indirect backward impacts refer to upstream effects at suppliers of the industries in focus in the CBA. A coefficient of 0.06 has been used, based on the estimations/results of the European Data Monitoring Tool for indirect impacts in the data industry/economy.[435]

[435] The European Data Monitoring implicitly includes several types of multipliers for the data economy/industry, including indirect and induced impacts, which estimate impacts on the supplier industries and the overall economy generated through additional income and consumption (both could be classically estimated using e.g. Input-Output models), as well as indirect forward impacts, which estimate the effects downstream in the economy. To stay conservative, we only included backward multipliers (for the upstream effects at suppliers of the industries in focus in the CBA). The European Data Monitoring Tool in this regard estimates coefficients of around 0.06 for the data economy/industry, which is rather insignificant.

Figure 30 – Overview of direct, indirect and induced impacts.


Value chain
Indirect Impact (backward)
-usually measured via Input-Output model
- Captures impacts on supplier industry
- Usually high for traditional industries (e.g. automotive)
- low for sectors at early stage of production
Direct Impact
-usually measured in terms of GDP
to consider:
- revenues not necessarily GDP
- societal impacts from CBA such as Value of Time or Life not GDP
Indirect Impact (forward)
- e.g. increased activities in downstream industries
- to be measured topic-specific
- avoid double counting: e.g. impacts on-data re-users probably already captured as direct impact
Induced Impact
- usually measured via Input-Output model
- - captures the additional GDP induced by additional Consumer Spending: GDP of direct, indirect impact increases Wages, Salaries, Household Income (as part of GDP) → this increases consumer spending

3.5.2 Macroeconomic impacts of the policy options

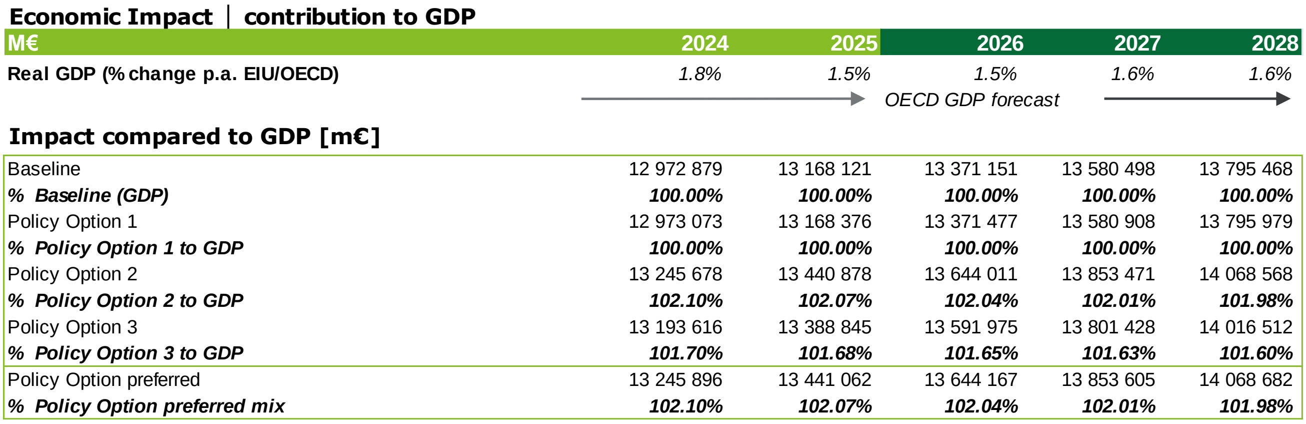
The Impact Assessment support study took as the baseline the total value of the GDP for the EU27 of around 11.5 trillion EUR in 2020. These numbers take into account a correction linked to Covid-19 impact on the overall EU economy.

The baseline scenario foresees an autonomous growth to around 13.8 trillion EUR (+20%) in 2028.

For 2028, our analysis indicates a potential annual addition of 273 billion EUR to GDP if the policy option 2 intervention was introduced. If policy option 3 is introduced, a potential annual addition of 221 billion EUR to GDP is estimated.

In 2028, the value of the GDP could increase from 13.8 trillion EUR to around 14.1 trillion EUR if the policy option 2 was introduced (plus 1.98% to the GDP). In 2028, the value of the GDP could increase from 13.8 trillion EUR to 14.0 trillion EUR if policy option 3 was introduced (plus 1.60% to the GDP). In case that a mix of preferred policy options according to the results of the Multi-Criteria-Analysis (section 3.4.1.) is implemented - i.e. policy option 1 for Measures supporting citizen empowerment ('human-centric data economy') a) and policy option 2 for Measures supporting citizen empowerment ('human-centric data economy') b), and Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing- the GDP could increase by 1.98% to 14.1 trillion EUR in 2028 compared to the baseline.

Figure 31 – Results of the bottom-up macroeconomic impact calculations



Economic Impact | contribution to GDP

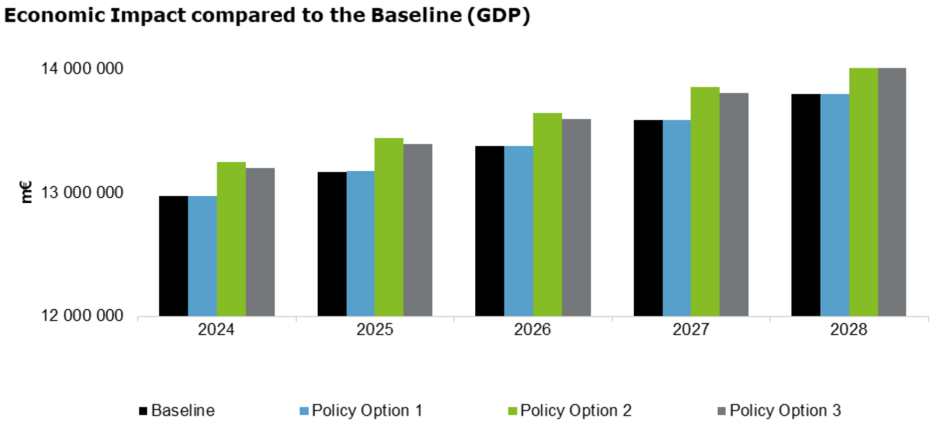
M€	2024	2025	2026	2027	2028
Real GDP (% change p.a. EIU/OECD)	*1.8%*	*1.5%*	*1.5%*	*1.6%*	*1.6%*
		→	*OECD GDP forecast*	→	

Impact compared to GDP [m€]

	2024	2025	2026	2027	2028
Baseline	12 972 879	13 168 121	13 371 151	13 580 498	13 795 468
% Baseline (GDP)	***100.00%***	***100.00%***	***100.00%***	***100.00%***	***100.00%***
Policy Option 1	12 973 073	13 168 376	13 371 477	13 580 908	13 795 979
% Policy Option 1 to GDP	***100.00%***	***100.00%***	***100.00%***	***100.00%***	***100.00%***
Policy Option 2	13 245 678	13 440 878	13 644 011	13 853 471	14 068 568
% Policy Option 2 to GDP	***102.10%***	***102.07%***	***102.04%***	***102.01%***	***101.98%***
Policy Option 3	13 193 616	13 388 845	13 591 975	13 801 428	14 016 512
% Policy Option 3 to GDP	***101.70%***	***101.68%***	***101.65%***	***101.63%***	***101.60%***
Policy Option preferred	13 245 896	13 441 062	13 644 167	13 853 605	14 068 682
% Policy Option preferred mix	***102.10%***	***102.07%***	***102.04%***	***102.01%***	***101.98%***

However, the preferred/mixed option is heavily dominated by the impact of Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, therefore we do not consider this mixed option in the following sections.

Figure 32 – Impact of the Economic Value compared to GDP



The following figures provide an overview of the share of contribution of the policy options to the economic impacts.

For policy option 1, it is expected that only Measures supporting citizen empowerment ('human-centric data economy') will have an impact. The economic impact for policy option 2 and 3 is dominated by Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing, which is expected to have a significant impact. However, it should be noted, that Measures supporting citizen empowerment ('human-centric data economy') a) (Smart home appliances) is estimated to have an adverse (negative) impacts under policy option 2 and 3. Consequently, in the figures below there is no (zero) impact under policy option 1 with regard to Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing. For policy option 2 and policy option 3 the impacts with regard to domains on smart home appliance and fitness trackers are not visible, because the impacts are insignificant compared to Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.

Figure 33 – Economic impact Policy Option 1 by domain

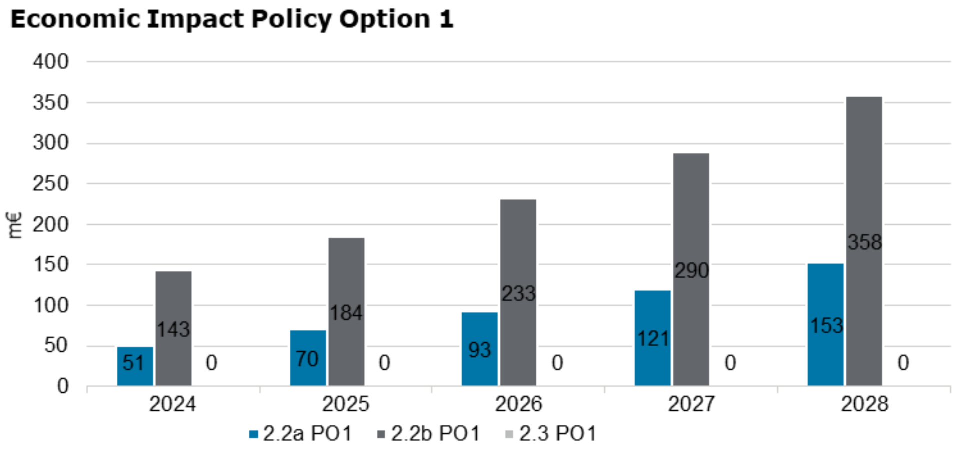


Figure 34 - Economic impact Policy Option 2 by domain

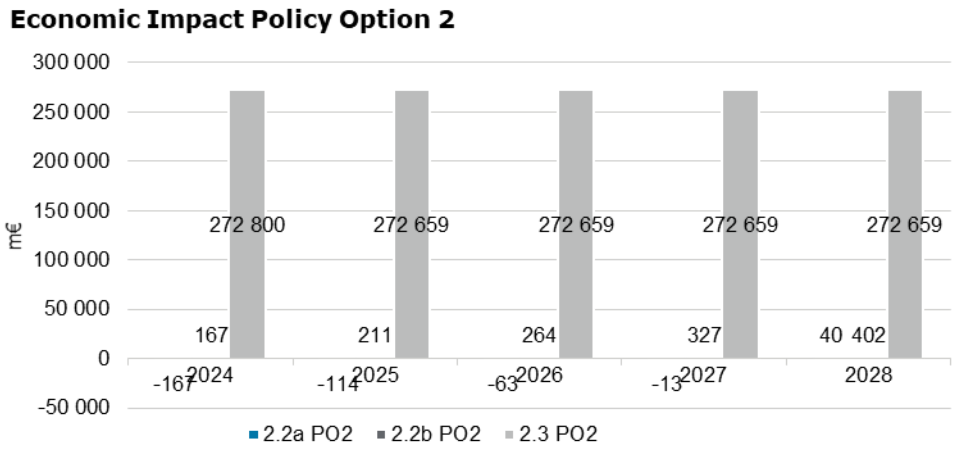
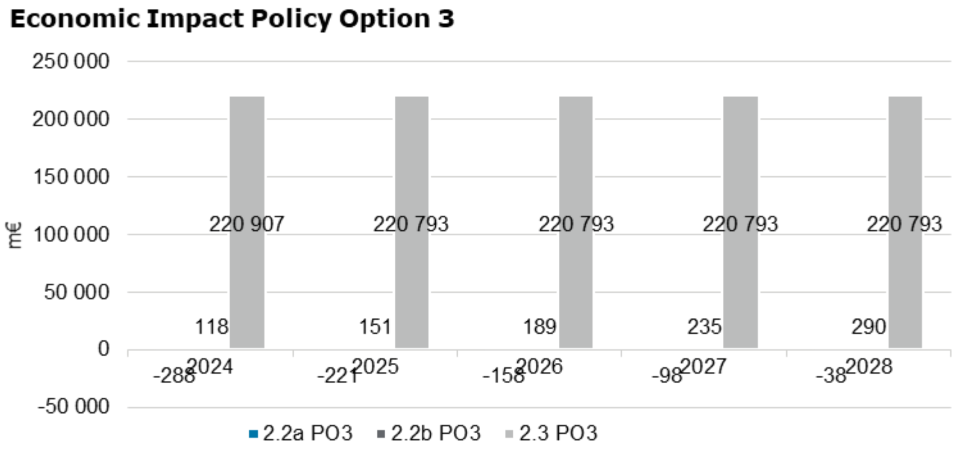


Figure 35 – Economic impact Policy Option 3 by domain



Our assessment is that overall positive impacts are expected at the macroeconomic level, by boosting the value of the total GDP for the EU27 from a projected autonomous growth from 11.5 trillion EUR in 2020 to

13.8 trillion EUR in 2028 to between 14.1 trillion EUR (policy option 2) and 14.0 trillion EUR (policy option 3) in 2028 (plus 1.98% to plus 1.60% of the GDP).

The policy option 2 creates the highest impact on the total economy. However, it should be noted, that Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing is expected to be the main driver and have a significant share in the overall impact (around 99-100%). Measures supporting citizen empowerment ('human-centric data economy') a) is expected to have a negative impact under policy option 2 and 3, whereas Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing is expected to have no impact under policy option 1.

3.5.2.1 Additional indicators

Based on the macroeconomic impacts we have estimated the impact of the policy options and policy packages on the following economic and socio-economic indicators:

- Employment (total number of additional persons employed, direct and indirect)
- Additional governmental revenues (total gross as % of GDP incl. SSC, taxes, subsidies, governmental revenues etc.)
- Additional investment activity

To estimate the impact on these indicators, coefficients in terms of GDP-ratios have been used based on official data provided by Eurostat. With regard to numbers of person's employment, the number of additional companies and additional investment activities the GDP-ratios of the ICT-sector have been applied. Governmental revenues were calculated based on the data on tax revenue and its relationship to gross domestic product (GDP) for the EU27 in general.

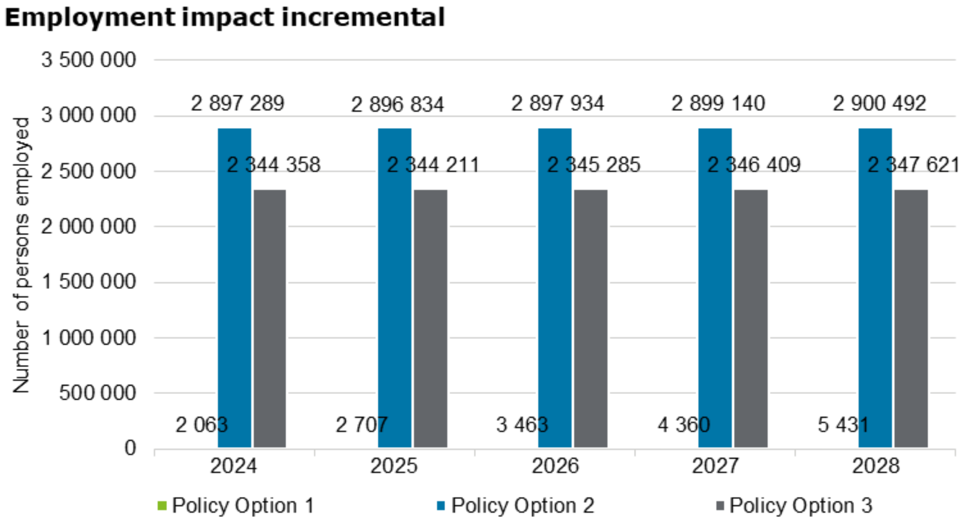
3.5.2.1.1 Employment

The first indicator, employment, indicates the total number of additional persons employed (directly and indirectly) in the case the respective Policy Option will be implemented. To calculate the total number of additionally employed people, the coefficient of employment as per mEUR gross value added (GVA) was determined. This coefficient was determined to be a weighted coefficient of the EU27 per mEUR GDP/GVA in the ICT services sector. Proceeding these calculations, a constant coefficient of 10.6 for the years 2024-2028 was applied.[436] The employment coefficient indicates the per-ratio increase in employment (number of persons employed) throughout the economy which result from an increase in GDP/GVA.

The following two figures provide a detailed overview of the employment impact incremental for the three policy options, based on the bottom-up calculation of the GDP impact. The impact is incremental compared to the baseline, meaning e.g. that in 2028 additional employment of up to around 2.9 million (PO2) can be achieved.

[436] The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the employment ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

Figure 36: Employment impact incremental compared to the baseline (bottom-up) in 2024-2028 for the different Policy Options

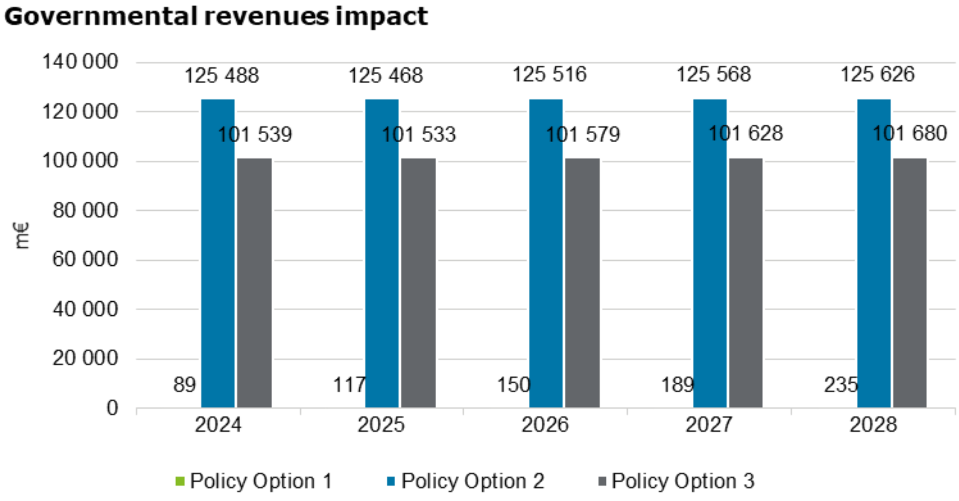


3.5.2.1.2 Governmental revenues

The second indicator to be included is the governmental revenues. According to the definition of Eurostat[437], the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Total taxes are composed of taxes on production and imports, current taxes on income and wealth and capital taxes. The net social contribution is composed of actual social contributions by employers and households and the imputed social contributions, households' social contribution supplements and social insurance scheme service charges. Other current revenues consist of the categories property income earned, other subsidies on production received and current transfers. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained. Following the calculations of Eurostat, this coefficient has the value of 46% of GDP for the EU27. It should be noted, that part of this is related to governmental output, including market output, output for own final use and payments for non-market output, which could be linked to increased economic activity, but does not represent governmental inflows from taxes, social security payments or similar revenues.

[437] Eurostat 2020, Statistics Explained, Glossary: government revenue and expenditure. https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government_revenue_and_expenditure

Figure 37 – Governmental revenue from 2024-2028 (bottom-up approach)

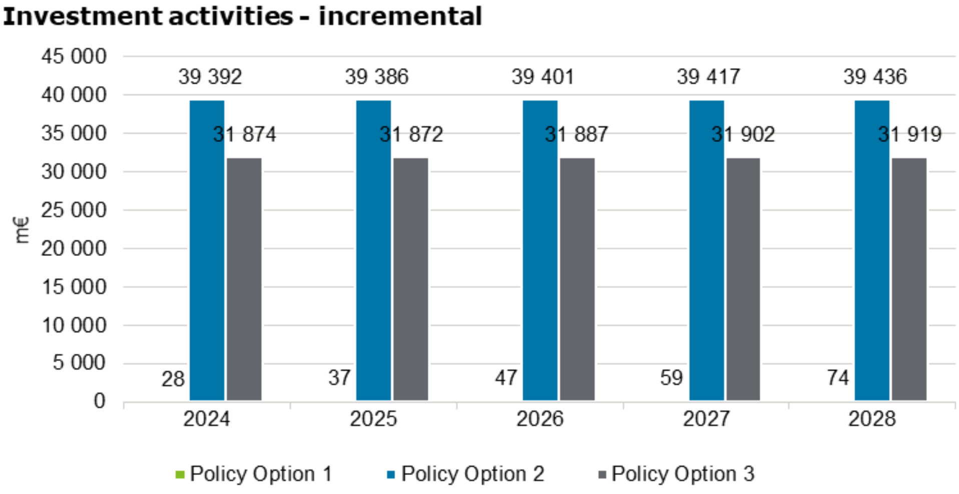


It must be noted, however, that this total governmental revenue includes – as defined in the European System of Accounts 2010 – also the market output, output for own final use and payments for non-market production. As this definition is a rather broad concept and as the macroeconomic effect of the introduction of the policy options depends on a lot yet unknown factors, market output, output for own final use and payments for non-market production cannot be predicted as precisely as the other variables of governmental revenues. Excluding the categories mentioned, the adjusted governmental revenues would lower to approximately 38% of GDP according to OECD estimates.[438]

3.5.2.1.3 Investment activity

As a further indicator to be added we suggest including investment activity. The investment rate is defined as the investment per value added at factor costs and is indicated as a percentage of the GDP of the EU27. The investment rate which was obtained by Eurostat[439] is at 14.4% of the GDP of the EU27 ICT-sector.

Figure 38 - Investment activities (bottom-up) for 2024-2028



[438] OECD, 2020, Comparative Statistics: Governmental Revenue. https://stats.oecd.org/Index.aspx?DataSetCode=REV

[439] Eurostat, 2020, Investment share of GDP. See: https://ec.europa.eu/eurostat/web/products-datasets/product?code=sdg_08_11

3.6 Conclusion

The chapter focused in particular on four key issues with regard to data sharing and re-use, namely:

- Aspects related to **Business to Government Data Sharing (B2G) for the public interest** (i.e. for the development of better policies and delivery of better public services).
- Possibilities for **empowering citizens** by facilitating their control of their data, in line with the General Data Protection Regulation[440] and establishing a human centric data economy.
- The question of **rights and control over co-generated data** (i.e. in the context of connected and Internet of Things devices) for enabling further business to business (B2B) data sharing.
- Aspects related to **conflict of laws at the international level** and possible obstacles for businesses subject to extra-territorial provisions and foreign jurisdictions.

For each of these key aspects, the current state of play in Europe is explored and what the impact of different policy options would be within the overall data ecosystem.

From a geographical perspective, the Part focused on **27 European Union Member States**, covering desk research, case studies, interviews and workshops. Examples and literature coming from third countries were also provided when relevant (i.e. experiences of B2G data sharing). From a stakeholder perspective, the study focused on the **relevant stakeholders in the data value chain** for each of the topics in scope, meaning on data holders, data intermediaries and data re-users (different stakeholders were emphasized based on the nature of the topics.

This study collected data from a range of sources, including desk research, stakeholder interviews, workshops and case studies. The data collection of the study faced challenges as both the public and private sectors are still relatively new to navigating the data economy and could only share insights regarding costs and benefits to a very limited extent. Therefore, while this study was able to collect qualitative feedback from the public and private sector on the different policy interventions discussed for each domain, it was more difficult to quantify their costs and benefits, e.g. because case numbers are still small or the data sharing practices are just emerging and stakeholders themselves do not yet know their scale and/or costs of making data available. This report should therefore be considered as an initial attempt at examining this topic and gathering the existing data on these subjects. The analysis above is based on the limited data available in an emerging data ecosystem, therefore it consists primarily of a qualitative overview of the costs and benefits for the different topics under scrutiny. The conclusions reached are based on independent judgement and specific to this study.

A macroeconomic analysis was performed for two domains to estimate the potential impact on the EU economy. In particular, in the context of co-generated data based on IoT solutions deployed across a wide variety of sectors the measures improving the sharing of data are expected to have a significant impact. The measures regarding empowering citizens to port the data they generate when using 'smart' devices (e.g. smart home devices, voice assistants, fitness trackers) also generate an economic impact although this is expected to be far more moderate overall. The low intensity policy measures (policy option 2) in each of these domains are estimated to generate an additional 273 billion EUR in GDP annually, adding 1.98% of GDP. The high intensity policy measures (policy option 3) for these domains are estimated to generate an additional 221 billion EUR in GDP annually, adding 1.60% of GDP. For the sharing of co-generated data based on IoT solutions, policy option 2 is preferred both in the multi-criteria assessment and from a macroeconomic impact perspective. The preferred intervention for data portability for citizens as consumers

[440] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

of smart devices is rather dependent on the sector concerned. For the smart home appliances sector, a non-regulatory, voluntary based, policy intervention with a reciprocity clause included is recommended, while for fitness trackers, a low intensity regulatory policy recommendation, with reciprocity clause included, is preferred. This is also reflected in the expected macroeconomic impact, where the former sector would experience a negative net impact under both low and high intensity measures (policy options 2 and 3), while the latter is expected to experience a positive net impact for these measures. Overall, in this domain the net impact is highest under the non-regulatory policy intervention.

Regarding the policy interventions in the area of business-to-government data sharing for the public interest the low intensity regulatory intervention (policy option 2) is preferred. The level of costs incurred under the low or high intensity regulatory measures is estimated, however, as evidenced by the assessment of a hypothetical scenario that will depend on a number of assumptions and the establishment of actual B2G use cases. Similarly, the level of compensation of these costs (e.g. free, covering marginal cost, fair return on investment) would depend on the actual use case and agreements made in the national governance context in this regard. The socio-economic impacts that could derive from better access to data to generate insights that help address societal issues (e.g. environmental challenges, health emergencies, improved public service delivery and evidence-based policy-making) has great potential in bringing significant benefits to society, an order of magnitude that can be higher than the associated costs if strategic use cases are established the enable reaping such benefits.

Finally, the research confirms that there is a clear risk for the data of European customers (public administrations, companies and citizens alike) when entrusting that data to cloud providers with a non-European nexus of activity – i.e. to cloud providers which are established outside of the EU, which process data outside of the EU (e.g. due to data centre locations), or which have a non-European mother company. Such data may be accessed by non-European public authorities on the basis of foreign legislation that does not systematically adhere to European values of independent supervision, and that does not provide a reasonable level of protection of fundamental rights. Countermeasures are available on the market that mitigate these risks, either by increasing transparency on risks and available safeguards, or by implementing technical, legal and operational measures that reduce the likelihood or impact of non-European access requests. While no single measure is universally feasible or perfectly effective against each individual threat, policy intervention that requires cloud providers to elevate their level of security by implementing some of these safeguards (policy option 3), as appropriate for the service, would be beneficial and effective in providing a level playing field for non-personal data, and would contribute to European data sovereignty.

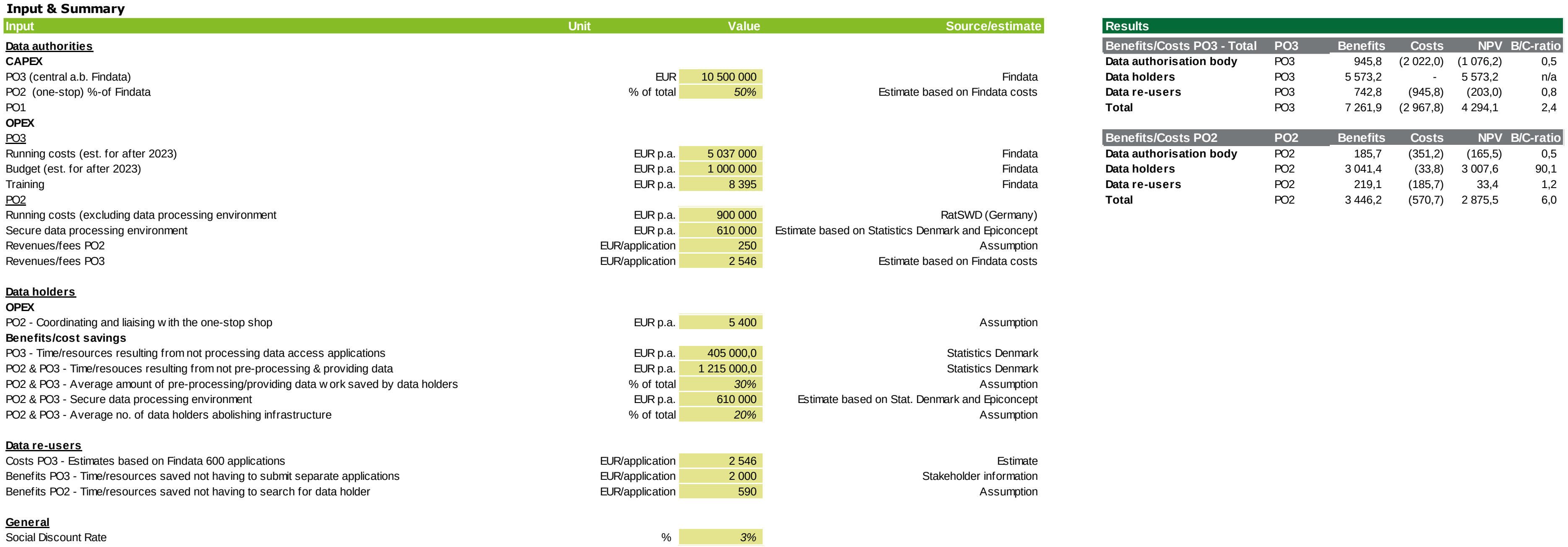
4 Annex

4.1 Annex I - Measures to enhance data governance

4.1.1 Cost-Benefit Analysis

4.1.1.1 Measures facilitating secondary use of sensitive data held by the public sector

The figure below presents the input summary for the cost-benefit analysis for Measures facilitating secondary use of sensitive data held by the public sector.



Input & Summary

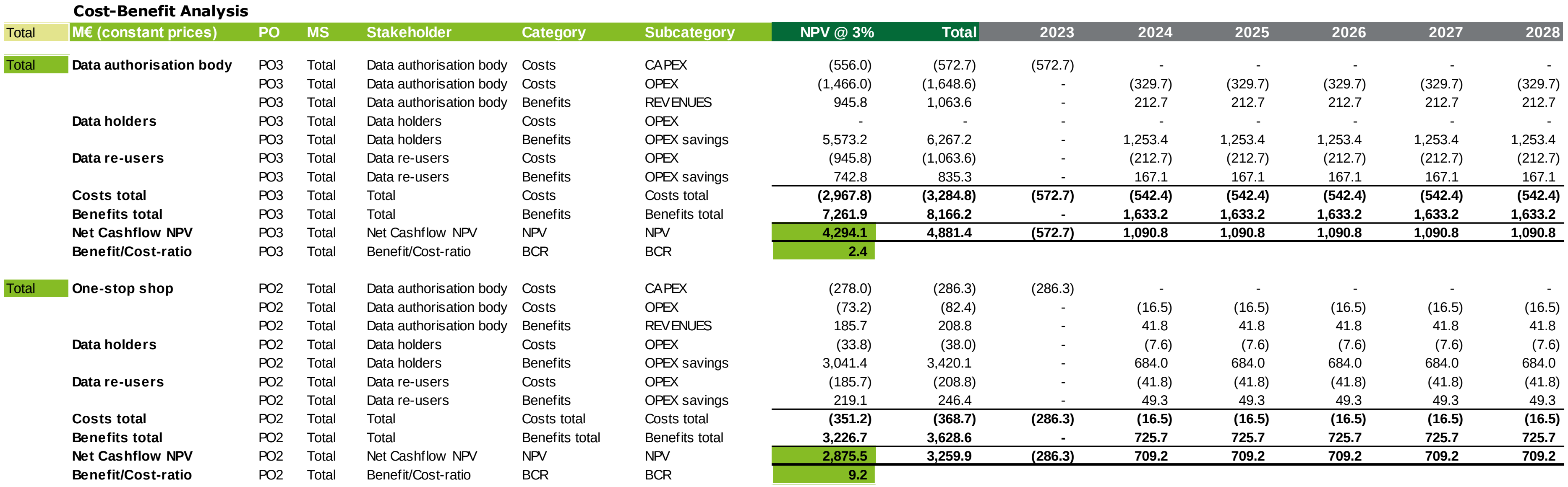
Input	Unit	Value	Source/estimate
Data authorities			
CAPEX			
PO3 (central a.b. Findata)	EUR	10 500 000	Findata
PO2 (one-stop) %-of Findata	% of total	*50%*	Estimate based on Findata costs
PO1			
OPEX			
PO3			
Running costs (est. for after 2023)	EUR p.a.	5 037 000	Findata
Budget (est. for after 2023)	EUR p.a.	1 000 000	Findata
Training	EUR p.a.	8 395	Findata
PO2			
Running costs (excluding data processing environment	EUR p.a.	900 000	RatSWD (Germany)
Secure data processing environment	EUR p.a.	610 000	Estimate based on Statistics Denmark and Epiconcept
Revenues/fees PO2	EUR/application	250	Assumption
Revenues/fees PO3	EUR/application	2 546	Estimate based on Findata costs
Data holders			
OPEX			
PO2 - Coordinating and liaising with the one-stop shop	EUR p.a.	5 400	Assumption
Benefits/cost savings			
PO3 - Time/resources resulting from not processing data access applications	EUR p.a.	405 000,0	Statistics Denmark
PO2 & PO3 - Time/resouces resulting from not pre-processing & providing data	EUR p.a.	1 215 000,0	Statistics Denmark
PO2 & PO3 - Average amount of pre-processing/providing data work saved by data holders	% of total	*30%*	Assumption
PO2 & PO3 - Secure data processing environment	EUR p.a.	610 000	Estimate based on Stat. Denmark and Epiconcept
PO2 & PO3 - Average no. of data holders abolishing infrastructure	% of total	*20%*	Assumption
Data re-users			
Costs PO3 - Estimates based on Findata 600 applications	EUR/application	2 546	Estimate
Benefits PO3 - Time/resources saved not having to submit separate applications	EUR/application	2 000	Stakeholder information
Benefits PO2 - Time/resources saved not having to search for data holder	EUR/application	590	Assumption
General			
Social Discount Rate	%	*3%*	

Results

Benefits/Costs PO3 - Total	PO3	Benefits	Costs	NPV	B/C-ratio
Data authorisation body	PO3	945,8	(2 022,0)	(1 076,2)	0,5
Data holders	PO3	5 573,2	-	5 573,2	n/a
Data re-users	PO3	742,8	(945,8)	(203,0)	0,8
Total	PO3	7 261,9	(2 967,8)	4 294,1	2,4

Benefits/Costs PO2	PO2	Benefits	Costs	NPV	B/C-ratio
Data authorisation body	PO2	185,7	(351,2)	(165,5)	0,5
Data holders	PO2	3 041,4	(33,8)	3 007,6	90,1
Data re-users	PO2	219,1	(185,7)	33,4	1,2
Total	PO2	3 446,2	(570,7)	2 875,5	6,0

The figure below presents the cost-benefit analysis for Measures facilitating secondary use of sensitive data held by the public sector

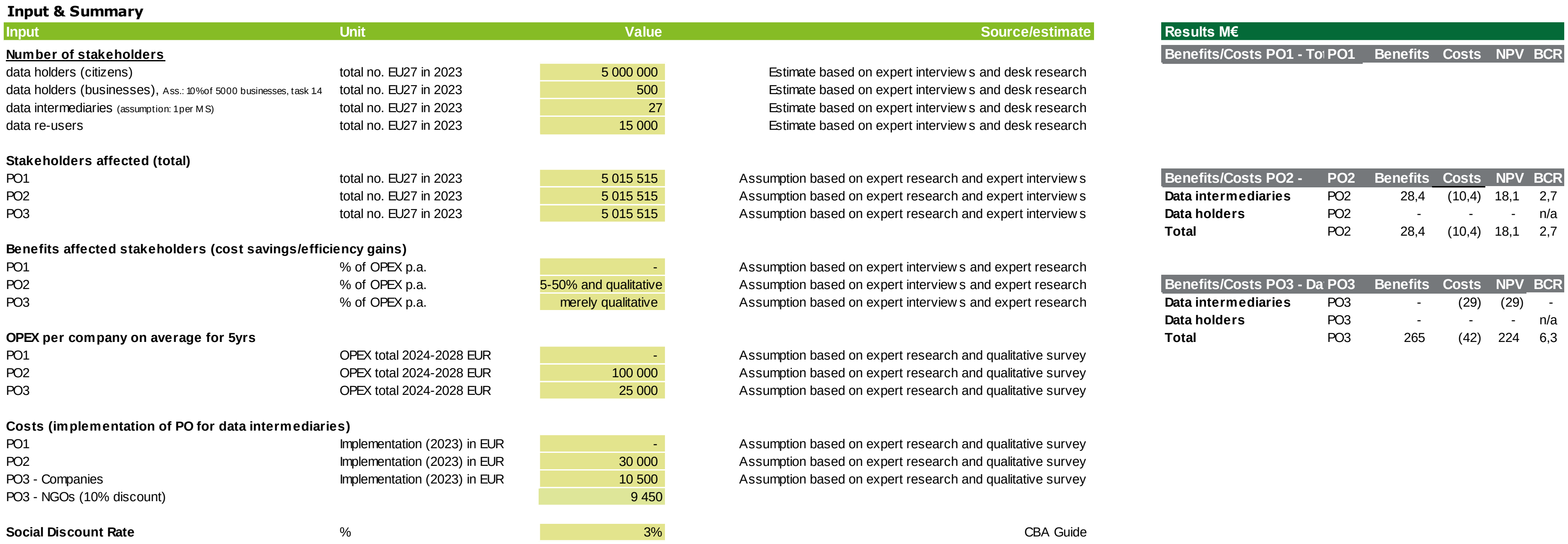


Cost-Benefit Analysis

Total	M€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Data authorisation body**	PO3	Total	Data authorisation body	Costs	CAPEX	(556.0)	(572.7)	(572.7)	-	-	-	-	-
		PO3	Total	Data authorisation body	Costs	OPEX	(1,466.0)	(1,648.6)	-	(329.7)	(329.7)	(329.7)	(329.7)	(329.7)
		PO3	Total	Data authorisation body	Benefits	REVENUES	945.8	1,063.6	-	212.7	212.7	212.7	212.7	212.7
	Data holders	PO3	Total	Data holders	Costs	OPEX	-	-	-	-	-	-	-	-
		PO3	Total	Data holders	Benefits	OPEX savings	5,573.2	6,267.2	-	1,253.4	1,253.4	1,253.4	1,253.4	1,253.4
	Data re-users	PO3	Total	Data re-users	Costs	OPEX	(945.8)	(1,063.6)	-	(212.7)	(212.7)	(212.7)	(212.7)	(212.7)
		PO3	Total	Data re-users	Benefits	OPEX savings	742.8	835.3	-	167.1	167.1	167.1	167.1	167.1
	Costs total	PO3	Total	Total	Costs	Costs total	**(2,967.8)**	**(3,284.8)**	**(572.7)**	**(542.4)**	**(542.4)**	**(542.4)**	**(542.4)**	**(542.4)**
	Benefits total	PO3	Total	Total	Benefits	Benefits total	**7,261.9**	**8,166.2**	**-**	**1,633.2**	**1,633.2**	**1,633.2**	**1,633.2**	**1,633.2**
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**4,294.1**	**4,881.4**	**(572.7)**	**1,090.8**	**1,090.8**	**1,090.8**	**1,090.8**	**1,090.8**
	Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	**2.4**							
Total	**One-stop shop**	PO2	Total	Data authorisation body	Costs	CAPEX	(278.0)	(286.3)	(286.3)	-	-	-	-	-
		PO2	Total	Data authorisation body	Costs	OPEX	(73.2)	(82.4)	-	(16.5)	(16.5)	(16.5)	(16.5)	(16.5)
		PO2	Total	Data authorisation body	Benefits	REVENUES	185.7	208.8	-	41.8	41.8	41.8	41.8	41.8
	Data holders	PO2	Total	Data holders	Costs	OPEX	(33.8)	(38.0)	-	(7.6)	(7.6)	(7.6)	(7.6)	(7.6)
		PO2	Total	Data holders	Benefits	OPEX savings	3,041.4	3,420.1	-	684.0	684.0	684.0	684.0	684.0
	Data re-users	PO2	Total	Data re-users	Costs	OPEX	(185.7)	(208.8)	-	(41.8)	(41.8)	(41.8)	(41.8)	(41.8)
		PO2	Total	Data re-users	Benefits	OPEX savings	219.1	246.4	-	49.3	49.3	49.3	49.3	49.3
	Costs total	PO2	Total	Total	Costs total	Costs total	**(351.2)**	**(368.7)**	**(286.3)**	**(16.5)**	**(16.5)**	**(16.5)**	**(16.5)**	**(16.5)**
	Benefits total	PO2	Total	Total	Benefits total	Benefits total	**3,226.7**	**3,628.6**	**-**	**725.7**	**725.7**	**725.7**	**725.7**	**725.7**
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**2,875.5**	**3,259.9**	**(286.3)**	**709.2**	**709.2**	**709.2**	**709.2**	**709.2**
	Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	**9.2**							

4.1.1.2 Establishing a certification scheme for data altruism mechanisms

The figure below presents the input summary for the cost-benefit analysis for Establishing a certification scheme for data altruism mechanisms.



Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders (citizens)	total no. EU27 in 2023	5 000 000	Estimate based on expert interviews and desk research
data holders (businesses), Ass.: 10%of 5000 businesses, task 14	total no. EU27 in 2023	500	Estimate based on expert interviews and desk research
data intermediaries (assumption: 1 per MS)	total no. EU27 in 2023	27	Estimate based on expert interviews and desk research
data re-users	total no. EU27 in 2023	15 000	Estimate based on expert interviews and desk research
Stakeholders affected (total)			
PO1	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
PO2	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
PO3	total no. EU27 in 2023	5 015 515	Assumption based on expert research and expert interviews
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	-	Assumption based on expert interviews and expert research
PO2	% of OPEX p.a.	5-50% and qualitative	Assumption based on expert interviews and expert research
PO3	% of OPEX p.a.	merely qualitative	Assumption based on expert interviews and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	-	Assumption based on expert research and qualitative survey
PO2	OPEX total 2024-2028 EUR	100 000	Assumption based on expert research and qualitative survey
PO3	OPEX total 2024-2028 EUR	25 000	Assumption based on expert research and qualitative survey
Costs (implementation of PO for data intermediaries)			
PO1	Implementation (2023) in EUR	-	Assumption based on expert research and qualitative survey
PO2	Implementation (2023) in EUR	30 000	Assumption based on expert research and qualitative survey
PO3 - Companies	Implementation (2023) in EUR	10 500	Assumption based on expert research and qualitative survey
PO3 - NGOs (10% discount)		9 450	
Social Discount Rate	%	3%	CBA Guide

Results M€

Benefits/Costs PO1 - To	PO1	Benefits	Costs	NPV	BCR

Benefits/Costs PO2 -	PO2	Benefits	Costs	NPV	BCR
Data intermediaries	PO2	28,4	(10,4)	18,1	2,7
Data holders	PO2	-	-	-	n/a
Total	PO2	28,4	(10,4)	18,1	2,7

Benefits/Costs PO3 - Da	PO3	Benefits	Costs	NPV	BCR
Data intermediaries	PO3	-	(29)	(29)	-
Data holders	PO3	-	-	-	n/a
Total	PO3	265	(42)	224	6,3

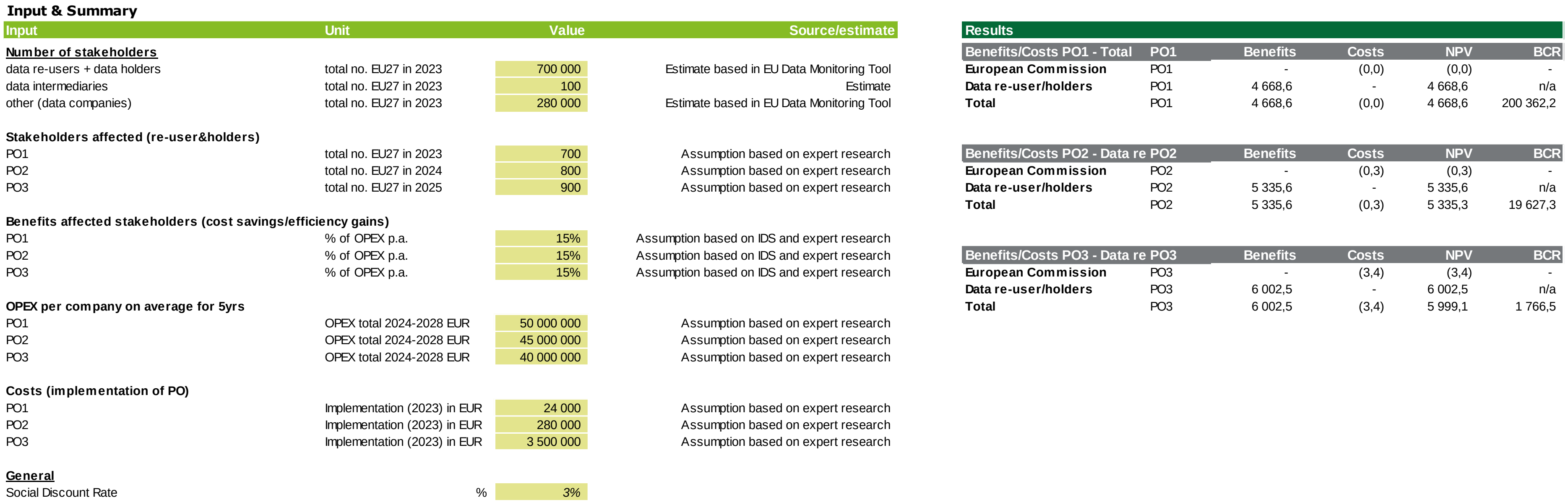
The figure below presents the cost-benefit analysis for Establishing a certification scheme for data altruism mechanisms.

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	Benefits	PO2	Total	Data intermediaries	no.	total intermediaries	n/a	n/a	1 255	1 255	1 255	1 255	1 255	1 255
		PO2	Total	Data intermediaries	no.	participating in the voluntary scheme			125	11	125	125	125	125
		PO2	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	10%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	revenues	total additional revenues	8 966 675	10 220 000	-	220 000	2 500 000	2 500 000	2 500 000	2 500 000
		PO2	Total	Data intermediaries	add.revenues	aver. revenues per intermediary	88 926	100 000	-	20 000	20 000	20 000	20 000	20 000
		PO2	Total	Data re-users	no.	no	n/a	n/a	15 000	15 000	15 000	15 000	15 000	15 000
		PO2	Total	Data re-users	Benefits	total value of data**	19 460 297	22 002 250	-	3 500 350	4 000 400	4 500 450	5 000 500	5 000 550
		PO2	Total	Data re-users	Benefits	average value per unit of data	46	50		10	10	10	10	10
		PO2	Total	Data holders	no.	total holders (Citizens)	n/a	n/a	5 000 000	5 000 000	5 000 000	5 000 000	5 000 000	5 000 000
		PO2	Total	Data holders	no.	participating in the voluntary scheme	n/a	n/a	300 000	350 000	400 000	450 000	500 000	500 000
		PO2	Total	Data holders	no.	total holders (Companies)	n/a	n/a	500	500	500	500	500	500
				Data holders	no.	participating in the voluntary scheme	n/a	n/a	30	35	40	45	50	55
		PO2	Total	Data intermediaries	Client base	number of clients %			-	40%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	Client base	number of use cases			-	35%	10%	10%	10%	10%
		PO2	Total	Data intermediaries	Increased competition	B2B Market			-	-	3%	3%	3%	3%
		PO2	Total	Data intermediaries	Increased competition	C2B Market			-	10%	5%	5%	5%	5%
		PO2	Total	Data intermediaries	Benefits	Benefits	-							
	Costs	PO2	Total	Data re-users	Costs	Maintenance costs*	n/a	n/a	-	-	-	-	-	-
		PO2	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification for the first time	(3 640 777)	(3 750 000)	(3 750 000)	-	-	-	-	-
		PO2	Total	Data intermediaries	Costs	Compliance recurrent costs of renew ing certification	(6 725 007)	(7 665 000)	-	(165 000)	(1 875 000)	(1 875 000)	(1 875 000)	(1 875 000)
		PO2	Total	Data intermediaries	Reduced Competition	B2B Market			-25%					
	Costs total	PO2	Total	Total	Costs	Costs total	**(10 365 783)**	**(11 415 000)**	**(3 750 000)**	**(165 000)**	**(1 875 000)**	**(1 875 000)**	**(1 875 000)**	**(1 875 000)**
	Benefits total	PO2	Total	Total	Benefits	Benefits total	**28 426 972**	**10 220 000**	**-**	**220 000**	**2 500 000**	**2 500 000**	**2 500 000**	**2 500 000**
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**(1 399 108)**	**(1 195 000)**	**(3 750 000)**	**55 000**	**625 000**	**625 000**	**625 000**	**625 000**
	Benefit/Cost-ratio	PO2	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**2,7**							

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	Benefits	PO3	Total	Data intermediaries	no.	Companies	n/a	n/a	1 255	1 255	1 255	1 255	1 320	1 350
		PO3	Total	Data intermediaries	no.	NGOs			55	55	55	60	65	75
		PO3	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	10%	10%	10%	10%	10%	10%
		PO3	Total	Data re-users	no.	no.	n/a	n/a	15 000	15 000	15 500	15 750	16 000	16 500
		PO3	Total	Data re-users	Benefits	total value of data**	265 491 825	300 030 000	-	50 005 000	55 005 500	60 006 000	65 006 500	70 007 000
		PO3	Total	Data re-users	Benefits	average value per unit of data	44	50	-	10,00	10,00	10,00	10,00	10,00
		PO3	Total	Data holders*	no.	Citizens	n/a	n/a	5 000 000	5 000 000	5 500 000	6 000 000	6 500 000	7 000 000
		PO3	Total	Data holders*	no.	Companies	n/a	n/a	500	500	550	600	650	700
		PO3	Total	Data holders	Benefits	Qualitative Benefit: increased trust	n/a	n/a						
	Costs	PO3	Total	Data re-users	Costs	Maintenance costs*	n/a	n/a	[already included in task 1.1]					
		PO3	Total	Data intermediaries	Companies	One-off authorisation costs per company	(10 194)	(10 500)	(10 500)	-	-	-	-	-
		PO3	Total	Data intermediaries	Companies	total one-off authorisation costs	(12 793 689)	(13 177 500)	(13 177 500)					
		PO3	Total	Data intermediaries	NGOs	One-off authorisation costs per NGO	(9 175)	(9 450)	(9 450)	-	-	-	-	-
		PO3	Total	Data intermediaries	NGOs	total one-off authorisation costs	(504 612)	(519 750)	(519 750)					
		PO3	Total	Data intermediaries	Costs	recurrent costs/maintenance costs per intermediary	(22 899)	(25 000)		(5 000)	(5 000)	(5 000)	(5 000)	(5 000)
		PO3	Total	Data intermediaries	Costs	total recurrent costs/maintenance costs	(28 578 796)	(32 175 000)	-	(6 275 000)	(6 275 000)	(6 275 000)	(6 600 000)	(6 750 000)
	Costs total	PO3	Total	Total	Costs	Costs total	**(41 877 097)**	**(45 872 250)**	**(13 697 250)**	**(6 275 000)**	**(6 275 000)**	**(6 275 000)**	**(6 600 000)**	**(6 750 000)**
	Benefits total	PO3	Total	Total	Benefits	Benefits total	**265 491 825**	**300 030 000**	**-**	**50 005 000**	**55 005 500**	**60 006 000**	**65 006 500**	**70 007 000**
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**223 614 728**	**254 157 750**	**(13 697 250)**	**43 730 000**	**48 730 500**	**53 731 000**	**58 406 500**	**63 257 000**
	Benefit/Cost-ratio	PO3	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**6,3**							

4.1.1.3 Establishing a European structure for governance aspects of data sharing

The figure below presents the input summary for the cost-benefit analysis for Establishing a European structure for governance aspects of data sharing.



Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data re-users + data holders	total no. EU27 in 2023	700 000	Estimate based in EU Data Monitoring Tool
data intermediaries	total no. EU27 in 2023	100	Estimate
other (data companies)	total no. EU27 in 2023	280 000	Estimate based in EU Data Monitoring Tool
Stakeholders affected (re-user&holders)			
PO1	total no. EU27 in 2023	700	Assumption based on expert research
PO2	total no. EU27 in 2024	800	Assumption based on expert research
PO3	total no. EU27 in 2025	900	Assumption based on expert research
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	15%	Assumption based on IDS and expert research
PO2	% of OPEX p.a.	15%	Assumption based on IDS and expert research
PO3	% of OPEX p.a.	15%	Assumption based on IDS and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	50 000 000	Assumption based on expert research
PO2	OPEX total 2024-2028 EUR	45 000 000	Assumption based on expert research
PO3	OPEX total 2024-2028 EUR	40 000 000	Assumption based on expert research
Costs (implementation of PO)			
PO1	Implementation (2023) in EUR	24 000	Assumption based on expert research
PO2	Implementation (2023) in EUR	280 000	Assumption based on expert research
PO3	Implementation (2023) in EUR	3 500 000	Assumption based on expert research
General			
Social Discount Rate	%	*3%*	

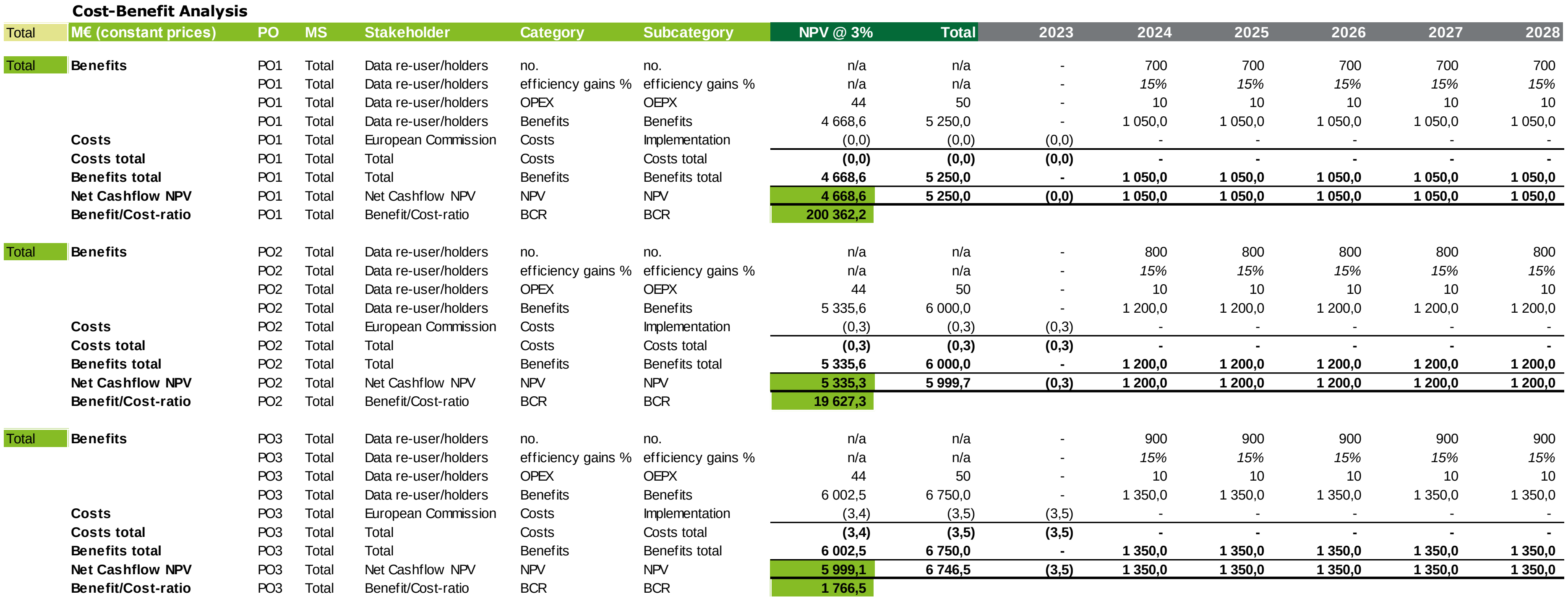
Results

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
European Commission	PO1	-	(0,0)	(0,0)	-
Data re-user/holders	PO1	4 668,6	-	4 668,6	n/a
Total	PO1	4 668,6	(0,0)	4 668,6	200 362,2

Benefits/Costs PO2 - Data re	PO2	Benefits	Costs	NPV	BCR
European Commission	PO2	-	(0,3)	(0,3)	-
Data re-user/holders	PO2	5 335,6	-	5 335,6	n/a
Total	PO2	5 335,6	(0,3)	5 335,3	19 627,3

Benefits/Costs PO3 - Data re	PO3	Benefits	Costs	NPV	BCR
European Commission	PO3	-	(3,4)	(3,4)	-
Data re-user/holders	PO3	6 002,5	-	6 002,5	n/a
Total	PO3	6 002,5	(3,4)	5 999,1	1 766,5

The figure below presents the cost-benefit analysis for Establishing a European structure for governance aspects of data sharing.

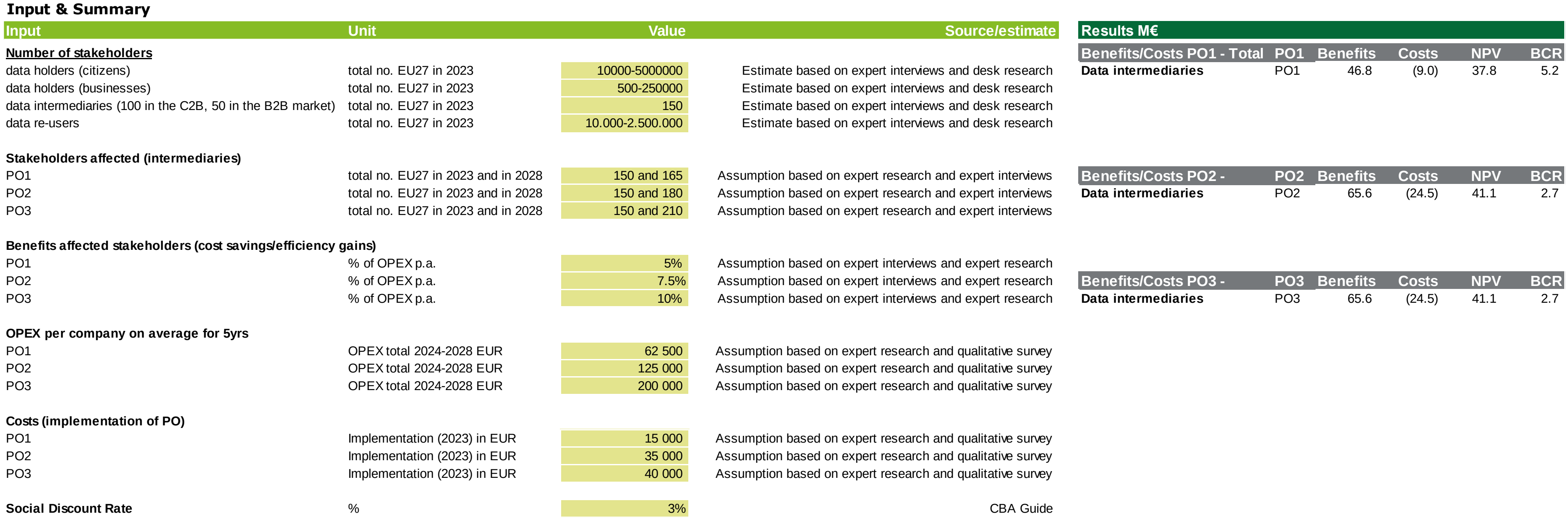


Cost-Benefit Analysis

Total	M€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO1	Total	Data re-user/holders	no.	no.	n/a	n/a	-	700	700	700	700	700
		PO1	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO1	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO1	Total	Data re-user/holders	Benefits	Benefits	4 668,6	5 250,0	-	1 050,0	1 050,0	1 050,0	1 050,0	1 050,0
	Costs	PO1	Total	European Commission	Costs	Implementation	(0,0)	(0,0)	(0,0)	-	-	-	-	-
	Costs total	PO1	Total	Total	Costs	Costs total	**(0,0)**	**(0,0)**	**(0,0)**	**-**	**-**	**-**	**-**	**-**
	Benefits total	PO1	Total	Total	Benefits	Benefits total	**4 668,6**	**5 250,0**	**-**	**1 050,0**	**1 050,0**	**1 050,0**	**1 050,0**	**1 050,0**
	Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	**4 668,6**	**5 250,0**	**(0,0)**	**1 050,0**	**1 050,0**	**1 050,0**	**1 050,0**	**1 050,0**
	Benefit/Cost-ratio	PO1	Total	Benefit/Cost-ratio	BCR	BCR	**200 362,2**							
Total	**Benefits**	PO2	Total	Data re-user/holders	no.	no.	n/a	n/a	-	800	800	800	800	800
		PO2	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO2	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO2	Total	Data re-user/holders	Benefits	Benefits	5 335,6	6 000,0	-	1 200,0	1 200,0	1 200,0	1 200,0	1 200,0
	Costs	PO2	Total	European Commission	Costs	Implementation	(0,3)	(0,3)	(0,3)	-	-	-	-	-
	Costs total	PO2	Total	Total	Costs	Costs total	**(0,3)**	**(0,3)**	**(0,3)**	**-**	**-**	**-**	**-**	**-**
	Benefits total	PO2	Total	Total	Benefits	Benefits total	**5 335,6**	**6 000,0**	**-**	**1 200,0**	**1 200,0**	**1 200,0**	**1 200,0**	**1 200,0**
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**5 335,3**	**5 999,7**	**(0,3)**	**1 200,0**	**1 200,0**	**1 200,0**	**1 200,0**	**1 200,0**
	Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	**19 627,3**							
Total	**Benefits**	PO3	Total	Data re-user/holders	no.	no.	n/a	n/a	-	900	900	900	900	900
		PO3	Total	Data re-user/holders	efficiency gains %	efficiency gains %	n/a	n/a	-	*15%*	*15%*	*15%*	*15%*	*15%*
		PO3	Total	Data re-user/holders	OPEX	OEPX	44	50	-	10	10	10	10	10
		PO3	Total	Data re-user/holders	Benefits	Benefits	6 002,5	6 750,0	-	1 350,0	1 350,0	1 350,0	1 350,0	1 350,0
	Costs	PO3	Total	European Commission	Costs	Implementation	(3,4)	(3,5)	(3,5)	-	-	-	-	-
	Costs total	PO3	Total	Total	Costs	Costs total	**(3,4)**	**(3,5)**	**(3,5)**	**-**	**-**	**-**	**-**	**-**
	Benefits total	PO3	Total	Total	Benefits	Benefits total	**6 002,5**	**6 750,0**	**-**	**1 350,0**	**1 350,0**	**1 350,0**	**1 350,0**	**1 350,0**
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**5 999,1**	**6 746,5**	**(3,5)**	**1 350,0**	**1 350,0**	**1 350,0**	**1 350,0**	**1 350,0**
	Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	**1 766,5**							

4.1.1.4 Establishing a certification framework for data intermediaries

The figure below presents the input summary for the cost-benefit analysis for Establishing a certification framework for data intermediaries.



Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders (citizens)	total no. EU27 in 2023	10000-5000000	Estimate based on expert interviews and desk research
data holders (businesses)	total no. EU27 in 2023	500-250000	Estimate based on expert interviews and desk research
data intermediaries (100 in the C2B, 50 in the B2B market)	total no. EU27 in 2023	150	Estimate based on expert interviews and desk research
data re-users	total no. EU27 in 2023	10.000-2.500.000	Estimate based on expert interviews and desk research
Stakeholders affected (intermediaries)			
PO1	total no. EU27 in 2023 and in 2028	150 and 165	Assumption based on expert research and expert interviews
PO2	total no. EU27 in 2023 and in 2028	150 and 180	Assumption based on expert research and expert interviews
PO3	total no. EU27 in 2023 and in 2028	150 and 210	Assumption based on expert research and expert interviews
Benefits affected stakeholders (cost savings/efficiency gains)			
PO1	% of OPEX p.a.	5%	Assumption based on expert interviews and expert research
PO2	% of OPEX p.a.	7.5%	Assumption based on expert interviews and expert research
PO3	% of OPEX p.a.	10%	Assumption based on expert interviews and expert research
OPEX per company on average for 5yrs			
PO1	OPEX total 2024-2028 EUR	62 500	Assumption based on expert research and qualitative survey
PO2	OPEX total 2024-2028 EUR	125 000	Assumption based on expert research and qualitative survey
PO3	OPEX total 2024-2028 EUR	200 000	Assumption based on expert research and qualitative survey
Costs (implementation of PO)			
PO1	Implementation (2023) in EUR	15 000	Assumption based on expert research and qualitative survey
PO2	Implementation (2023) in EUR	35 000	Assumption based on expert research and qualitative survey
PO3	Implementation (2023) in EUR	40 000	Assumption based on expert research and qualitative survey
Social Discount Rate	%	3%	CBA Guide

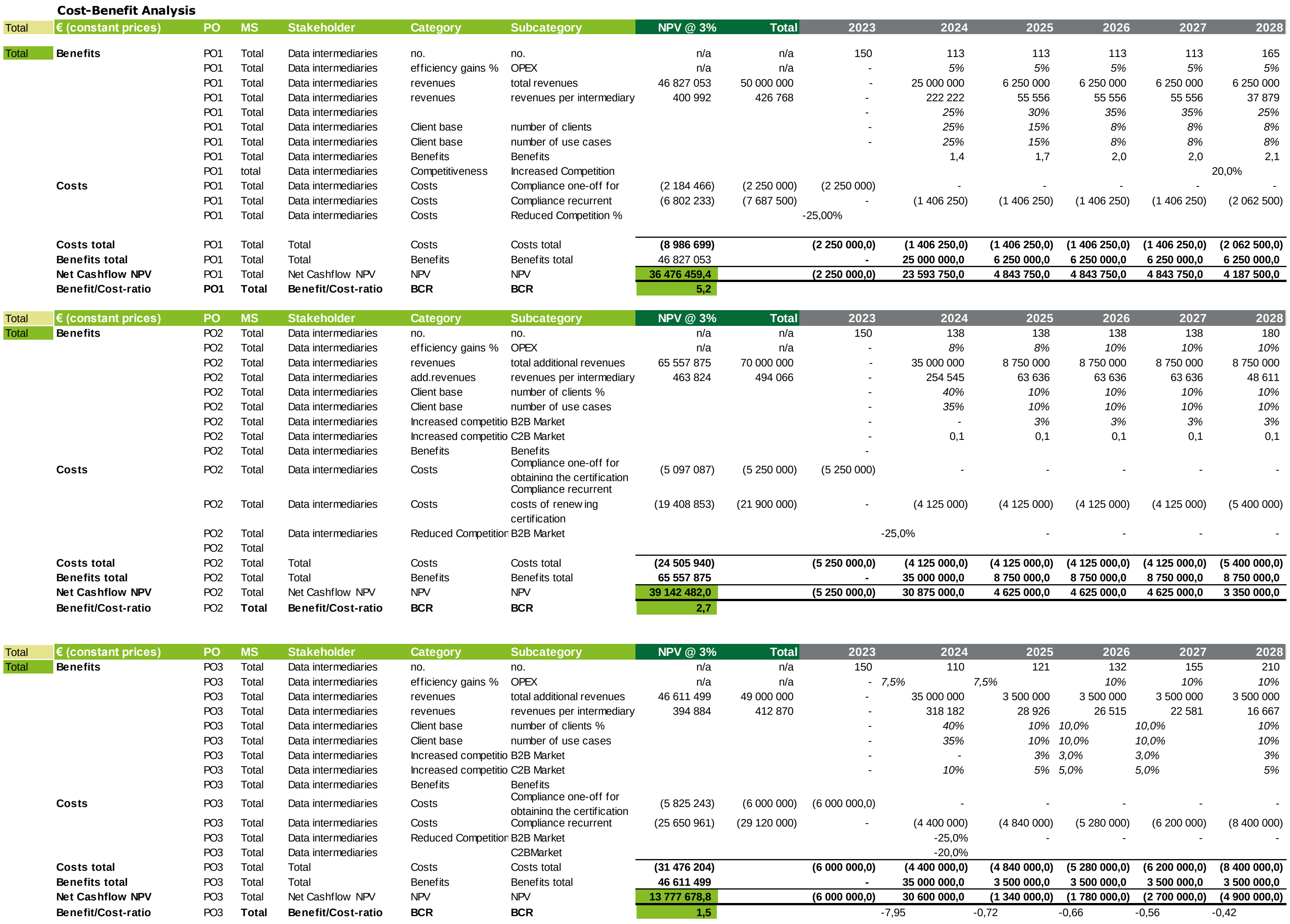
Results M€

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
Data intermediaries	PO1	46.8	(9.0)	37.8	5.2

Benefits/Costs PO2 -	PO2	Benefits	Costs	NPV	BCR
Data intermediaries	PO2	65.6	(24.5)	41.1	2.7

Benefits/Costs PO3 -	PO3	Benefits	Costs	NPV	BCR
Data intermediaries	PO3	65.6	(24.5)	41.1	2.7

The figures below presents the cost-benefit analysis for Establishing a certification framework for data intermediaries.



Cost-Benefit Analysis

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO1	Total	Data intermediaries	no.	no.	n/a	n/a	150	113	113	113	113	165
		PO1	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	*5%*	*5%*	*5%*	*5%*	*5%*
		PO1	Total	Data intermediaries	revenues	total revenues	46 827 053	50 000 000	-	25 000 000	6 250 000	6 250 000	6 250 000	6 250 000
		PO1	Total	Data intermediaries	revenues	revenues per intermediary	400 992	426 768	-	222 222	55 556	55 556	55 556	37 879
		PO1	Total	Data intermediaries					-	*25%*	*30%*	*35%*	*35%*	*25%*
		PO1	Total	Data intermediaries	Client base	number of clients			-	*25%*	*15%*	*8%*	*8%*	*8%*
		PO1	Total	Data intermediaries	Client base	number of use cases			-	*25%*	*15%*	*8%*	*8%*	*8%*
		PO1	Total	Data intermediaries	Benefits	Benefits				1,4	1,7	2,0	2,0	2,1
		PO1	total	Data intermediaries	Competitiveness	Increased Competition								20,0%
	Costs	PO1	Total	Data intermediaries	Costs	Compliance one-off for	(2 184 466)	(2 250 000)	(2 250 000)	-	-	-	-	-
		PO1	Total	Data intermediaries	Costs	Compliance recurrent	(6 802 233)	(7 687 500)	-	(1 406 250)	(1 406 250)	(1 406 250)	(1 406 250)	(2 062 500)
		PO1	Total	Data intermediaries	Costs	Reduced Competition %			-25,00%					
	Costs total	PO1	Total	Total	Costs	Costs total	**(8 986 699)**		**(2 250 000,0)**	**(1 406 250,0)**	**(1 406 250,0)**	**(1 406 250,0)**	**(1 406 250,0)**	**(2 062 500,0)**
	Benefits total	PO1	Total	Total	Benefits	Benefits total	46 827 053		**-**	**25 000 000,0**	**6 250 000,0**	**6 250 000,0**	**6 250 000,0**	**6 250 000,0**
	Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	**36 476 459,4**		**(2 250 000,0)**	**23 593 750,0**	**4 843 750,0**	**4 843 750,0**	**4 843 750,0**	**4 187 500,0**
	Benefit/Cost-ratio	**PO1**	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**5,2**							

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO2	Total	Data intermediaries	no.	no.	n/a	n/a	150	138	138	138	138	180
		PO2	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	*8%*	*8%*	*10%*	*10%*	*10%*
		PO2	Total	Data intermediaries	revenues	total additional revenues	65 557 875	70 000 000	-	35 000 000	8 750 000	8 750 000	8 750 000	8 750 000
		PO2	Total	Data intermediaries	add.revenues	revenues per intermediary	463 824	494 066	-	254 545	63 636	63 636	63 636	48 611
		PO2	Total	Data intermediaries	Client base	number of clients %			-	*40%*	*10%*	*10%*	*10%*	*10%*
		PO2	Total	Data intermediaries	Client base	number of use cases			-	*35%*	*10%*	*10%*	*10%*	*10%*
		PO2	Total	Data intermediaries	Increased competitio	B2B Market			-	-	*3%*	*3%*	*3%*	*3%*
		PO2	Total	Data intermediaries	Increased competitio	C2B Market			-	0,1	0,1	0,1	0,1	0,1
		PO2	Total	Data intermediaries	Benefits	Benefits			-					
	Costs	PO2	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification	(5 097 087)	(5 250 000)	(5 250 000)	-	-	-	-	-
		PO2	Total	Data intermediaries	Costs	Compliance recurrent costs of renewing certification	(19 408 853)	(21 900 000)	-	(4 125 000)	(4 125 000)	(4 125 000)	(4 125 000)	(5 400 000)
		PO2	Total	Data intermediaries	Reduced Competition	B2B Market				-25,0%	-	-	-	-
		PO2	Total											
	Costs total	PO2	Total	Total	Costs	Costs total	**(24 505 940)**		**(5 250 000,0)**	**(4 125 000,0)**	**(4 125 000,0)**	**(4 125 000,0)**	**(4 125 000,0)**	**(5 400 000,0)**
	Benefits total	PO2	Total	Total	Benefits	Benefits total	**65 557 875**		**-**	**35 000 000,0**	**8 750 000,0**	**8 750 000,0**	**8 750 000,0**	**8 750 000,0**
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**39 142 482,0**		**(5 250 000,0)**	**30 875 000,0**	**4 625 000,0**	**4 625 000,0**	**4 625 000,0**	**3 350 000,0**
	Benefit/Cost-ratio	PO2	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**2,7**							

Total	€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	**Benefits**	PO3	Total	Data intermediaries	no.	no.	n/a	n/a	150	110	121	132	155	210
		PO3	Total	Data intermediaries	efficiency gains %	OPEX	n/a	n/a	-	*7,5%*	*7,5%*	*10%*	*10%*	*10%*
		PO3	Total	Data intermediaries	revenues	total additional revenues	46 611 499	49 000 000	-	35 000 000	3 500 000	3 500 000	3 500 000	3 500 000
		PO3	Total	Data intermediaries	revenues	revenues per intermediary	394 884	412 870	-	318 182	28 926	26 515	22 581	16 667
		PO3	Total	Data intermediaries	Client base	number of clients %			-	*40%*	*10%*	*10,0%*	*10,0%*	*10%*
		PO3	Total	Data intermediaries	Client base	number of use cases			-	*35%*	*10%*	*10,0%*	*10,0%*	*10%*
		PO3	Total	Data intermediaries	Increased competitio	B2B Market			-	-	*3%*	*3,0%*	*3,0%*	*3%*
		PO3	Total	Data intermediaries	Increased competitio	C2B Market			-	*10%*	*5%*	*5,0%*	*5,0%*	*5%*
		PO3	Total	Data intermediaries	Benefits	Benefits								
	Costs	PO3	Total	Data intermediaries	Costs	Compliance one-off for obtaining the certification	(5 825 243)	(6 000 000)	(6 000 000,0)	-	-	-	-	-
		PO3	Total	Data intermediaries	Costs	Compliance recurrent	(25 650 961)	(29 120 000)	-	(4 400 000)	(4 840 000)	(5 280 000)	(6 200 000)	(8 400 000)
		PO3	Total	Data intermediaries	Reduced Competition	B2B Market				-25,0%	-	-	-	-
		PO3	Total	Data intermediaries		C2BMarket				-20,0%				
	Costs total	PO3	Total	Total	Costs	Costs total	**(31 476 204)**		**(6 000 000,0)**	**(4 400 000,0)**	**(4 840 000,0)**	**(5 280 000,0)**	**(6 200 000,0)**	**(8 400 000,0)**
	Benefits total	PO3	Total	Total	Benefits	Benefits total	**46 611 499**		**-**	**35 000 000,0**	**3 500 000,0**	**3 500 000,0**	**3 500 000,0**	**3 500 000,0**
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**13 777 678,8**		**(6 000 000,0)**	**30 600 000,0**	**(1 340 000,0)**	**(1 780 000,0)**	**(2 700 000,0)**	**(4 900 000,0)**
	Benefit/Cost-ratio	PO3	**Total**	**Benefit/Cost-ratio**	**BCR**	**BCR**	**1,5**			-7,95	-0,72	-0,66	-0,56	-0,42

4.1.2 Measures facilitating secondary use of sensitive data held by the public sector: Case Studies

This Annex contains the case studies carried out for Measures facilitating secondary use of sensitive data held by the public sector, namely Findata, RatSWD, the Scottish national Safe Haven, Statistics Denmark, OpenSAFELY, and Statbel.

4.1.2.1 Findata

4.1.2.1.1 Introduction

Finland has a long history of collecting extensive data in registers but making use of the data has been difficult and inefficient. In 2019 a new **Act on Secondary Use of Health and Social Data** entered in force in Finland. With the new enabling legislation, Finland has become the first country in the world to successfully enact a law on the secondary use of well-being data that meets the requirements of the European General Data Protection Regulation (GDPR).

The new legislation **enables and expands the use of social and healthcare data from the traditional areas of scientific research and statistics** to those of management/control of social welfare and healthcare, development and innovations, knowledge management, education, authorities' planning and forecasting tasks, and steering and supervision of work. The new Act facilitates the establishment of a **new central data permit authority** in Finland, Findata.

The objectives of establishing a new centralised body (Findata as Data Permit Authority) devoted to the implementation of the secondary use of health and social data have been mainly:

- **To enable efficient and secure processing of personal data collected during the provision of social and health care** as well as personal data collected for the purpose of steering, supervision, researching and collecting statistics on the social and health care sector, in full compliance with GDPR prescriptions;
- **To allow the collected personal data to be combined** with the personal data held by Social Insurance Institution of Finland, Population Register Centre, Statistics Finland and Finnish Centre for Pensions;
- To secure the legitimate expectations, rights and freedoms of individuals when processing personal data.

Findata is the **one-stop-shop** responsible for streamlining and securing the secondary use of social and health data. It guarantees a flourishing ecosystem around the secondary use of social and health data streamlining the processes for the **issuing of research permits** and data collection and ensuring that data is being used in **secure environments**, thereby maintaining the trust that the general public have in authorities and the public sector.

This case study details how Findata operates in practice, using the GOFA model.

4.1.2.1.2 Governance

Findata operates directly under the **Ministry of Social Affairs and Health** and is a separate legal entity functioning as part of the **National Institute of Health and Welfare** (THL). Findata's operations are supervised by the **Parliamentary Ombudsman and the Data Protection Ombudsman**, among others. The **National Supervisory Authority for Welfare and Health Valvira** monitors Findata's data secure user environments. In addition, Findata must give an annual report to the Data Protection Ombudsman regarding the processing of health and social data and the related log data.

To steer the operations of Findata and to develop the cooperation, the Ministry of Social Affairs and Health organises a **steering committee** every three years and elects a **chair person** for the commitee. The members of the steering group have been choosen from:

- the Ministry of Social Affairs and Health,
- the Finnish Institute for Health and Welfare,
- the Social Insurance Institution of Finland,
- the Finnish Centre for Pensions,
- the Populaton Register Centre,
- Statistics Finland,
- the Finnish Institute of Occupational Health,
- the Finnish Medicines Agency Fimea, and
- representatives of social welfare and health care service providers.

The **task of the steering committee** is to process and make proposals to the National Institute for Health and Welfare and the Ministry of Social Affairs and Health on:

- the annual action plan of the Data Permit Authority and the associated budget;
- the report on operations and financial statements as applicable to the Data Permit Authority;
- the joint development of controllers and the resources allocated to the task;
- the resources allocated for the development of information systems and cooperation;

Additionally, the steering committee is responsible to:

- set goal indicators for the processes of the Data Permit Authority and initiate external audits on the processes;
- if necessary, make a proposal to the National Institute for Health and Welfare and the Ministry of Social Affairs and Health on the improvement of the Data Permit Authority’s operations;

The Ministry of Social Affairs and Health established a **high-level expert group** for Findata. The task of the group is to create guidelines on anonymisation, data protection and data security for the Data Permit Authority’s operations. The expert group must have an expert on each of the following fields: artificial intelligence, data analytics, data security, data protection, suitable research, statistics and statistical service as well as a representative of the Data Permit Authority.

4.1.2.1.3 Operations

Findata grants permits to allow the secure and easy use of social and health data for the purposes laid down by the Act on the Secondary Use of Health and Social Data. As such, Findata is a one-stop-shop for data, centralising the decision procedures and access to the data.

Thanks to Findata, retrieving combined health and social data from different sources is easier, faster and possible with just one permit application, removing the need to approach each authority and data source separately. Previously, obtaining the permits and data has taken as long as up to two or three years. The Act guarantees the provision of a permit within just three months. For exceptionally complex data requests that can cover several data registries, the data permit authority can extend the time it takes to obtain a permit by a maximum of three extra months. In addition, the data is provided with little delay, no later than within 60 business days after the permit has been approved.

Findata is responsible for ensuring the ethically sustainable use of data. It makes decisions on data permits concerning data held by other controllers, and is responsible for the collection, combination, pre-processing and disclosure of data for secondary use, in accordance with the Act. Furthermore, the data permit authority maintains a data request management system to forward and process data requests and permit applications. Findata also maintains a secure hosting service for receiving or disclosing personal data and a secure operating environment, in which the permit holder may process the personal data he/she has been disclosed on the basis of data permit. It also supervises compliance with the terms and conditions of the permit it has issued. The data permit may be revoked if the permit holder fails to comply with the law or the terms and conditions of the permit. Lastly, the data permit authority is responsible for the pseudonymisation and the anonymisation of personal data.

The secondary use of health and social data means that the data generated during health and social services are also used for other purposes, in addition to the primary purposes for which they were originally saved. Health and Social data were initially only used for traditional **scientific research and statistic in the health care domain**. Thanks to the Findata approach, it is possible now to activate new data usage, such as:

- **Development and innovations activities (R&D):** not only researchers but many diverse worldwide health technologies and life-science companies benefit from this new approach. Thanks to the access to social and healthcare data reserves, these companies can start to see opportunities in Finland and expand their R&D activities to the country. These activities however must be aimed at promoting national health or social security, at developing social welfare and healthcare services or service systems, or at protecting the health and well-being of individuals or securing for them the related rights and freedoms.
- **Knowledge management:** thanks to Findata, each organisation can improve its knowledge-management opportunities in social welfare and healthcare sectors with easier access to comprehensive data sources and new services around high-quality registered data.
- **Planning and forecasting of the activities and initiatives performed by social and health care Finnish authorities**: to transform the Finnish authorities in a data-driven organisations, the data collected from Findata can be used as a basis for the planning of central initiatives and programmes.
- **Governance and supervision of social and health care organisations:** the governance and control of organisations by social and health care **Finnish authorities** based on personal data and statistics and/ or on data received from case-studies, such as the National Institute for Health and Welfare or the Population Register centre.
- **Education:** higher education institutions, such as biomedical campus universities, can benefit from the data stored in Findata using data for the development of projects, publications, preparation of seminaries and other materials.

In 2019 a temporary steering group was put in place to prepare the launch of Findata operations. In the summer of 2019 the Findata director and staff were recruited. The Findata website has been opened since August 2019. From the 1 November 2019, the Findata help desk (website, e-mail and phone service) is available. Currently, it Findata is open to receive data requests for anonymised statistical data (since 1 January 2020). The system also collects data permit applications for individual-level data from 1 April 2020.

As regards future developments, Findata aims to guarantee a secure remote use environment for customers and improve data set descriptions, data management and methods.

In 2020 Findata counts 15 people. A first investment has been made to hire profiles with legal expertise and administrative skills. Additionally, several ICT profiles have been hired. The hiring process is still on-going, and in 2021, Findata expects to have 20 professionals. The hiring strategy aims to guarantee a good mix of skills that can enable the use of the new technology and methods, analytics skills and achieve a good understanding of research practices.

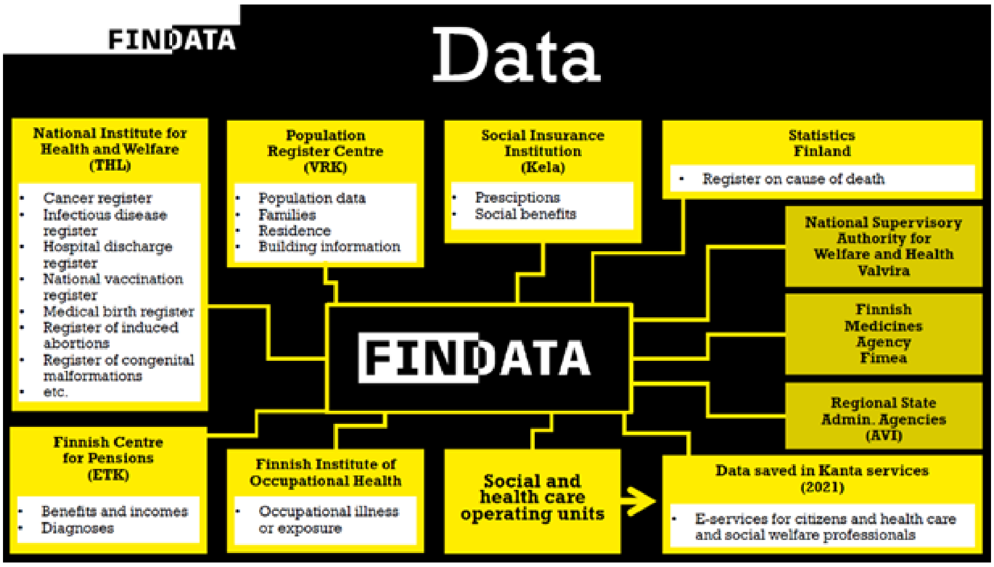
In addition to the citizens, the users who can access Findata also include:

- **Authorities:** among others Social Insurance Institution of Finland (KELA), Social Welfare Office and National Supervisory Authority for Welfare and Health – Valvira.
- **Institutes:** research institutes, universities and biomedical campuses.
- **Companies:** pharmaceutical companies, health technology and life-science companies.
- **Professionals:** Healthcare and social welfare professionals, professors and PhDs.

Findata is responsible for data permits and data requests when the data is combined from the following holders:

- Social and health care operating units;
- Finnish Institute for Health and Welfare (does not apply to data collected for statistical purposes);
- Social Insurance Institution of Finland Kela (benefits and prescriptions);
- Data saved in Kanta Services;
- Finnish Centre for Pensions (work and earnings data, benefits and the bases for them);
- National Supervisory Authority for Welfare and Health Valvira;
- Finnish Medicines Agency Fimea;
- Finnish Institute of Occupational Health (occupational illnesses, exposure tests);
- Regional state administrative agencies (matters related to social welfare and health care);
- Population Register Centre (individual's basic details, family relations, places of residence and building information);
- Statistics Finland (to the extent that access is required to data covered by the Act on Establishing the Cause of Death 459/1973).

Figure 39 - Data holders and related data[441]


FINDATA
Data
National Institute for Health and Welfare (THL)
Cancer register
Infectious disease register
Hospital discharge register
National vaccination register
Medical birth register
Register of induced abortions
Register of congenital malformations
etc.
Population Register Centre (VRK)
Population data
Families
Residence
Building information
Social Insurance Institution (Kela)
Presciptions
Social benefits
Statistics Finland
Register on cause of death
National Supervisory Authority for Welfare and Health Valvira
Finnish Medicines Agency Fimea
Regional State Admin. Agencies (AVI)
FINDATA
Finnish Centre for Pensions (ETK)
Benefits and incomes
Diagnoses
Finnish Institute of Occupational Health
Occupational illness or exposure
Social and health care operating units
Data saved in Kanta services (2021)
E-services for citizens and health care and social welfare professionals

4.1.2.1.4 Financing

Findata's 2020 budget, funded from public funds, is **EUR 5.2 million** – in addition to its Isaacus precursor project that had a budget of EUR 14 million (of which an approximate three quarters were linked to the establishment of Findata). This budget is likely to decrease in time, as Findata's operations gain in maturity and as fee-based revenues increase.

These fees are of a triple nature and consist of:

[441] Source: Implementation of the national Social and Health Data permit authority Findata. Johanna Seppänen, PhD, Director

- A **fixed fee of EUR 1,000** for a data permit or request decision (including for EEA re-users), or a EUR 350 fee to amend an existing permit; and
- A **processing** fee (for combining, pre-processing, pseudonymisation and anonymisation) of **EUR 115 per hour worked**;
- A fixed fee for the remote access environment package that ranges between EUR 2,250 and EUR 8,500 (excluding customisation of the environment for an additional fee of EUR 115/hour worked).

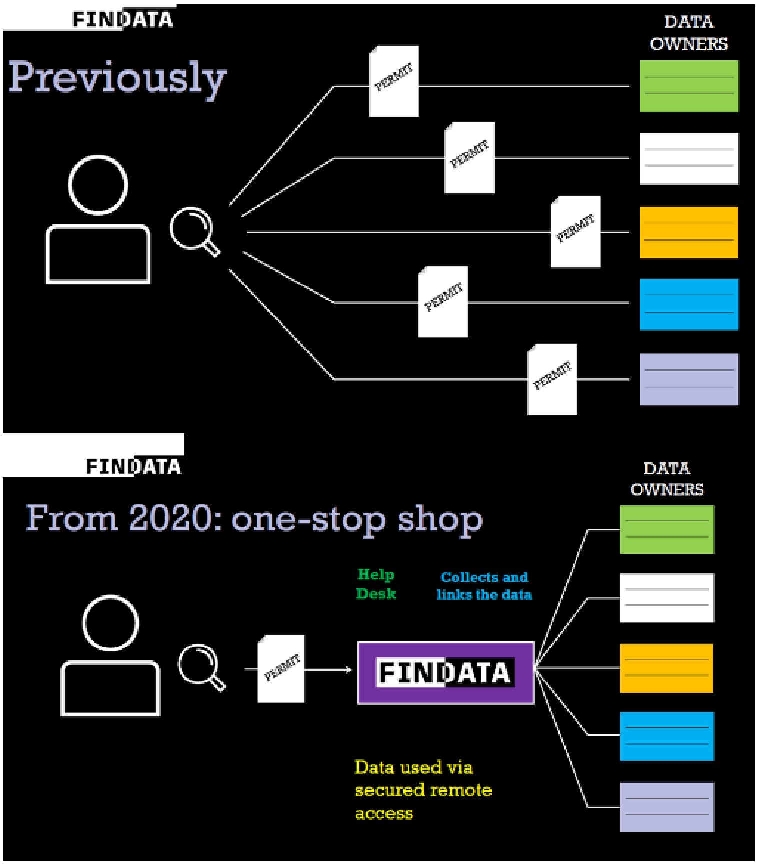
These fees are intended, on the long-term, to cover Findata's cost but not to make a profit.

Currently, Findata employs 15 people and aims to hire an additional 10 on the long-term. A rough estimate of costs for one FTE is EUR 75,000 per year for Findata (i.e. currently EUR 1.9 million).

4.1.2.1.5 Architecture

The data the use of which is subject to the rights of others is handled in a safe and secure environment. Access to data is controlled, and only the results of the analytics can be used externally.

Figure 40 - The evolution of the data provisioning[442]



As described in figure 2, previously, obtaining the permits and data was a difficult and expensive process in terms of time. Indeed the user needed to approach each authority and data source separately. Today instead of having to apply for separate permissions from several different data owners, a single central operator/ service operator (Findata) issues and grants research permits[443], including ethical evaluation. After granting the permission to use data, the service operator collects relevant data from different registers and edits,

[442] Source: Implementation of the national Social and Health Data permit authority Findata. Johanna Seppänen, PhD, Director

[443] Findata's permit service; https://lupa.findata.csc.fi/

combines and anonymises the data before distributing it to the user. As also depicted in the figure 3, Findata ensures that data handling and transfer of data occurs in a secure environment and that the process meets all the requirements defined by Finnish law. To do this Findata uses a data management system including a secure remote user environment with associated tools. Additionally, a data description system serves as a centralised place for saving the metadata of available materials. The solution includes, among other things, a metadata editor for editing and updating data descriptions.

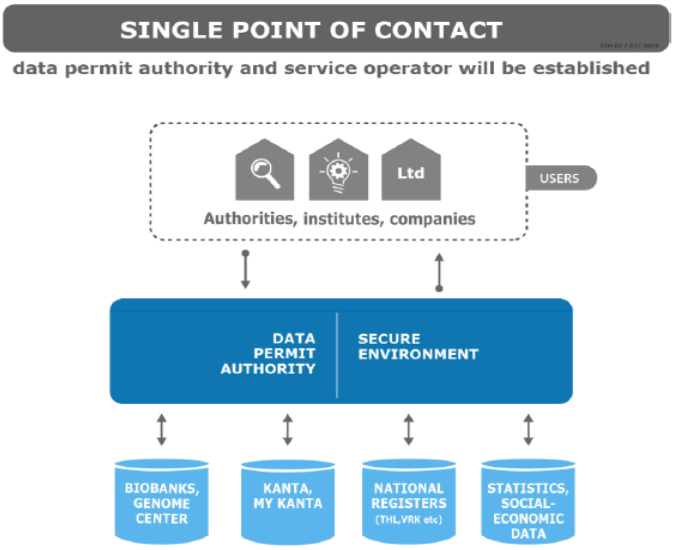

SINGLE POINT OF CONTACT
STM 07 / 03/ 2018
data permit authority and service operator will be established
Ltd
Authorities, institutes, companies
USERS
DATA PERMIT AUTHORITY
SECURE ENVIRONMENT
BIOBANKS, GENOME CENTER
KANTA, MY KANTA
NATIONAL REGISTERS (THL,VRK etc)
STATISTICS, SOCIAL-ECONOMIC DATA

Figure 41 -Access to data[444]

There are two different levels of data and different ways to access related datasets:

- **Individual level data.** The data of this level can be used for scientific research, statistics, education, authorities' steering, supervision, planning and forecasting. This data is available in a remote access environment for a set period. The data has been anonymised or pseudonymised. A data utilisation plan is required for access to data sets.
- **Statistical level data.** The data of this level can be used for the aforementioned purposes and, in addition, for development and innovation and knowledge management. This kind of data are directly delivered to customers.

Health and social data are stored in various national and local databanks. There is a large variety of different kinds of patient record, well-being, social wellness and other data available. The usage of a unique national person ID-number makes it possible to combine personal records.

The following architecture represents the databases in scope and two types of data lakes:

- Local data lakes: County hospitals, local social and health care providers etc. have enormous data in various systems. In many places, the data is now gathered into data lakes.
- National data lakes: Social care data, Patient data and Prescription data, Personal Health Record and social data are stored in national data lakes.

[444] Source: Finland – Most advanced ecosystem for healthcare innovation. Nora Kaarela, Head of Industry, Health & Wellbeing, Invest in Finland, Business Finland.

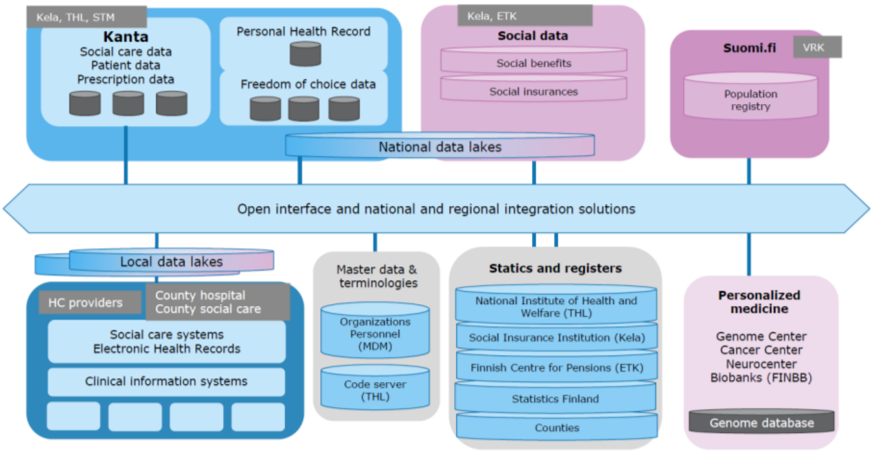


Figure 42 - Data sources and data lakes[445]

4.1.2.2 RatSWD

4.1.2.2.1 Introduction

The **German Data Forum (*Rat für Sozial- und WirtschaftsDaten*)** is a **public advisory council to the German federal government** and was founded in 2004. The RatSWD aims at sustainably **improving the research data infrastructure** that underlies empirical research and at contributing to the international competitiveness of said research.

It is made up of an **independent body of researchers and representatives of data holders**, and acts as an institution of exchange and of mediation between the interests of science and data producers. As such, it is an important platform for **communication and coordination**.

Although RatSWD itself does not make data available to re-users, it is an **intermediary responsible for the accreditation of Germany's Research Data Centres** (RDCs), which act as data holders and sometimes also as data re-users for research purposes. It coordinates these RDCs via a **Standing Committee Research Data Infrastructure (FDI Committee)** established in 2009.

4.1.2.2.2 Governance

The RatSWD was established by the **Federal Ministry of Education and Research** in 2004. It operates under rules of procedure determined by the Federal Ministry. These **rules of procedure** govern operations, tasks and competencies of the chair and the business office. The rules of procedure can be changed with a two-third majority of the members of the RatSWD and the consent of the Federal Ministry responsible for research.

The RatSWD is evaluated by the **German Council of Science and Humanities** – the last such evaluation revealed that the RatSWD succeeded in opening up and improving access to data and in creating synergies between scientific community and data holders.

The RatSWD consists of **16 members**. Of these, **8 are representatives from Germany's research community**, and are elected at the Conference for Social and Economic Data, held every 3 years. The remaining **8 members originate from data holders**, specifically from:

- The Federal Statistical Office;
- A (State) Statistical Office of the Länder;
- The IAB Institute for Employment Research of the German Federal Employment Agency;

[445] Source: Secondary Use of Health and Social Data in Finland, Joni Komulainen Ministrial adviser Master of Laws.

- An institution from the German social security system;
- An institution from the area of official health data;
- An institution from the area of official financial data; and
- An institution from the area of science-based data production.

The **members from data holders are appointed by the Federal Ministry for Education and Research** based on proposals made by their respective institutions in accordance with the Law for the Composition of Federal Committees (*Bundesgremienbesetzungsgesetz*) and on request from the Federal Ministry for Education and Research.

In addition to these members, up to two elected representatives from the Standing Committee Research Data Infrastructure have a permanent right to attend RatSWD meetings. Likewise, two representatives of the Federal Government, as well as two representatives of the Länder have a right to attend and are entitled to bring forward motions. Further, the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) has a permanent right to attend.

4.1.2.2.3 Operations

The RatSWD performs an advisory function, initiates new development and secures quality with regards to standardisation and data quality, and to development of RDCs and data service centres.

Its core tasks are the following:

- To issue recommendations on further improving the data infrastructure, specifically:
- Recommendations on how to secure and further improve data access, particularly by establishing and evaluating research data centers and data service centers according to a set of clear standards;
- Recommendations on how to improve data use through the provision of scientific and statistical data (research data portal; metadata) and appropriate documentation;
- Recommendations on research topics and research tasks pertaining to the conceptual development of data infrastructures on the national, European and international level;
- Recommendations on how to optimise the production and provision of research-relevant data;
- To advise science and policy, specifically:
- Advising the Federal Ministry for Research and the Länder governments on the development of the research-based data infrastructure;
- Advising public and private data producers;
- Advising data producers that are institutionally unaffiliated with independent scientific research on how to receive certification as a scientific research institution (certification);
- To monitor legal and technological developments, specifically:
- Monitoring national and legal developments in data provision;
- Monitoring technological developments, e.g. virtual research environments; and
- To organise and host the Conference for Social and Economic Data every three years.

The Conference serves as a **platform for discussing** topics pertaining to empirical social, behavioural and economic data as well as process-produced or survey-based data production. Participation is open to all interested researchers. The German Data Forum furthermore hosts **colloquia, panel discussions and workshops** that foster an ongoing exchange between researchers and data producers.

The RatSWD currently boasts 11 working groups divided into 3 themes.[446] These are:

- New data sources and data access for researchers
- Access to Big Data;

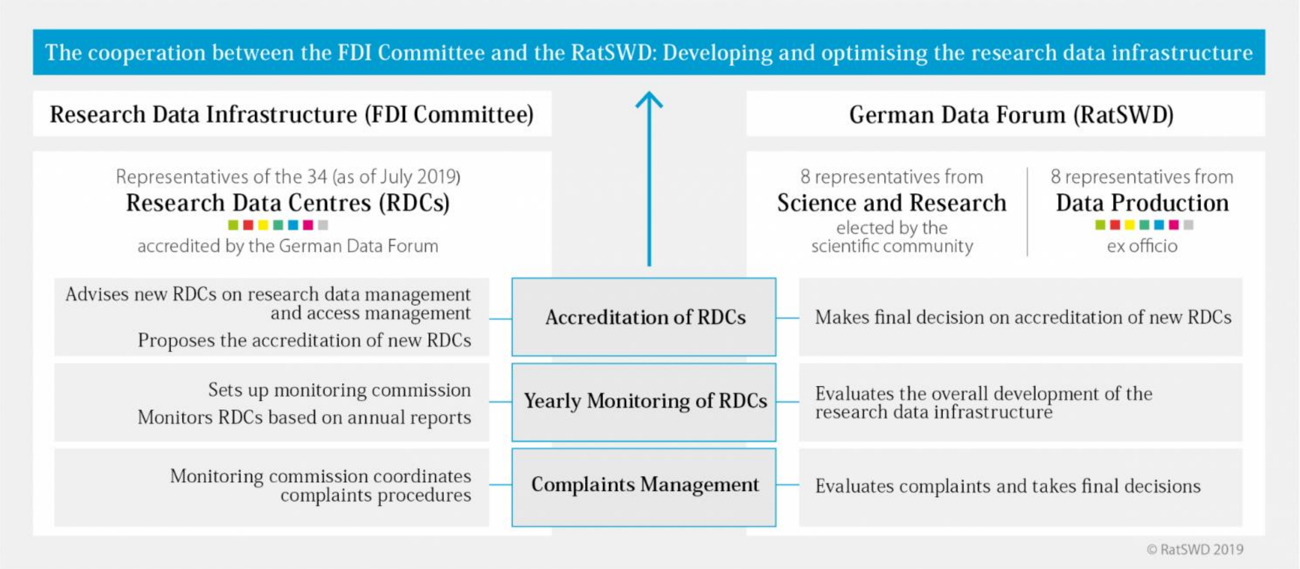
[446] https://www.ratswd.de/en/activities/working-groups

- Tax and Wealth Data;
- Further Development of Crime Statistics and Legal Data;
- Archiving and Access to Qualitative Data;
- Data Collection With New Information Technology; and
- Remote Access to Data from Official Statistics Agencies.
- Further development of the research data infrastructure
- Common guidelines in research data centres;
- Improving access to existing data in research data centres;
- Decentral archiving structure at research data centres; and
- Skills development in research data centres (RDCs).
- Advising of legislators and policy-makers
- The social sciences in roadmap processes.

4.1.2.2.4 FDI Committee

The RatSWD coordinates 38 accredited RDCs, which work together in a Standing Committee Research Data Infrastructure (FDI Committee) established in 2009.[447] The FDI committee produces recommendations for the RatSWD, and coordinates cooperation among RDCs with a view to continuously improve the research data infrastructure and to facilitating data access for researchers. The Committee also serves as a peer-review/peer-pressure mechanism to ensure the adherence by RDCs to the quality criteria these must uphold; as a complaints body to reach amicable settlements. It also attempts to standardise usage conditions such as charging practices among RDCs.

Figure 43 - The FDI Committee within the RatSWD


The cooperation between the FDI Committee and the RatSWD: Developing and optimising the research data infrastructure
Research Data Infrastructure (FDI Committee)
German Data Forum (RatSWD)
Representatives of the 34 (as of July 2019)
Research Data Centres (RDCs)
accredited by the German Data Forum
8 representatives from
Science and Research
elected by the scientific community
8 representatives from
Data Production
ex officio
Advises new RDCs on research data management and access management
Proposes the accreditation of new RDCs
Accreditation of RDCs
Makes final decision on accreditation of new RDCs
Sets up monitoring commission
Monitors RDCs based on annual reports
Yearly Monitoring of RDCs
Evaluates the overall development of the research data infrastructure
Monitoring commission coordinates complaints procedures
Complaints Management
Evaluates complaints and takes final decisions
© RatSWD 2019

The FDI Committee lists a number of key activities for 2017-2020.[448] In particular, the Committee aims for:

- the harmonisation of processes in the RDCs (e.g. data use contracts);
- the expansion of access to research data from RDCs, for example, via guest researcher work stations of other RDCs;
- the opening up of the existing data infrastructure for scientific and official data producers; and
- the advancement of skill development training for the RDCs' employees and data users.

[447] https://www.ratswd.de/en/data-infrastructure/fdi
[448] https://www.ratswd.de/en/data-infrastructure/fdi

In order to be accredited by the RatSWD, the RDCs need to meet three core criteria.[449] They must:

- Provide at least one data access path;
- Provide sufficient data documentation; and
- Ensure the long-term availability of the data.

In addition, RDCs must be "fully operational", which is defined as having been in operation for at least six months, and having at least three external data re-users. Lastly, initial accreditation requires compliance with additional information criteria used to assess the scope and quality of the RDC's operation. These information criteria relate to the:

- Scope and development of the social, behavioural, and economic data provided
- Method for timely data provision
- Provision of tools
- Quality assurance of datasets
- Data protection safeguards in due consideration of the interests of researchers
- Service concept
- Single entity comprising institution and research data centre
- Provision of all datasets relevant to research
- Overlap and distinct features compared to existing RDCs
- Research activities
- Multiple provision of the same data (multiple hosting, not hosting at multiple sites)
- Time to process applications
- Staff
- Infrastructure development

All accredited RDCs contribute to annual reporting by completing a questionnaire. As with accreditation, the questionnaire is based on the mandatory and information criteria. The FDI Committee elects a monitoring commission for a three-year term concurrent with the German Data Forum's (RatSWD) appointment period. The main task of the monitoring commission is to collect and assess the research data centres' annual reports. Moreover, the commission handles complaints regarding RDC accreditation criteria and provisional accreditations.[450]

4.1.2.2.5 Financing

RatSWD is funded by a grant from the Federal Ministry of Education and Research.[451] In its early years, the annual cost of running RatSWD ranged between EUR 200,000 and 300,000. This has significantly increased however: RatSWD regularly holds meeting with representatives of the accredited RDCs. These have grown in number, and RatSWD is responsible for their travel.

Overall, RatSWD's current **annual budget amounts to EUR 900,000**, of which:

- Half (roughly EUR 450,000) corresponds to human resources;
- EUR 300,000 corresponds to expenses linked to RDCs meetings;
- EUR 75,000 go to renting RatSWD's premises and to technical support;
- EUR 30,000 go to translation and ad-hoc legal advice on some publications; and
- EUR 25,000 cover travel expenses of RatSWD's staff.

[449] https://www.ratswd.de/en/info/accreditation
[450] https://www.ratswd.de/en/info/monitoring-and-complaints-management
[451] Figures, particularly regarding the RatSWD's budget, will be collected via interviews.

Of the 30 RDCs accredited up until 2016, 21 report not charging any fees for providing access to data (see Fig. 6). The fees charged by the nine other RDCs are low (in the two-digit euro realm) and mostly used to cover the costs for media and contracting.[452] RatSWD itself is a free service.

4.1.2.2.6 Architecture

Partly due to data sensitivity and respective legal regulations, the data offered by the RDCs must remain at their respective data producing institutions. Therefore, the decentralised structure of the research data infrastructure is a tried and tested way to satisfy the demands of data producers, data users in science and research, and data protection.

RDCs' main responsibilities are:

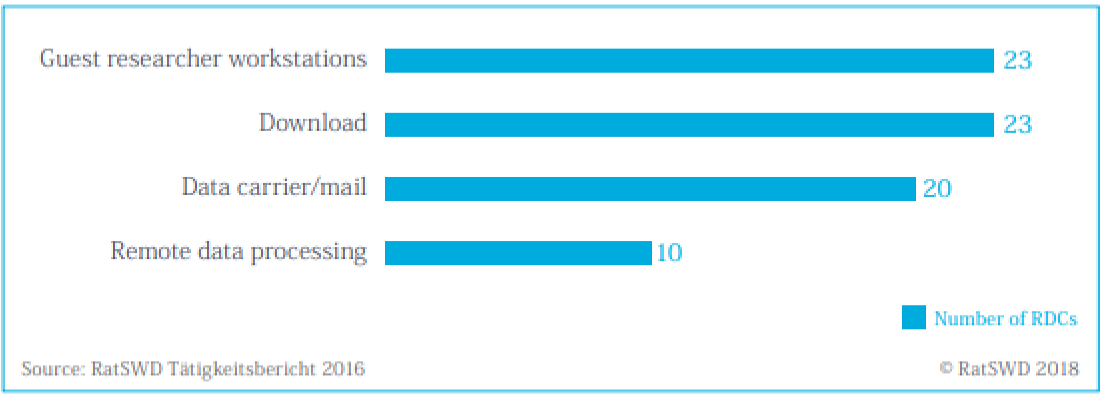
- **Providing researchers with user-friendly, transparent, and high-quality access to data**. So far, this data has concerned mostly microdata that can be analysed statistically. The data are collected as part of official statistics, administrative operations, research projects, or scientific survey programmes. In their capacity as mediators, the RDCs help improve cooperation between data users and data producers.
- Ensuring that data users comply with federal data protection policies and, if applicable, with policies specific to individual research areas, by taking appropriate technical and organisational measures. Depending on the level of anonymisation (see info box 1), datasets are offered for off-site use (via download or mail order) in the form of Scientific Use Files (SUF), Public Use Files (PUF), or Campus Files (CF). Moreover, the generation of synthetic data can be an option to support research needs. To facilitate access to highly sensitive microdata, the RDCs offer the option of on-site use. In this case, users can access the data at a guest researcher workstation on the premises of the RDC.
- **Ensuring equal treatment of all data users by means of transparent and standardised application and access policies**. Incoming applications are not assessed with regard to the content of the proposed research; they are only reviewed in terms of their compliance with contractual or data protection policies.
- **Creating easy-to-analyse data products featuring quality-assured, standard-compliant metadata and comprehensive documentation**. The RDCs present information on their respective data services via their websites, in data and method reports, at scientific events, or in individual advising sessions.
- **Conducting independent research using the data they offer**. This helps ensure that each RDC has strong expertise regarding the data and their quality. At the same time, ongoing scientific discussions about methods and contents can inform the advising services provided to data users. Research activities by RDC staff do not involve any exclusive access to data products

Depending on the content or the unit of observation, making such data available is subject to various legal requirements, most notably by the EU General Data Protection Regulation and the German Federal Data Protection Act but also state data protection legislation, the German Social Code, and the Federal Statistics Act.

In 2016, the RDCs offered a total of 3,214 datasets. The RDCs provide a wide range of access paths to their data. Generally, there are two distinct basic ways of accessing the data: on-site (i.e. on the physical premises of the RDC) and off-site (i.e. outside the RDC). Several RDCs offer multiple paths of access.[453]

[452] https://www.ratswd.de/dl/RatSWD_Output1.6_QualityMgmt.pdf
[453] https://www.ratswd.de/dl/RatSWD_Output1.6_QualityMgmt.pdf

Figure 44 – Data access paths in different RDCs



Note: These figures are from 2016 and only include 30 RDCs.

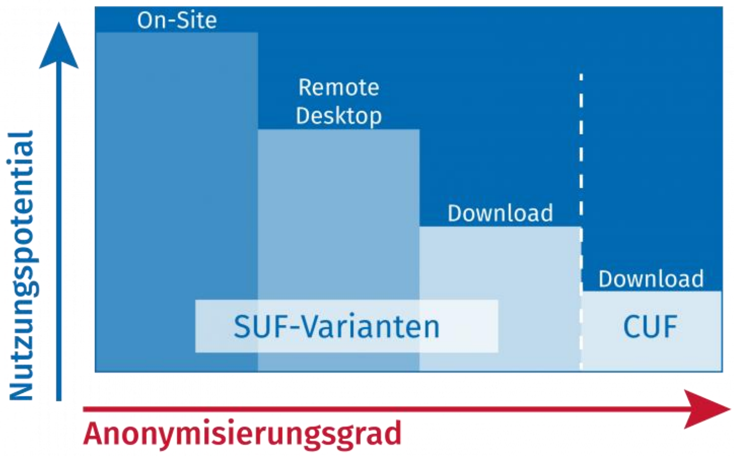
The RatSWD provides a search engine for data provided by most of the 38 RDCs.[454] This engine is da|ra, the registration agency for social and economic data that enables holders to register their data in order for it to be stored and easily identifiable on the long-term by researchers.[455] Still, one stakeholder pointed to the difficulty to ensure good quality metadata, especially with respect to historic data.

In addition to the Forschungsdatenzentrum, the Research Data Centre of the German federal Employment Agency is one other accredited Research Data Centre.[456] It provides researchers with micro-data on social security and employment for research purposes. This data may be accessed using Scientific Use Files,[457] remotely, and on-site. In addition, it offers advice on data selection, access, handling, analysis potential, scope and validity; it regularly updates datasets; produces research; and organises workshops and conferences. Access to data requires an application and the conclusion of a user agreement.

Likewise, the Research Data Centre for Higher Education Research and Science Studies makes available quantitative and qualitative data from universities and research institutes to researchers, following an application process.[458] Information collected as part of this process includes the focus on the study, the data collection methods and types of data used, and the duration and status of the study. Access to data is done either via Scientific Use Files or Campus Use Files (anonymous data sets for teaching and exercise purposes at universities). Access can be on-site in a controlled environment, remotely via a virtual desktop, or through downloading highly anonymised datasets (see figure below).

[454] https://www.ratswd.de/en/researchdata/search
[455] http://www.da-ra.de/en/home/
[456] https://fdz.iab.de/en.aspx
[457] These are de facto anonymized data records for scientific purposes.
[458] https://www.fdz.dzhw.eu/de

Figure 45 - Usage potential (*y* axis) and level of anonymisation (*x* axis) of microdata at the RDC for Higher Education Research and Science Studies



4.1.2.3 National Safe Haven (Scotland)

4.1.2.3.1 Introduction

The UK National Health Service (NHS) **National Services Scotland (NSS) national safe haven service** – allows data from electronic records to be used to support research when it is not practicable to obtain individual patient consent, while protecting patient identity and privacy. It provides **secure file transfer and submission services to data providers** and additional services (e.g. analytics platforms) to researchers.

Data safe havens provide a **secure and safe environment**, supported by trained staff and agreed processes to facilitate statistics and research work on **sensitive data**, including medical data (e.g. patient records and MRI images) and social data (e.g. census, government or police data). Health data can be processed and linked with other health data (and/or non-health related data) and made available in a de-identified form for analysis. Save havens serve as safeguards for confidential information which is being used for research purposes. Any researchers applying for access to health data must adhere to the Safe Haven principles,[459] namely:

- The ultimate aim of information sharing is to provide care;
- Citizens should have the choice about the use of their data; and
- Dialogue with the public should be maintained.

Safe Havens in Scotland were established as part of a national need for delivering research excellence and the need for rapid access to high-quality health data for research purposes. They were developed in line with the SHIP blueprint which outlined a programme for a Scotland-wide research platform for the collation, management, dissemination and analysis of anonymised Electronic Patient Records (EPRs). The agreed principles and standards to which the Safe Havens are required to operate are set out in the Safe Haven Charter.[460]

[459] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven. NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens. EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service

[460] NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens.

NHS Scotland allows research using routinely collected, unconsented patient data. Additionally, these data can be linked to social data such as education. The research that this enables can have an enormous public benefit but the use of this data must be managed very carefully to safeguard privacy and maintain public trust and support.[461]

This case study details how Scotland's National Safe Haven operates in practice.

4.1.2.3.2 Governance

Safe havens are subject to strong national information governance policies.[462] The Scottish National Safe Haven within Information Services Division (ISD) is an expert unit within the Common Services Agency (known as NHS National Services Scotland).[463] The National Safe Haven is part of the Scottish Informatics Linkage Collaboration (SILC) that also includes the electronic Data Research and Innovation Service (eDRIS) and the National Records of Scotland indexing service. SILC facilitates linkage for research and statistical activities across many sectors including the NHS.

The NHS Research Scotland (NRS) nodes received funding, through NRS infrastructure allocations from the Chief Scientist Office, to help establish the Safe Havens.[464] Together, the National Safe Haven within SILC and the four NRS Safe Havens have formed a federated network of Safe Havens in order to work collaboratively to support health informatics research across Scotland. All the Safe Havens have the individual responsibility to operate at all times in full compliance with all relevant codes of practice, legislation, statutory orders and in accordance with current good professional practice. Each Safe Haven may also work independently to provide advice and assistance to researchers as well as secure environments to enable health informatics research on the pseudonymised research datasets they create. The governance of SILC involves representatives of the Safe Havens.[465]

EPCC at the University of Edinburgh, under contract with the NHS is now the operator of the new NHS National Services Scotland (NSS) national safe haven and responsible also for building, supporting, maintaining and hosting it, in collaboration with the Farr Institute of Health Informatics Research which provides the infrastructure. EPCC continues to develop the infrastructure and software to further enhance the service.[466]

4.1.2.3.3 Operations

The new NHS National Services Scotland (NSS) National Safe Haven service implementation work started in September 2015 with the live service rolled out during December and January 2016. Now fully operational, the safe haven is both physical and remote. It offers a secure file transfer and submission service for data providers and a range of access methods and analytics platforms and tools for researchers.[467] The Scottish

[461] EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

[462] EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service

[463] Scottish Government, A Charter for Safe Havens in Scotland. See: https://www.abdn.ac.uk/iahs/documents/00489000.pdf

[464] Scottish Government, Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. See: https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/

[465] *Ibid*.

[466] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven.
EUDAT Slides, Data safe havens: a future EOSC service? See: https://www.slideshare.net/EUDAT/data-safe-havens-a-future-eosc-service
EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

[467] EPCC, NHS National Services Scotland (NSS) national safe haven. See: https://www.epcc.ed.ac.uk/projects-portfolio/nhs-national-services-scotland-nss-national-safe-haven.

national safe haven, as all Data Safe Havens, provides a platform for the use of NHS electronic data in research feasibility, delivery and pharmacovigilance offering:

- A secure environment, with trained staff operating to agreed principles and standards;
- A federated network across Scotland working collaboratively to deliver and develop health informatics capability;
- Access to a wide range of anonymised datasets (including national datasets through to specialised local datasets);
- Robust governance procedures to protect the confidentiality of the data;
- Single costing and contracting; and
- Daily updates.[468]

Working to agreed principles and standards, Safe Havens in Scotland provide access to health data and services to enable research while protecting the confidentiality of the data. Data remains under the control of the NHS and complies with legislation and NHS policies.[469]

The National Safe Haven has proven very successful and has supported more than 200 research projects over the two years it has been running. EPCC is currently working with the Health Informatics Centre at the University of Dundee to provide researchers with access to the Scottish NHS imaging data (X-ray, CT, MRI, ultrasound etc.) dating from 2010 onwards. Such data offers tremendous opportunities for a wide variety of research including examining early/preclinical diagnosis, disease progression, personalised medicine, genotype-phenotype associations, and the development of novel computer vision and machine learning algorithms. Additionally, the ability to link this image data with patient outcome data is unique to Scotland and offers even greater potential for world leading research that will have a major contribution to the future health of the nation.[470]

eDRIS was established as a specific ISD function within NSS and provides a single point of contact for advice on research project design and development as well as access via the National Safe Haven to a wide range of national datasets. Given the well-established close working relationship with ISD (the Data Controller of national NHS Scotland datasets), the National Safe Haven may be best placed to take the lead when research requires the processing and linkage of national datasets.[471]

4.1.2.3.4 Financing

The NRS nodes received funding through NRS infrastructure allocations from the Chief Scientist Office to help establish Safe Havens.[472] In particular, Chief Scientist Office (CSO) has invested significantly in Safe Havens in each of the four NRS nodes, providing a platform for the use of NHS electronic data in research feasibility, delivery and pharmacovigilance.[473]

There is no particular formula used to allocate funding to the National Safe Haven (NHS NSS) but this is being done based on an assessment of need.[474] According to the Chief Scientist Office Outturn Summary for 2018-19, the amount of £150,000 was allocated to ISD Safe Haven [NHS National Services Scotland] in

[468] NHS Scotland, Data Safe Haven. See: https://www.nhsresearchscotland.org.uk/research-in-scotland/data/safe-havens

[469] *Ibid*.

[470] EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

[471] Scottish Government, Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. See: https://www.gov.scot/publications/charter-safe-havens-scotland-handling-unconsented-data-national-health-service-patient-records-support-research-statistics/pages/4/

[472] *Ibid.*

[473] Chief Scientist Office, Initiatives. See: https://www.cso.scot.nhs.uk/about/initiatives/

[474] Scottish Government, Formula used to allocate funding to NHS Boards: FOI release. See: https://www.gov.scot/publications/foi-18-02515/

order to support a national data safe haven for the safe and secure provision and linkage of de-identified data from national health data sets for approved research. Additionally, CSO invests around GBP 40 million each year to support NHS Research Scotland to conduct research. The purpose of the principal funding streams is to support, among others, NRS infrastructure including NRS biorepositories and data safe havens.[475]

4.1.2.3.5 Architecture

The Scottish National Safe Haven is a secure environment where the project data is uploaded and accessed. It offers a high powered computing service, secure analytic environment, secure file transfer, and a range of analytic software including SPSS, STATA, SAS and R. The IT infrastructure is provided by the EPCC at Edinburgh University.

To use the unconsented NHS data, researchers must apply to the Public Benefit and Privacy Panel specifying the data that is required and the purpose of the research. If granted permission, the data will be selected, anonymised and linked (a process called pseudo-anonymisation) before being placed within the National Services Scotland (NSS) National Safe Haven. Researchers must process their data from within this National Safe Haven infrastructure.[476]

A data controller may require that their data is only accessed through a secure access point to ensure data security. A secure access point is a dedicated computer in a physically secure area where no external devices can be used or connected. The secure access point does not connect to the internet nor can it be accessed remotely. In case of work in a secure access point, a username and password for the linked data files will be given. The location of the secure access points may depend on the data to be accessed, being either at the offices of BioQuarter at the Royal Infirmary of Edinburgh or at selected universities across Scotland.

In some cases, data controllers may allow the researchers remote access to the data. This will be via a VPN (virtual private network). To remotely access the safe haven there is a 2 factor authentication process, the first part of which will be receipt of an access code via mobile phone. To access the National Safe Haven, researchers must use remote desktop software and log in using a high-security protocol. Researchers are then able to use the remote desktop session to access, process and analyse the data they have requested. The remote computer is installed with several statistics packages that the researcher can use. Crucially, the remote computer offers no access to the Internet neither to receive nor send data. Researchers can request files to be transferred into, or out of, the Safe Haven but such requests are subject to a manual verification process to ensure privacy is never breached.[477]

Outputs from the analyses are only released for the agreed purpose of the research. Data cannot be used in any other way. Outputs cannot be released until the Research Co-ordinator has assessed them for statistical disclosure control in line with the data controller's requirements specified for the study. The objective of this is to ensure that an output does not contain information which could be used either on its own or in conjunction with other data to breach an individual's privacy. Data is held in archive for a specific

[475] Chief Scientist office & Scottish Government, Chief Scientist Office Outturn Summary 2018-19. See: https://www.cso.scot.nhs.uk/wp-content/uploads/CSO1819OTsummary.pdf

[476] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/
EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

[477] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/
EPCC, The National Safe Haven for research using unconsented NHS data. See: http://www.epcc.ed.ac.uk/blog/2017/11/21/national-safe-haven-research-using-unconsented-nhs-data

period set by the data controller, after which the study data are being deleted (upon prior confirmation). It is not possible to restore study data, as deletion is permanent.[478]

4.1.2.4 Statistics Denmark

4.1.2.4.1 Introduction

Statistics Denmark is the central authority on Danish statistics. Its mission is to collect, compile and publish impartial statistics on the Danish society, as a basis for democracy and the economy, covering a range of subjects, including:[479]

- Population and elections (including population and population projections; immigrants and their descendants; births; deaths and life expectancy; households, families and children; marriages and divorces; migrations; names; elections);
- Labour, income and wealth (including labour force participation; employment; unemployment; commuting; absence and work stoppages; earnings and labour costs; income; wealth and debt);
- Living conditions (including gender equality; quality of life; survey on living conditions; housing; health; childcare; persons receiving public benefits; social conditions; traffic accidents; criminal offences);
- Education and knowledge (including population by status of education; full-time education; courses and adult education; educational transitions; research, development and innovation; information society);
- Business Sector in general (including enterprises in general; enterprise development; accounts; globalisation; organic production and trade; tendency surveys); and
- Particular business sectors (including agriculture, horticulture and forestry; fishery and aquaculture; manufacturing industries; construction; transport; distributive trades; tourism; services sector).

This case study details how Statistics Denmark operates in practice.

4.1.2.4.2 Governance

Statistics Denmark is a state institution under the Danish Ministry of Social Affairs and the Interior. The executive board of Statistics Denmark is composed of the Director General and five Directors. The head of the management office also attends the weekly management meetings.

Statistics Denmark is managed by a supervisory board with the National Statistician as chair, and it includes seven other members. The supervisory board establishes its own rules of procedure and appoints one of its members as vice-chair. By law, the supervisory board chaired by the National Statistician is responsible for the following:

- The overall strategy and financial management of Statistics Denmark, while the National Statistician has the sole responsibility for defining the professional criteria for development, collection, compilation and dissemination of Statistics Denmark's statistical output.
- The professional independence of the official statistics and of the institution of Statistics Denmark.
- Consideration and decisions in matters of interest to the strategic management of the institution, including work programme, statistical programme and budget. Decisions as to the extent and ways of collecting data from the business community, including for the purpose of implementing EU and

[478] Public Health Scotland, Use of the National Safe Haven. See: https://www.isdscotland.org/products-and-services/edris/use-of-the-national-safe-haven/
[479] Statistics Denmark, About us. See: https://www.dst.dk/en/OmDS

national legislation. In this way, the supervisory board is responsible for the reporting task imposed by Statistics Denmark on the business community.

- Via the work programme and statistical programme, decisions about the data that public authorities and institutions must submit to Statistics Denmark.[480]

4.1.2.4.3 Operations

Authorisations to access de-identified microdata may be granted to researchers pre-approved by Statistics Denmark. These can be from public sector research organisations (such as ministries or universities) and from private sector non-profit foundations, NGOs and consultancies. Foreign researches cannot obtain access, unless via a Danish organisation.

The Division of Research Services is the one responsible for the provision of statistical microdata for research purposes. Annually, there are around 2,000 applications for access to de-identified data – of which 75% originate from the public sector, and 25% from the private sector.

The main tasks that Statistics Denmark perform with regards to microdata are:

- Examining and assessing applications to access data;
- De-identifying the data;
- Standardising past data (i.e. data dating back decades) to make it usable in combination with recent data;
- Fetching the requested data and merging different datasets; and
- Answering questions from reusers on specific datasets.

This is in addition to the back-office work, including research and development and coordination with other data holders.

4.1.2.4.4 Financing

Statistics Denmark's overall annual budget is DKK 35 million (i.e. EUR 4,7 million), of which DKK 7 million (EUR 940,000) is public funding from the Danish Ministry of Science. The remaining amount is fetched via user fees, and the share of budget covered by user fees is increasing.

Statistics Denmark applies different fees depending on the status of the applicant: public sector researchers pay a lower amount than researchers from the private sector, and this difference corresponds to the DKK 7 million of public funding. In other words, these public funds are used to subsidise public sector research.

Of the total DKK 35 million budget, a tenth (DKK 3,5 million, or EUR 470,000) covers the cost of maintaining Statistics Denmark's secure data processing environment.

4.1.2.4.5 Architecture

Microdata is not handed over to researchers, but rather accessed via a research server at Statistics Denmark. It is separate from the other networks and contains exclusively de-identified microdata for research purposes. Remote (and encrypted) access is possible via the internet, following an agreement with Statistics Denmark.

This agreement states that all work on the microdata must take place on the server and prohibits attempts to remove microdata from the server or to identify individuals or businesses. However, aggregated data may be removed from the server.

[480] Statistics Denmark, Supervisory Board. See: https://www.dst.dk/en/OmDS/organisation/bestyrelsen

All aggregated results from the researchers´ computer can be stored in a special file and such printouts are sent to the researchers by e-mail. This is a continuous process (every five minutes) and has shown to be quite effective. The advantage for Statistics Denmark is that all e-mails are logged at Statistics Denmark and checked by the Research Service Unit.

Several computer packets are available on the research server, such as SAS, SPSS, STATA, GAUSS and R. The programs are frequently updated with new versions.[481]

4.1.2.5 OpenSAFELY

4.1.2.5.1 Introduction

The OpenSAFELY project was developed in the United Kingdom in view of the global COVID-19 emergency, as a collaboration between the DataLab at the University of Oxford, the EHR group at London School of Hygiene and Tropical Medicine, TPP and other electronic health record software companies (who already manage NHS patients' records), working on behalf of NHS England and NHSX, with a growing list of broader collaborations including ICNARC. The team is composed of software developers, clinicians, and epidemiologists, all pooling diverse skills and knowledge to deliver high performance, highly secure and accurate health data analytics, using modern open software development techniques. The project has delivered because of its mixed skillset software developers, and "developer-epidemiologists", who can speak the same language as the technical teams within EHR system suppliers.[482]

This case study details how OpenSAFELY operates in practice.

4.1.2.5.2 Governance

The OpenSAFELY team works on behalf of NHS England, who is acting as Data Controller for the purposes of this project, while each EHR holder acts as Data Processor. The Secretary of State for Health issued NHS England/Improvement a notice under the Health Service (Control of Patient Information) Regulations 2002 3(4) which enabled NHS England to collect the data required from GP practices directly from their EHR vendor. All information governance for this urgent project is handled by NHS England.

The Data Protection Impact Assessments approving data flows and access approves linking GP data to outcomes data from the new NHS England and NHSX data store and other sources including COVID–19 Patient Notification System (CPNS) deaths data; Intensive Care National Audit & Research Centre (ICNARC) ITU admissions data; Second Generation Surveillance System (SGSS) PHE test data; Emergency Care Data Set (ECDS) Accident and Emergency (A&E) patient-level data; and Office for National Statistics (ONS) death data.[483]

4.1.2.5.3 Operations

OpenSAFELY deployed a new statistical analysis platform during the Covid-19 emergency to deliver urgent answers on key clinical and public health questions. It is successfully delivering analyses covering 40% of practices in the country and process data on over 24 million patients including their previous medical history, investigations, and current or past medications. Its first analysis identifies which patients are most at risk of death in hospital from COVID-19, with more accuracy than any previous analyses by an order of magnitude. The team has an extensive ongoing collaborations across the scientific community, running analyses to identify which patients are most at risk, and why, which treatments increase or decrease risk.

[481] Statistics Denmark, The Danish System for Access to Micro Data. See: https://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark_2014--pdf
[482] OpenSAFELY, Home. See: https://opensafely.org/#appendix-2-further-detail
[483] Ibid.

The project is also supporting modellers to understand, evaluate and predict the spread of the disease and pressure on NHS services, using hyperlocal real-world data.[484]

4.1.2.5.4 Financing

OpenSAFELY team has developed and deployed a fully functional platform in five weeks with no funding. With modest financial resources, the team will sustain, accelerate, and expand its work. OpenSAFELY currently has funding applications under review with UK National Institute for Health Research (NIHR) and with UK Research and Innovation (UKRI).[485]

4.1.2.5.5 Architecture

OpenSAFELY is a secure analytics platform for electronic health records in the NHS England, delivering pseudonymised analyses. Its analytic software is open for security review, scientific review, and re-use. OpenSAFELY uses a new model for enhanced security and timely access to data: it does not remove large volumes of potentially disclosive (i.e. allowing for identification) pseudonymised patient data from the secure environments managed by the electronic health record software company; instead, trusted EHR analysts can run large scale computation across near real-time pseudonymised patient records inside the data centre of the electronic health records software company. This pragmatic and secure approach has allowed to deliver the first analyses in just five weeks from project start. [486]

All data that carries any privacy risk (even a theoretical risk, and even when pseudonymised) remains within the secure data centre of the electronic health record holder, where it already resides. This also means that all activity is logged for independent review. All processing takes place in the same secure data centre, where the patients' electronic records were already stored. The only information to ever leave the data centre is summary tables (with low numbers suppressed) from statistical models. Within the data centre, all pseudonymised data is stored in a tiered system of increasingly less disclosive data stores tailored to each analysis.

All underlying software and research code is open to review for security profiling, scientific evaluation, and to re-use as open source tools improving science across the community. Overall, this approach is therefore highly secure, and supports high quality science: in contrast to working on intermittent "data extracts", OpenSAFELY's approach also ensures that the statistical models run across up-to-date records, which is vital during a global health emergency.

OpenSAFELY's approach to privacy and security exceeds standards for many other current EHR analysis projects. It severely restricts SQL query access to the "event-level" data, which would otherwise present the highest theoretical privacy risk. It then abstracts the key clinical features of each patient for each analysis into a "feature store" for statistical analysis: this summary data is perfectly matched to the needs of each project, but substantially less vulnerable to re-identification attacks; it is nonetheless still managed to the highest privacy standards, as if it were security-critical event-level data. All access to the secure platform is over highly secure VPN from specific IP addresses and MAC addresses for a very small number of highly trusted, named and experienced analysts whose activity is all fully logged. By building our analytics platform inside the originating EHR vendors' data centre, the team completely avoids transporting large raw primary care datasets which would otherwise present a substantial privacy risk, even when pseudonymised. [487]

The data linked include the full coded primary care record containing all previous medical history, test results, diagnoses, medications, treatments, and more; A&E attendance data; hospital death from Covid-19; ITU data; ONS death data including cause of death. The team is able to rapidly map and link new datasets where

[484] *Ibid.*
[485] *Ibid.*
[486] *Ibid.*
[487] *Ibid.*

required. This big data approach with an unusually large volume of primary care data is necessary to get sufficient statistical power to detect associations with specific medications and medical conditions as early as possible during the pandemic and thereby save lives by modifying patient, clinician, and population behaviour. All code for the platform is compliant with open standards and designed to be portable, so that it can run against any platform produced by the NHS in the future to securely store rich and linked primary and secondary care patient data. [488]

4.1.2.6 Statbel

4.1.2.6.1 Introduction

Statbel, the Belgian statistical office, collects, produces and disseminates reliable and relevant figures on the Belgian economy, society and territory to frame complex issues and dilemmas and provide some sort of support for the society. As an official statistical institution, it offers a huge range of figures in terms of economy, population and demography, labour market, poverty, agriculture, industry, services, real estate, transport and traffic, environment, etc. These figures are available at national, regional, provincial, municipal and even more detailed level, as well as within a European context.[489] Statbel also makes microdata available for research to public institution and research centres. This case study details how Statbel operates in practice.

4.1.2.6.2 Governance

The Belgian statistical office, Statbel, is part of the general management of the **Federal Public Service of Economy**.[490] It is one of the partners of the Institution for the National Accounts.

Provision of statistical microdata for research is done in accordance with article 15 of the law of 4 July 1962 on public statistics.

4.1.2.6.3 Operations

The following microdata is available to researchers from public bodies and research institutes:

- Road accidents;
- 2011 Census;
- Causes of death;
- Household budgets;
- Infant deaths (under one year old);
- Divorces;
- Workforce;
- Marriage;
- Birth; and
- Income and living standards.

These are pseudonymised. To obtain access, researchers must contact Statbel or a specific statistician via email, and submit a formal request for data using a dedicated form. These must be submitted along with a preliminary contract via email. The requester is required to indicate which measures it has in place with regards to IT and physical security of the data, as well as the measures taken to ensure compliance with the GDPR when applicable. The requester must also clearly and in a detailed way indicate how it will use the data requested, as well as who within their organisation will access them

[488] *Ibid.*

[489] Statbel, Who we are. See: https://statbel.fgov.be/en/about-statbel/who-we-are

[490] Statbel, Mission, Vision and Objectives. See: https://statbel.fgov.be/en/about-statbel/who-we-are/mission-vision-and-objectives

The request is then examined, and the final decision rests with Statbel's Director-General. Should the request be accepted, Statbel will draft and send a final contract that must be signed, and will deliver the data upon reception of the signed contract.

The result of the research, including analysis, studies and statistics produced, must be made available free of charge to Statbel, which may use them as it sees fit. Statbel may also forbid their publication.

4.1.2.6.4 Financing

The global budget of Statbel in 2020 was EUR 24 million, of which EUR 20.8 million or about 87% went to human resource costs. On December 1st 2019, Statbel employed 356 people, of which 40% were highly skilled employees, although there is a rising trend in this percentage. In addition to its own staff, there are another 350 surveyors conducting surveys among citizens for Statbel.[491] It must be noted that these numbers are for the whole of Statbel and thus not solely the budget and workforce needed for the handling of sensitive data: because the microdata for research service is not a service, organisationally speaking, the budget cannot be estimated. However, approximately 1 FTE works on providing microdata for researchers, according to Statbel.

The provision of microdata is fee-based: costs are intended to cover the cost of the production and delivery of the data (including the administrative work this entails), as well as to limit the number of requests to only those that cannot be fulfilled using the freely available aggregated datasets. This fixed fee of EUR 500 only applies to entities that are not government (national, regional, or local) entities, i.e. universities, research centres and others, but not ministries or cities.

4.1.2.6.5 Architecture

Due to historical reasons, Statbel's way of providing access to microdata for researchers differs from that of the other statistical offices examined in this study.

Indeed, after a request is approved and the necessary administrative steps taken, a Statbel statistician makes the requested anonymised microdata available in Statbel's data warehouse, and sends a copy of the file to the researcher's organisation (independent researchers may not receive microdata) via an SFTP connection. There are safeguards associated with this approach and specified in the data request form,[492] namely:

- The user may not obtain more copies of the data than are need for the objectives of the study;
- The user may not forward the data to third parties unless agreed by Statbel;
- The user may not use the data after the agreed timeline of the research conducted;
- Following the end of the research, the user must delete all copies of the data (or following the completion of the research objectives, whichever comes sooner);
- The user must ensure the data is used only by the personnel of the organisation they belong to;
- The data must remain on the servers of the organisation to which the user belongs; and
- That organisation must detail which security measures it has in place and demonstrate how it will securely handle the microdata.

According to Statbel, there have so far not been any issues or breaches with this approach, i.e. researchers always respect the contract.

[491] Statbel, 2020 Statistical programme. See: https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/ProgrammeStatistiqueStatbel2020_nl.pdf

[492] Statbel, Demande de micro données pour finalités statistiques. See: https://statbel.fgov.be/sites/default/files/files/documents/Formulaire%20de%20demande%20micro%20donn%C3%A9es%20(FR).docx

In certain, very rare (under once a year) cases, a researcher may request access to microdata where they may be an identification risk (where anonymity of respondents cannot be guaranteed). In these cases, the researcher must use the microdata in Statbel’s secure processing environment – i.e. the same approach taken by other statistical offices – and under the constant supervision of a statistician.

4.1.3 Establishing a certification framework for data intermediaries: One pagers

Digi.me

Intermediary Category	PIMS/PDS
Data Sharing Scenario	C2B
Type of data sharing	Personal Data *(however there is technical capacity also for industrial data)*
Year of Establishment	Founded in 2009 as a legal entity; operational in its current form since 2013
Stage	Growth
Country/-ies of Establishment	United Kingdom (HQ); France, Netherlands; Bosnia; United States; Australia
Profit/Non-Profit Driven	For profit
Number of Employees	50-60
Revenue/turnover	N/A
Funding	$30M Private Sources (including i.a. Swiss RE, Omidyar Network and many notable and other HNWs)
Business Model/Functionalities	Digi.me empowers the individual to share more & better data, to enable businesses to provide more & better value, with full privacy, security and consent. • Individuals benefit from the value in their data • Businesses get a complete view of their customer When using the digi.me app, the individual holds their own data and digi.me does not see, touch or hold user data. • Data encryption and normalization happens inside the app without digi.me ever being able to see or access user data. • Only the user has the credentials to access their digi.me library and must provide credentials directly to data sources. • Digi.me stores no user data. The user chooses their own location (e.g. dropbox, google drive, onedrive) where encrypted data is stored. When businesses/services request data from an individual they use the digi.me consent system and certificate for explicit and informed consent as per GDPR (certificate number goes into the API and the certificate is displayed to the individual). Upon individual's consent to data being transferred to the business/service, then digi.me charges the business/service $0.10 with a cap of $3 per individual per business/service per year to be paid by 'postal fee' on data transfer. Use of the digi.me app by the individual is free as the individual's cost to digi.me is negligible.
Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	• Horizontal client base (including i.a. health and wellbeing, finance, retail, insurance banking, government and IoT, research) • More than 700,000 users of digi.me app over time in 140 countries and 1,000s of sources of data

DAWEX

Intermediary Category	Data Marketplace & Data Exchange Solution Provider
Data Sharing Scenario	B2B
Type of data sharing	All types of data (industrial data, anonymized data, personal data)
Year of Establishment	2015
Stage	Late
Country/-ies of Establishment	France; US; Canada
Profit/Non-Profit Driven	For profit
Number of Employees	Approximately 50
Revenue/turnover	N/A
Funding	Approximately global funding of €18M from private sources and loans
Business Model/Functionalities	**Data Exchange Platform** technology that facilitates data sharing, data sourcing, data commercialization, data orchestrating, providing in particular the following functionalities: • Facilitate the connectivity between the parties: place where data providers and data users meet and allow cross-border and cross-sector data exchange; automatic matching of supply and demand; • Allow the parties to interact with each other and keep track of these interactions; • For data providers/holders: packaging and describing data offering which can be very diverse (e.g API based, combination of several files, one-time transaction or subscription, description of product and definition of terms and conditions for making data available, licencing terms and pricing) • For data users/buyers: easy search, discovery and filtering certain data products and offering; assessment of offering, negotiation of transaction terms and offering • For Data Exchange Platform orchestrator: fully featured administration console including various automation features for managing participants and stimulate activity, at scale. **Global data marketplace**, built upon Data Exchange Platform technology and providing, among others, the following features and services: • Data monetisation or free-of-charge data exchange • Full control given to data providers and data users over the terms of the exchange • Volume, variety and speed of data exchange • Configurable contract license as well as open data license supported • Data visualization algorithms for quick evaluation and promotion of data quality • Visualization of representative data samples • Recurring transaction services • Integrated payment processing • Administration and traceability services for the data providers an data users • Three subscription plans to access the service (Free, Business, Enterprise) meeting various usage levels and support requirements
Data volume	N/A
Architecture	Flexible architecture allowing users to choose centralized and/or decentralized models for exchanging data depending on the use case
Client Base/Use Cases	• Broad horizontal client base that includes more than 10,000 organizations in more than more than 50 countries and more than 20 industries (incl. i.a. automotive; agriculture; bank, air and space; insurance & financial services; energy; environment; health; banking; telecommunications; retail and consumer goods; public sector; tourism and sports; shipping and logistics) • Customized implementations of Data Exchange Platform technology (white label) target corporates, consortiums and governments (sectorial or regional scope)

International Data Spaces Association

Intermediary Category	Trusted third party providing a reference architecture
Data Sharing Scenario	B2B
Type of data sharing	Industrial data
Year of Establishment	2016
Stage	Growth
Country/-ies of Establishment	Germany
Profit/Non-Profit Driven	Non-profit
Number of Employees	20
Revenue/turnover	N/A
Members	117
Funding	Sources: public and private: Industrial Data Space was created in a research project funded by the German Federal Ministry of Education and Research (BMBF) involving multiple Fraunhofer institutes. It currently works with membership fees.
Business Model/Functionalities	The IDSA reference architecture forms the basis for a variety of certifiable software solutions, smart services and business models. The business model of the IDSA itself is based on an annual membership fee that depends on the size and type of the organisation[493]
Data volume	Not applicable because the International Data Spaces positions itself as an architecture to link different cloud platforms through secure exchange and trusted sharing of data, i.e. through data sovereignty
Architecture	Decentralized
Client Base/Use Cases	Horizontal Client base with 50 use cases

[493] https://www.internationaldataspaces.org/wp-content/uploads/2020/06/IDSA-MembershipFeeRegulations-2020.pdf

Inrupt (Solid)

Intermediary Category	Personal Data Store
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2017
Stage	Growth
Country/-ies of Establishment	USA (Boston, Massachusetts); UK (London)
Profit/Non-Profit Driven	Non-profit
Number of Employees	Around 25
Revenue/turnover	N/A
Funding	US$ 6.4M from private sources (including Octopus Ventures; Glasswing Ventures)
Business Model/Functionalities	Business model is currently evolving. Solid is a set of technical agreements that enable an ecosystem of data. Aim of Inrupt is to grow the Solid ecosystem for multiple companies. Based on Solid open-source software built to decentralize the web by organizing data, applications, and identities, while focusing on universality by building on existing web/open standards, Inrupt offers software and services for: a. *Organizations and Developers:* • Design a new breed of applications and get better value from data • Inrupt maintains an open source SDK and other tools, hosts a Pod server (Node Solid Server), an open-source software on inrupt.net in order to support development work, and is working on a suite of tools to help enterprise developers. Inrupt.net is a cloud-hosted instance of the open source software Node Solid Server. This software allows users to establish Pods — virtual places where they can store all kinds of data and choose how to share it with applications or other users. The data on all Solid Pods is organized in a common, machine-readable format called Linked Data. This means any Solid app can read and write data to any Solid Pod, decoupling applications from backends. b. *End users through Solid Pods, can:* • store personal data from multiple sources and choose how to share or use personal data. • control permissions on their data and draw value from an ecosystem of beneficent applications On top of the core Solid specification functionality, the Inrupt Enterprise Solid Server (ESS) also provides a number of additional features, including: • **Enhanced Security** — Better protect your data with advanced Auditing, end-to-end TLS encryption, and OIDC/OAuth Access Control features and support. • **Operational Tooling** — Operate a production system with confidence via native monitoring, distributed logging, backup/restore, and simple integration with industry leading ops platforms. • **SLAs** — ESS's microservices architecture enables simple scaling, high performance, and support for highly available deployment configurations. • **Support** — Inrupt offers 24/7 high SLA support for operators and developers with a commercial license for ESS..
Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	• Horizontal client base (including health, government, media, finance) • Aim of Solid concept is to make data available to be used in every possible domain • Currently around 20-30 companies having tried and considering the open Solid technology

Meeco

Intermediary Category	N/A *(it could fall under different categories depending on the use case)*
Data Sharing Scenario	Both C2B *(Me2B)* and B2B *(+B2B2C)*
Type of data sharing	Both personal and industrial data (mainly personal)
Year of Establishment	2012
Stage	Growth
Country/-ies of Establishment	Belgium; UK; Australia
Profit/Non-Profit Driven	For profit
Number of Employees	Around 20
Revenue/turnover	N/A *(790% increase within the last 12 months)*
Funding	A$17.2M from Private Sources (including i.a SVX Group, Present Group Developments, A$11.5M Cash, $750K in kind, $5M Assets)
Business Model/Functionalities	Meeco platform provides various functionalities for: A. *Enterprises. Meeco's Privacy by Design tools:* • Enable customers to control their personal data. Deploy our personal data APIs to develop customer centric Privacy by Design applications. Deliver B2B, B2B2C, C2C & Me2B use cases, always with audit and consent. • Meeco API and Consent Engine a) allows 2-way access to data and verified attributes directly with the customer; b) offers full data and attribute control to your customers aligned with new data regulations; c) introduce new services for customers to delegate authority and manage data access; d) Gain customer consent in all data exchange journeys supported by a fully auditable event log • Meeco Labs program provides organisations an opportunity to test hypotheses and prove business value prior to making substantial investments. It is a custom designed process and pathway to new products, services, experiences and business models. B. *Individuals. Manage and share personal data:* • Gain access and store your personal data across your digital life, encrypted and securely available from any device. • Control whom to share data with. If your information changes, update it once and it will be distributed to your connections. • Delegate permission to enable the people and organisations you trust to act on your behalf. • Share data on your terms and maintain a permanent record of your explicit consent. You can change or revoke access at any time. C. *Developers. Developer Portal utilise Meeco's APIs to generate:* • Privacy by Design data store/wallet/vault include SDKs and extensive support documentation • Consent Engine for permissioned access by duration and date • Key Encryption Store to support Zero Value Knowledge D. *Blockchain & distributed ledger. Implement Standards based solutions for:* • Decentralised and Self Sovereign Identity SSID • Verified Credentials/Claims • DID generation and resolution E. *Data Integration partners.* Meeco's product range includes integrations with a broad range of data sources including social, financial health and IoT, including industry leading partners such as: Xero – accounting data; Class Super – superannuation data; Core Logic (RP Data) — property data; Yodlee — banking, transaction and financial data; XPLAN — financial and estate planning; TopDocs — legal data; Suitebox — compliance data; Citrix RightSignature — consent data.
Data volume	N/A
Architecture	It varies to centralized or decentralized model depending on the use case
Client Base/Use Cases	Horizontal client base with use cases in, among others, banking, government, health, retail, airline, financial services, accounting, education, children with special needs, early children development

MIDATA

Intermediary Category	Data Cooperative
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2015
Stage	Growth
Country/-ies of Establishment	Switzerland
Profit/Non-Profit Driven	Non-profit
Number of Employees	3
Revenue/turnover	N/A
Funding	Funding is limited and comes from: • Foundations • Research projects (research grants) • Limited revenues so far, coming from partner organizations (universities, pharmaceutical companies)
Business Model/Functionalities	• MIDATA operates a data platform, acts as a trustee for data collection and guarantees the sovereignty of citizens over the use of their data, showing how data can be used for the common good, while at the same time ensuring the citizens' control over their personal data. • The MIDATA model is designed for international application: MIDATA Switzerland supports the foundation of regional or national MIDATA cooperatives that share the data platform infrastructure. • At present, MIDATA focuses on health data and smartphone app based services. Startups, IT providers and research groups can connect mobile apps to the platform. The apps may offer data-based services and collect data for analysis. • Owners of a data account at MIDATA may actively contribute to medical research and clinical studies by granting selective access to their personal data. Members own and control the cooperative by governing it at the general assembly. Members write the statutes and decide how the profits will be allocated. Financial benefits/incentives for the members are excluded. • Personal data are stored on the MIDATA platform. Data account holders can participate in app-based research projects and benefit from app-based services. • All datasets are encrypted; only data account holders have access to their individual data. Each access to data is logged. To enable global research and clinical studies, secure data access via individual national cooperatives will be implemented, while at the same time maintaining account holders' full control over their personal data.
Data volume	N/A
Architecture	Cloud infrastructure; All datasets of a member are encrypted into one account under the same password
Client Base/Use Cases	• MIDATA currently has 20.000 users of the platform and 60 members • Horizontal approach of use cases, currently focusing on health and education sector • Around 5-10 partners including, among others, Zurich University hospital and Bern University hospital, ETH Zurich and Bern University of Applied Science in Bern, Leiden University Medical Centre

MindSphere (Siemens)

Intermediary Category	Industrial Data Platform
Data Sharing Scenario	B2B
Type of data sharing	Industrial Data
Year of Establishment	2016 (since 2017 in its current version)
Stage	Growth
Country/-ies of Establishment	Germany; available worldwide
Profit/Non-Profit Driven	For profit
Number of Employees	N/A
Revenue/turnover	N/A
Funding	Developed and funded by Siemens
Business Model/Functionalities	Industrial IOT Operating System, an open platform running on a infrastructure as a service solution: • Connect assets and upload data to the cloud; data may come from all kind of devices, other platforms and databases • Collect, monitor, and analyze data in real-time • Gain insights that improve efficiency and profitability • Add apps that increase the business value of the data An Ecosystem for Developers and Makers: • Open environment for development and operations • Ready-to-use APIs and services • Operating on AWS, Azure & Alibaba infrastructures as public cloud solution; also available as private cloud solution • Thriving community of developers and corporate partners (incl. strategic partners, technology partners, connectivity partners either offering new ways of connecting things or help implementing connectivity, partners creating their own applications) In particular, MindSphere platform offers the following functionalities: • For users: Connecting and monitoring assets and systems and performing advanced analytics; • For developers: Developing and delivering industry IoT applications; • For operators: Deploying and monitoring running applications and see what customers are doing; • For sellers: Marketing solutions to a growing, worldwide MindSphere user base.
Data volume	• 1.4 M # of Connected Devices • 262% increase in Connected Assets at Siemens within the last 1 year
Architecture	Cloud based centralized architecture
Client Base/Use Cases	• More than 6.1000 customers • More than 500 ecosystem partners • Horizontal client base covering broad range of domains (e.g.. manufacturing, machine builders, campuses and cities)

Nallian

Intermediary Category	Industrial Data Platform *(not developed by any industry dominant player)*
Data Sharing Scenario	B2B
Type of data sharing	Industrial data *(sometimes personal data might also be involved)*
Year of Establishment	2012
Stage	Growth
Country/-ies of Establishment	Belgium *(+ Asia establishment to be opened soon)*
Profit/Non-Profit Driven	For profit
Number of Employees	23
Revenue/turnover	N/A *(recurring revenue which comes from the usage of the platform has doubled over the past two years)*
Funding	€1.3M from both public (subsidies by Belgian government) and private sources (incl. i.a. Newion)
Business Model/Functionalities	Nallian platform empowers the different stakeholders at logistic hubs to efficiently align, coordinate their cross-company processes and operate as one (one representation of logistics worldwide). The Open Data Sharing Platform underpins a rich ecosystem of collaborative applications that are developed with, for and by air cargo communities. Tailored to the reality of the logistic hub, they: • enable efficient landside management: Empowering ground handlers, freight forwarder and trucking companies to streamline freight pick-up and delivery from A-Z • provide granular insights and analytics: Providing granular levels of insights that help take informed decisions and fuel user's strategy • facilitate regulatory processes: Streamlining planning, communication and data-exchange with regulatory and governmental instances, such as customs or federal food agencies • enable end-to-end track & trace: Gain control and adopt a pro-active approach with end-to-end traceability and visibility on your shipment's journey • allow seamless data sharing across processes, easy access to innovation, without vendor lock-in and short time to market, as well as solidintegration with existing systems and processes. The platform also presents the following features: • Open approach: Working with legacy systems, easy to add (3rd party) apps as the user grows • Data owner in control: The source always stays in control of who sees which part of his data in which context • Flexible & configurable: Starting with the functionality needed, adding another as the user grows • Single version of truth: Avoiding duplicate data entry and manual errors by sharing a single version of truth (data processing and correlation) • Community-led: Community members decide which use case to participate in, allowing gradual adoption
Data volume	N/A
Architecture	It varies to centralized or decentralized model depending on the use case
Client Base/Use Cases	• Horizontal client base [including i.a. logistics hubs: air cargo (airport authorities, ground handling agents, freight forwarders, trucking companies, regulatory institutions); maritime and shippers; consignees] • More than 100% increase of client base number within the last two years • Examples of clients include among others airports in Brussels, Luxembourg, Vienna, London, Dallas (Texas), Asia, ground handlers such as Swissport, WFS and dnata, forwarders such as DHL global forwarding, and recently also airlines as well as actors in the chemical supply chain such as BASF.

Ocean Protocol

Intermediary Category	N/A (Protocol that allows decentralized exchange of data and digital assets)
Data Sharing Scenario	Both B2B and C2B
Type of data sharing	Generic protocol that can be used for both personal and industrial data
Year of Establishment	2017
Stage	Growth
Country/-ies of Establishment	Germany, Berlin (BigChainDB GmbH); Singapore (Ocean Protocol Foundation)
Profit/Non-Profit Driven	Non-profit
Number of Employees	20-25 core team members | Around 40 advisors
Revenue/turnover	N/A
Funding	$28.1M Private sources/individuals (including Kosmos Capital, Fabric Ventures, Outlier Ventures, IOSG Ventures, Zeroth.AI, Julian Sarokin, Synapse Capital)
Business Model/Functionalities	• Ocean protocol is an open source software that acts as an enabler, helping its partners to build and provide data sharing services (i.a. computational services, storage) • Protocol designing technology that allows decentralized exchange of data and digital assets and helps developers build **marketplaces** and other apps to privately and securely publish, exchange, and consume data. Using Ocean software components, connected to the decentralized Ocean data sharing network: o Data providers can monetize data while preserving privacy and control; o Data consumers can access private data that they could not get before. • As data assets are exchanged, **blockchain technology** provides the security, privacy, and control benefits of Ocean Protocol and makes sellers and buyers benefit from the auditability of purchase transactions. (Tokenized environment) • **Web3 Access Control** (blockchain-enabled access control) allows maintaining control and granting access over the data set, without a centralized intermediary. Each dataset registered in an Ocean Protocol marketplace has a Web3 account as owner attached to it, based on the account used to publish the data set. Only that owner can modify or transfer ownership for the data set. Likewise, only Web3 accounts can consume your data, once they have been granted access to it. • **Compute-to-data** resolves the tradeoff between the benefits of using private data, and the risks of exposing it. It lets the data stay on-premise, yet allows 3rd parties to run specific compute jobs on it to get useful compute results like averaging or building an AI model. • Ocean Protocol's marketplaces and Compute-to-Data help data scientists & AI practitioners get more data, including private data. The non-profit Ocean Protocol Foundation commissioned BigChainDB GmbH to design and develop the core architecture and components of the existing Ocean Protocol software. Both Ocean Protocol Foundation and BigChainDB are actively looking for partnerships to help drive adoption and grow the open source ecosystem engaged in implementing the software for real life data exchange use cases.
Data transaction volume	N/A
Architecture	Decentralized
Client Base/Use Cases	• Horizontal client base mainly in the industries of automotive, logistics and healthcare, ranging from small SMEs to large MNCs • Relatively small client base number at the moment

Polypoly

Intermediary Category	PIMS; Data cooperative (Foundation of three companies including one cooperative)
Data Sharing Scenario	C2B
Type of data sharing	Personal Data
Year of Establishment	2019
Stage	Early
Country/-ies of Establishment	Liechtenstein; Germany (Berlin); South Africa
Profit/Non-Profit Driven	Polypoly Gmbh and Polypoly Cooperative SCE for profit / Polypoly Foundation non-profit
Number of Employees	Around 45
Revenue/turnover	Recently closed deals with three blue chips with an estimated 250.000€ MRR
Funding	Around 20M EUR, from both public and private sources with three pillars: • entrepreneurs/industrial crowdfunding • public sources (including both national and EU funding) • citizens' donations/memberships
Business Model/Functionalities	Polypoly software is not acting in the name of a 3rd party/user, as an intermediate would do – the software is running on the device of the user, providing them the capacity to act on their own. Aim of this software is to make GDPR rights executable for its users. Hosting of data is free of charge for both companies and individuals. Application called polyPod, offering services for: *a. Individuals* • Automation of the process of retrieving data from data providers (i.e Facebook, Amazon, WhatsApp, Google) based on GDPR provisions; finding and managing personal user data available on the Internet and storing it securely on their own electronic devices; • Gain financial benefit from data in an anonymous way; • No collection nor selling of the users's data by Polypoly. *b. Organisations (companies, governmental organizations and NGOs)* • personalised services using the data stored in individual polyPods (by giving data back to the users), all with the user's consent and without it leaving the user's device; All computations are being done on the end-users device and nobody is needed to enforce end-users' rights; • access to better and cheaper data sets, while upholding GDPR compliance, and reducing costs related to storing data and keeping them up to date, IT and cybersecurity, research, product development and improvement, GDPR compliance costs; • Aim for auditing, contract manufacturing, training, and big data services in the future. A European Cooperative (Polypoly SCE) has recently been set up: • Owned, driven and controlled by the users/members, by installing the software with an optional charge of 5 EUR; No legal body can buy a share of the cooperative • Changes in the fundamental element of the software require the agreement of the users; 1 vote corresponds to 1 user no matter how many shares they own..
Data volume	N/A
Architecture	Decentralized
Client Base/Use Cases	Polypoly GmbH: • Horizontal Use Cases (including i.a banks, insurance companies, mobility providers) Cooperative Polypoly SCE: • No users yet (planned go live date in early 2021) • Aim for horizontal approach of use cases (including i.a.neutral payment infrastructure, banks, messaging infrastructure, federated AI) Polypoly Foundation • Building Data Coops outside of Europe (Franchise model), running negations with India, South Africa, Canada, Switzerland, UK and the USA • Running polyPedia: open source database about the data behavior of several thousand companies.

Streamr

Intermediary Category	Data Union; Data Marketplace
Data Sharing Approach	Both C2B and B2B (mainly C2B)
Type of data sharing	Both personal and industrial (mainly personal)
Year of Establishment	Founded in 2015; operational since 2017
Stage	Growth
Country/-ies of Establishment	Finland; Switzerland
Profit/Non-Profit Driven	For profit
Number of Employees	Approximately 35
Revenue/turnover	N/A
Funding	27.7M EUR in 2017 Crowdfunded; both private and public sources (incl. Firestartr (Lead Investor); EIT Digital Accelerator; Fabric Ventures ; Andreas Schwartz)
Business Model/Functionalities	Provision of three sets of (neutral) functions rolled into one: • the ability to transport data from one platform to the end buyer of the data • the marketplace with two functions: ○ Discovery of the data ○ Aggregation of individuals' data within the data union framework into one product • Micropayments (for the payment of data union members)
Data volume	• 1000s messages/second through the Streamr Network • Financial value of data transactions: "small" transactions at the moment; much larger ones expected when unique data sets (e.g Swash) are viable and sold within a six-month to a year timeframe
Architecture	Decentralized peer to peer network
Client Base/Use Cases	Horizontal client base (incl. i.a. environment, IoT, transportation, energy, health, retail) • Swash (1,00+ members) • Tracey app (WWF Philippines - UnionBank - TX partnership) • 15 business partners (incl. Hewlett Packard Enterprise, WWF and DXC Technology, Bosch, Fastems, and Capita)

Smart Connected Supplier Network (SCSN)

Intermediary Category	Trusted third party that enables a fast, secure and interoperable exchange of information across company borders based on an open standard (called the SCSN standard) that builds on the IDSA (International Data Spaces Association) standard.
Data Sharing Scenario	B2B
Type of data sharing	Industrial data
Year of Establishment	The SCSN initiative started in 2015 as a project called " Connections in the chain" and partners cooperate in a field lab to develop the standard. The foundation SCSN is established in 2020.
Stage	Growth
Country/-ies of Establishment	The Netherlands
Profit/Non-Profit Driven	Non-profit
Number of Employees	Various partners contribute with FTE, so about 15 FTE are involved. However, the foundation does not have employees only a board and a supervisory board
Revenue/turnover	The foundation is expected to cost 200k (e.g. to maintain the standard). Potentially part of these costs will be covered by in-kind contributionsl via branche associations
Funding	Currently public-private funding
Business Model/Functionalities	• They will work with a participation fee in the future to maintain the data exchange standard. Some within SCSN involved service providers will also use a pay-per use fee
Data volume	• Not applicable because SCSN enables the data exchange based on a standard that is building on the IDSA. SCSN does not store the data of the users.
Architecture	Decentralized
Client Base/Use Cases	Horizontal client base (but a main focus on Manufacturing) The SCSN network involve the following partners: • OEMs first-tier suppliers, smaller second and third-tier suppliers, but also (steel) wholesalers and steel manufacturers are connected via the SCSN standard. • Service providers are IT partners who facilitate the connection to the SCSN network for manufacturing companies. Service Providers have set up various standard connections with a diverse portfolio of IT systems, so that they can easily connect manufacturing companies to the SCSN network. The idea is that a manufacturer only needs to be connected once to get access the whole supply network. • 9 service providers are currently involved in the SCSN network. These 9 service providers can be defined as data intermediaries. Manufacturing companies can chose the service provider they prefer. These are: *Fujitsu Glovia*, *Supplydrive*, *Tradecloud, ISAH Business Software*, *Trivest Connect*, *Ketenlink, Exact, Easy2Trade (INAD) , Attributes* • Almost 300 manufactures are currently connected via SCSN. The goal is to grow to 1000 connected partners in the coming year. • There is an onboarding program for both new manufacturers as well as new service providers.)

4.1.4 Macro economic analysis | Top/down

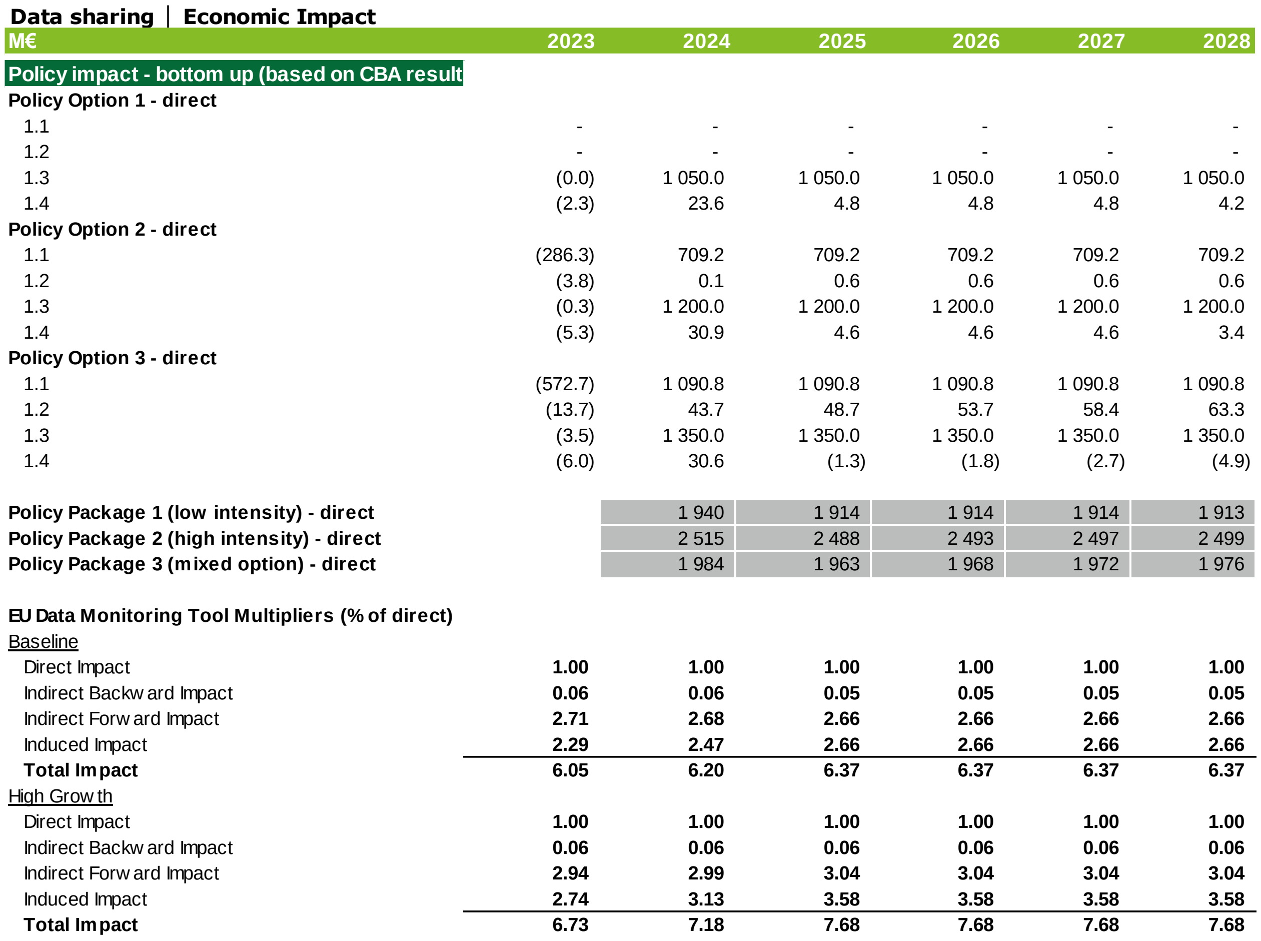
Data sharing | Economic Impact

M€	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Real GDP (% change p.a. EU/OECD)		*(7,8%)*	*5,3%*	*2,8%*	*2,1%*	*1,8%*	*1,5%*	*1,5%*	*1,6%*	*1,6%*
EU Data Monitoring Tool 2020 - baseline										
Data revenues	64 262	71 050	75 866	81 008	86 499	92 362	98 623	100 144	101 711	103 321
Data market value	58 214	62 244	65 795	69 584	73 628	77 948	82 564	83 837	85 149	86 497
Value of Data Economy										
Direct Impact	58 214	54 081	58 481	63 239	68 385	73 948	79 965	81 198	82 469	83 775
Indirect Backward Impact	3 197	3 105	3 324	3 559	3 811	4 081	4 369	4 436	4 506	4 577
Indirect Forward Impact	155 389	150 887	161 556	172 979	185 209	198 305	212 326	215 600	218 975	222 441
Induced Impact	108 058	98 853	115 213	134 280	156 502	182 402	212 589	215 867	219 246	222 717
Total Impact	**324 858**	**306 926**	**338 574**	**374 057**	**413 907**	**458 736**	**509 249**	**517 101**	**525 197**	**533 510**
EU Data Monitoring Tool 2020 - high growth										
Data revenues	64 262	71 050	80 943	92 215	105 055	119 684	136 350	138 453	140 620	142 846
Data market value	58 214	62 244	69 320	77 236	86 097	96 020	107 139	108 791	110 494	112 243
Value of Data Economy										
Direct Impact	58 214	54 081	62 005	71 090	81 505	93 447	107 139	108 791	110 494	112 243
Indirect Backward Impact	3 197	3 105	3 622	4 224	4 928	5 748	6 704	6 808	6 914	7 024
Indirect Forward Impact	155 389	150 887	176 002	205 296	239 467	279 324	325 817	330 840	336 020	341 339
Induced Impact	108 058	98 853	129 651	170 044	223 023	292 506	383 638	389 553	395 652	401 915
Total Impact	**324 858**	**306 926**	**371 279**	**450 655**	**548 922**	**671 026**	**823 298**	**835 992**	**849 081**	**862 521**
Data sharing [% of total Data Economy]		80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%
- share linked to trust [% of total Data Economy]		50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%
Data sharing [% linked to trust]		40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%	40,0%
Data sharing linked to trust - potential gap	**-**	**-**	**13 082**	**30 639**	**54 006**	**84 916**	**125 620**	**127 556**	**129 553**	**131 604**

Data sharing | Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - top-down estimation						
Policy Option 1 [% realisation of gap]						
1.1		0.15%	0.15%	0.15%	0.15%	0.15%
1.2		0.3%	0.3%	0.3%	0.3%	0.3%
1.3		1.0%	1.0%	1.0%	1.0%	1.0%
1.4		1.0%	1.0%	1.0%	1.0%	1.0%
Policy Option 2 [% realisation of gap]						
1.1		0.8%	0.8%	0.8%	0.8%	0.8%
1.2		1.2%	1.2%	1.2%	1.2%	1.2%
1.3		2.0%	2.0%	2.0%	2.0%	2.0%
1.4		4.0%	4.0%	4.0%	4.0%	4.0%
Policy Option 3 [% realisation of gap]						
1.1		1.0%	1.0%	1.0%	1.0%	1.0%
1.2		1.5%	1.5%	1.5%	1.5%	1.5%
1.3		3.0%	3.0%	3.0%	3.0%	3.0%
1.4		5.0%	5.0%	5.0%	5.0%	5.0%
Policy Option 1 [realisation of gap]						
1.1		127	188	191	194	197
1.2		255	377	383	389	395
1.3		849	1 256	1 276	1 296	1 316
1.4		849	1 256	1 276	1 296	1 316
Policy Option 2 [realisation of gap]						
1.1		679	1 005	1 020	1 036	1 053
1.2		1 019	1 507	1 531	1 555	1 579
1.3		1 698	2 512	2 551	2 591	2 632
1.4		3 397	5 025	5 102	5 182	5 264
Policy Option 3 [realisation of gap]						
1.1		849	1 256	1 276	1 296	1 316
1.2		1 274	1 884	1 913	1 943	1 974
1.3		2 547	3 769	3 827	3 887	3 948
1.4		4 246	6 281	6 378	6 478	6 580
Policy Package 1 (low intensity)		6 793	10 050	10 205	10 364	10 528
Policy Package 2 (high intensity)		8 916	13 190	13 393	13 603	13 818
Policy Package 3 (mixed option)		7 048	10 426	10 587	10 753	10 923

4.1.5 Macro economic analysis | Bottom/up



Data sharing | Economic Impact

M€	2023	2024	2025	2026	2027	2028
Policy impact - bottom up (based on CBA result						
Policy Option 1 - direct						
1.1	-	-	-	-	-	-
1.2	-	-	-	-	-	-
1.3	(0.0)	1 050.0	1 050.0	1 050.0	1 050.0	1 050.0
1.4	(2.3)	23.6	4.8	4.8	4.8	4.2
Policy Option 2 - direct						
1.1	(286.3)	709.2	709.2	709.2	709.2	709.2
1.2	(3.8)	0.1	0.6	0.6	0.6	0.6
1.3	(0.3)	1 200.0	1 200.0	1 200.0	1 200.0	1 200.0
1.4	(5.3)	30.9	4.6	4.6	4.6	3.4
Policy Option 3 - direct						
1.1	(572.7)	1 090.8	1 090.8	1 090.8	1 090.8	1 090.8
1.2	(13.7)	43.7	48.7	53.7	58.4	63.3
1.3	(3.5)	1 350.0	1 350.0	1 350.0	1 350.0	1 350.0
1.4	(6.0)	30.6	(1.3)	(1.8)	(2.7)	(4.9)
Policy Package 1 (low intensity) - direct		1 940	1 914	1 914	1 914	1 913
Policy Package 2 (high intensity) - direct		2 515	2 488	2 493	2 497	2 499
Policy Package 3 (mixed option) - direct		1 984	1 963	1 968	1 972	1 976
EU Data Monitoring Tool Multipliers (% of direct)						
Baseline						
Direct Impact	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Indirect Backward Impact	**0.06**	**0.06**	**0.05**	**0.05**	**0.05**	**0.05**
Indirect Forward Impact	**2.71**	**2.68**	**2.66**	**2.66**	**2.66**	**2.66**
Induced Impact	**2.29**	**2.47**	**2.66**	**2.66**	**2.66**	**2.66**
Total Impact	**6.05**	**6.20**	**6.37**	**6.37**	**6.37**	**6.37**
High Growth						
Direct Impact	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Indirect Backward Impact	**0.06**	**0.06**	**0.06**	**0.06**	**0.06**	**0.06**
Indirect Forward Impact	**2.94**	**2.99**	**3.04**	**3.04**	**3.04**	**3.04**
Induced Impact	**2.74**	**3.13**	**3.58**	**3.58**	**3.58**	**3.58**
Total Impact	**6.73**	**7.18**	**7.68**	**7.68**	**7.68**	**7.68**

Data sharing | Economic Impact

M€	2023	2024	2025	2026	2027	2028
Multiplier (indirect forward) applied						
Policy Option 1 - indirect forward		2.68	2.66	2.66	2.66	2.66
1.1		-	-	-	-	-
1.2		-	-	-	-	-
1.3		2 815.7	2 788.0	2 788.0	2 788.0	2 788.0
1.4		63.3	12.9	12.9	12.9	11.1
Policy Option 2 - indirect forward						
1.1		1 901.9	1 883.2	1 883.2	1 883.2	1 883.2
1.2		0.1	1.7	1.7	1.7	1.7
1.3		3 218.0	3 186.3	3 186.3	3 186.3	3 186.3
1.4		82.8	12.3	12.3	12.3	8.9
Policy Option 3 - indirect forward						
1.1		2 925.2	2 896.3	2 896.3	2 896.3	2 896.3
1.2		117.3	129.4	142.7	155.1	168.0
1.3		3 620.2	3 584.6	3 584.6	3 584.6	3 584.6
1.4		82.1	(3.6)	(4.7)	(7.2)	(13.0)
Policy Package 1 (low intensity) - indirect		5 203	5 083	5 083	5 083	5 080
Policy Package 2 (high intensity) - indirect		6 745	6 607	6 619	6 629	6 636
Policy Package 3 (mixed option) - indirect		5 320	5 211	5 224	5 237	5 246
Summary - results						
Top-down						
Policy Package 1 (low intensity) - total		6 793	10 050	10 205	10 364	10 528
Policy Package 2 (high intensity) - total		8 916	13 190	13 393	13 603	13 818
Policy Package 3 (mixed option) - total		7 048	10 426	10 587	10 753	10 923
Bottom-up						
Policy Package 1 (low intensity) - total		7 143	6 998	6 998	6 998	6 993
Policy Package 2 (high intensity) - total		9 260	9 095	9 112	9 125	9 135
Policy Package 3 (mixed option) - total		7 304	7 174	7 192	7 209	7 222
Average						
Policy Package 1 (low intensity) - total		6 968	8 524	8 601	8 681	8 761
Policy Package 2 (high intensity) - total		9 088	11 142	11 253	11 364	11 477
Policy Package 3 (mixed option) - total		7 176	8 800	8 890	8 981	9 073
GDP - EDM pre Covid	13 287 687	13 487 002	13 690 074	13 901 151	14 118 796	14 342 287
GDP - EDM post Covid (EIU/OECD correction)	12 742 133	12 972 879	13 168 121	13 371 151	13 580 498	13 795 468

Data sharing | Economic Impact

M€	2020	2021	2022	2023	2024	2025	2026	2027	2028
Impact on the Economic Value of the Data Economy compared to GDP [m€]									
Baseline	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.54%***	***3.87%***	***3.87%***	***3.87%***	***3.87%***
Policy Package 1 (top-dow n)	306 926	338 574	374 057	413 907	465 529	519 299	527 305	535 561	544 039
% Policy Package 1 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.94%***	***3.94%***	***3.94%***	***3.94%***
Policy Package 2 (top-dow n)	306 926	338 574	374 057	413 907	467 652	522 439	530 494	538 800	547 329
% Policy Package 2 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.60%***	***3.97%***	***3.97%***	***3.97%***	***3.97%***
Policy Package 3 (top-dow n)	306 926	338 574	374 057	413 907	465 784	519 675	527 688	535 950	544 433
% Policy Package 3 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.95%***	***3.95%***	***3.95%***	***3.95%***
Baseline	306 926	338 574	374 057	413 907	458 736	509 249	517 101	525 197	533 510
% Baseline to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.54%***	***3.87%***	***3.87%***	***3.87%***	***3.87%***
Policy Package 1 (bottom-up)	306 926	338 574	374 057	413 907	465 879	516 247	524 099	532 195	540 504
% Policy Package 1 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.92%***	***3.92%***	***3.92%***	***3.92%***
Policy Package 2 (bottom-up)	306 926	338 574	374 057	413 907	467 996	518 344	526 212	534 322	542 645
% Policy Package 2 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.61%***	***3.94%***	***3.94%***	***3.93%***	***3.93%***
Policy Package 3 (bottom-up)	306 926	338 574	374 057	413 907	466 040	516 423	524 293	532 406	540 732
% Policy Package 3 to GDP	***2.66%***	***2.79%***	***3.00%***	***3.25%***	***3.59%***	***3.92%***	***3.92%***	***3.92%***	***3.92%***

4.2 Annex II - Measures to foster data sharing and re-use

4.2.1 Cost-Benefit Analysis

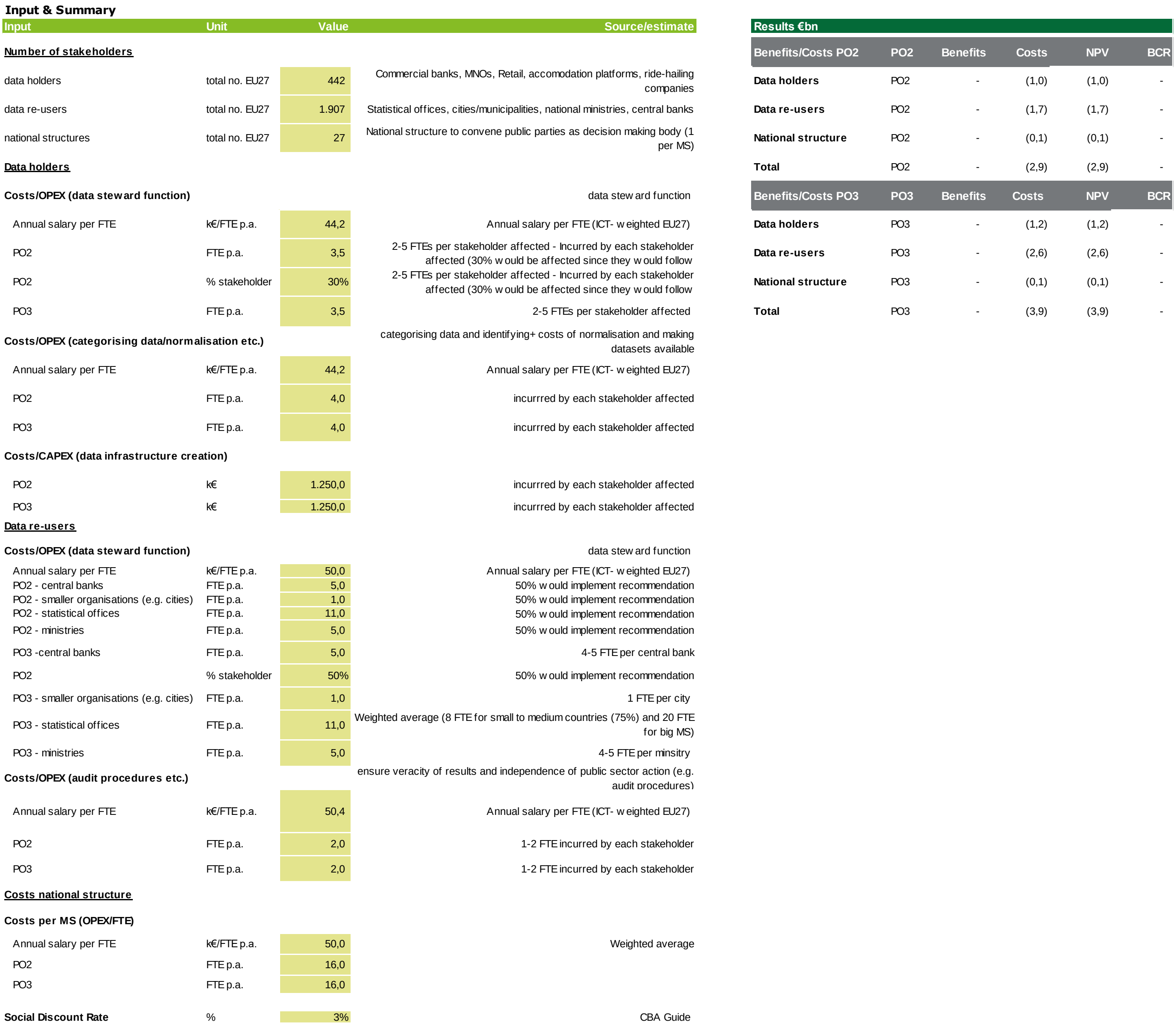
4.2.1.1 Summary of net benefits for domains Measures supporting citizen empowerment ('human-centric data economy'); and Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

Cost-Benefit Analysis - Summary 2.2 and 2.3

M€ (constant prices)	PO	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Costs total	PO1	(274.1)	(295.7)	(119.0)	(47.6)	(40.5)	(34.4)	(29.3)	(24.9)
Benefits total	PO1	1 570.6	1 786.6	-	231.9	282.3	343.8	418.7	509.9
Net Cashflow NPV	PO1	**1 296.4**	**1 490.9**	**(119.0)**	**184.2**	**241.8**	**309.4**	**389.4**	**485.0**
Benefit/Cost-ratio	PO1	**5.73**							
Costs total	PO2	(31 098.4)	(34 833.7)	(1 399.4)	(6 752.1)	(6 714.2)	(6 682.0)	(6 654.6)	(6 631.4)
Benefits total	PO2	1 180 131.1	1 327 110.1	-	265 287.0	265 341.8	265 407.8	265 488.0	265 585.5
Net Cashflow NPV	PO2	**1 149 032.7**	**1 292 276.4**	**(1 399.4)**	**258 534.9**	**258 627.6**	**258 725.8**	**258 833.4**	**258 954.2**
Benefit/Cost-ratio	PO2	**37.95**							
Costs total	PO3	(61 469.3)	(68 797.2)	(3 423.3)	(13 165.9)	(13 113.0)	(13 068.0)	(13 029.8)	(12 997.3)
Benefits total	PO3	989 136.1	1 112 326.2	-	222 361.2	222 403.0	222 454.0	222 516.1	222 591.8
Net Cashflow NPV	PO3	**927 666.7**	**1 043 528.9**	**(3 423.3)**	**209 195.4**	**209 290.1**	**209 386.0**	**209 486.3**	**209 594.5**
Benefit/Cost-ratio	PO3	**16.09**							

4.2.1.2 Business-to-Government (B2G) data sharing for the public interest

The figure below presents the input summary for the cost-benefit analysis for Business-to-Government (B2G) data sharing for the public interest.



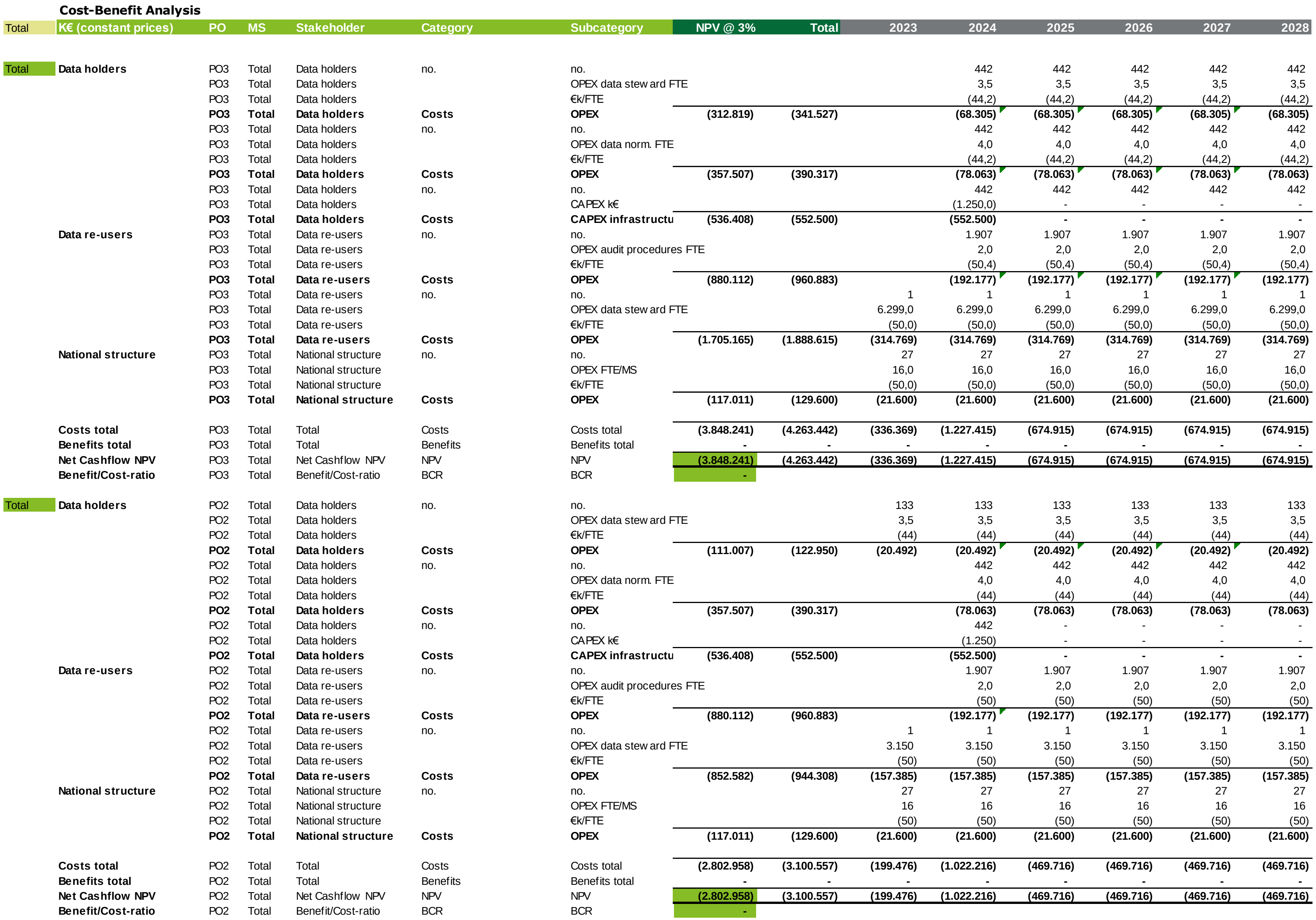
Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders	total no. EU27	442	Commercial banks, MNOs, Retail, accomodation platforms, ride-hailing companies
data re-users	total no. EU27	1.907	Statistical offices, cities/municipalities, national ministries, central banks
national structures	total no. EU27	27	National structure to convene public parties as decision making body (1 per MS)
Data holders			
Costs/OPEX (data steward function)			data steward function
Annual salary per FTE	k€/FTE p.a.	44,2	Annual salary per FTE (ICT- weighted EU27)
PO2	FTE p.a.	3,5	2-5 FTEs per stakeholder affected - Incurred by each stakeholder affected (30% would be affected since they would follow
PO2	% stakeholder	30%	2-5 FTEs per stakeholder affected - Incurred by each stakeholder affected (30% would be affected since they would follow
PO3	FTE p.a.	3,5	2-5 FTEs per stakeholder affected
Costs/OPEX (categorising data/normalisation etc.)			categorising data and identifying+ costs of normalisation and making datasets available
Annual salary per FTE	k€/FTE p.a.	44,2	Annual salary per FTE (ICT- weighted EU27)
PO2	FTE p.a.	4,0	incurrred by each stakeholder affected
PO3	FTE p.a.	4,0	incurrred by each stakeholder affected
Costs/CAPEX (data infrastructure creation)			
PO2	k€	1.250,0	incurrred by each stakeholder affected
PO3	k€	1.250,0	incurrred by each stakeholder affected
Data re-users			
Costs/OPEX (data steward function)			data steward function
Annual salary per FTE	k€/FTE p.a.	50,0	Annual salary per FTE (ICT- weighted EU27)
PO2 - central banks	FTE p.a.	5,0	50% would implement recommendation
PO2 - smaller organisations (e.g. cities)	FTE p.a.	1,0	50% would implement recommendation
PO2 - statistical offices	FTE p.a.	11,0	50% would implement recommendation
PO2 - ministries	FTE p.a.	5,0	50% would implement recommendation
PO3 -central banks	FTE p.a.	5,0	4-5 FTE per central bank
PO2	% stakeholder	50%	50% would implement recommendation
PO3 - smaller organisations (e.g. cities)	FTE p.a.	1,0	1 FTE per city
PO3 - statistical offices	FTE p.a.	11,0	Weighted average (8 FTE for small to medium countries (75%) and 20 FTE for big MS)
PO3 - ministries	FTE p.a.	5,0	4-5 FTE per minsitry
Costs/OPEX (audit procedures etc.)			ensure veracity of results and independence of public sector action (e.g. audit procedures)
Annual salary per FTE	k€/FTE p.a.	50,4	Annual salary per FTE (ICT- weighted EU27)
PO2	FTE p.a.	2,0	1-2 FTE incurred by each stakeholder
PO3	FTE p.a.	2,0	1-2 FTE incurred by each stakeholder
Costs national structure			
Costs per MS (OPEX/FTE)			
Annual salary per FTE	k€/FTE p.a.	50,0	Weighted average
PO2	FTE p.a.	16,0	
PO3	FTE p.a.	16,0	
Social Discount Rate	%	3%	CBA Guide

Results €bn

Benefits/Costs PO2	PO2	Benefits	Costs	NPV	BCR
Data holders	PO2	-	(1,0)	(1,0)	-
Data re-users	PO2	-	(1,7)	(1,7)	-
National structure	PO2	-	(0,1)	(0,1)	-
Total	PO2	-	(2,9)	(2,9)	-
Benefits/Costs PO3	**PO3**	**Benefits**	**Costs**	**NPV**	**BCR**
Data holders	PO3	-	(1,2)	(1,2)	-
Data re-users	PO3	-	(2,6)	(2,6)	-
National structure	PO3	-	(0,1)	(0,1)	-
Total	PO3	-	(3,9)	(3,9)	-

The figure below presents the cost-benefit analysis for Business-to-Government (B2G) data sharing for the public interest.

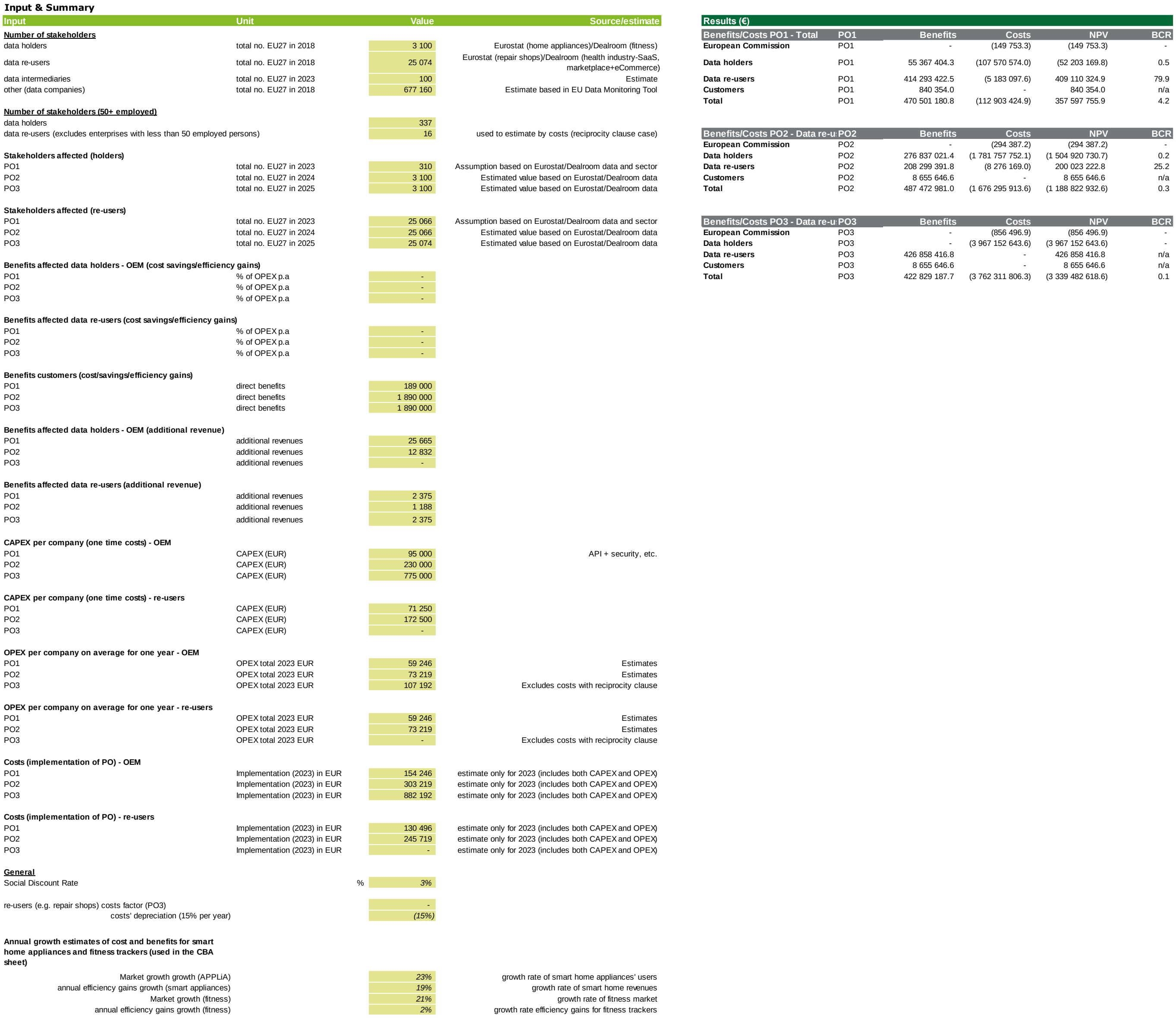


Cost-Benefit Analysis

Total	K€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Total	Data holders	PO3	Total	Data holders	no.	no.				442	442	442	442	442
		PO3	Total	Data holders		OPEX data stew ard FTE				3,5	3,5	3,5	3,5	3,5
		PO3	Total	Data holders		€k/FTE				(44,2)	(44,2)	(44,2)	(44,2)	(44,2)
		PO3	Total	Data holders	Costs	OPEX	(312.819)	(341.527)		(68.305)	(68.305)	(68.305)	(68.305)	(68.305)
		PO3	Total	Data holders	no.	no.				442	442	442	442	442
		PO3	Total	Data holders		OPEX data norm. FTE				4,0	4,0	4,0	4,0	4,0
		PO3	Total	Data holders		€k/FTE				(44,2)	(44,2)	(44,2)	(44,2)	(44,2)
		PO3	Total	Data holders	Costs	OPEX	(357.507)	(390.317)		(78.063)	(78.063)	(78.063)	(78.063)	(78.063)
		PO3	Total	Data holders	no.	no.				442	442	442	442	442
		PO3	Total	Data holders		CAPEX k€				(1.250,0)	-	-	-	-
		PO3	Total	Data holders	Costs	CAPEX infrastructu	(536.408)	(552.500)		(552.500)	-	-	-	-
	Data re-users	PO3	Total	Data re-users	no.	no.				1.907	1.907	1.907	1.907	1.907
		PO3	Total	Data re-users		OPEX audit procedures FTE				2,0	2,0	2,0	2,0	2,0
		PO3	Total	Data re-users		€k/FTE				(50,4)	(50,4)	(50,4)	(50,4)	(50,4)
		PO3	Total	Data re-users	Costs	OPEX	(880.112)	(960.883)		(192.177)	(192.177)	(192.177)	(192.177)	(192.177)
		PO3	Total	Data re-users	no.	no.			1	1	1	1	1	1
		PO3	Total	Data re-users		OPEX data stew ard FTE			6.299,0	6.299,0	6.299,0	6.299,0	6.299,0	6.299,0
		PO3	Total	Data re-users		€k/FTE			(50,0)	(50,0)	(50,0)	(50,0)	(50,0)	(50,0)
		PO3	Total	Data re-users	Costs	OPEX	(1.705.165)	(1.888.615)	(314.769)	(314.769)	(314.769)	(314.769)	(314.769)	(314.769)
	National structure	PO3	Total	National structure	no.	no.			27	27	27	27	27	27
		PO3	Total	National structure		OPEX FTE/MS			16,0	16,0	16,0	16,0	16,0	16,0
		PO3	Total	National structure		€k/FTE			(50,0)	(50,0)	(50,0)	(50,0)	(50,0)	(50,0)
		PO3	Total	National structure	Costs	OPEX	(117.011)	(129.600)	(21.600)	(21.600)	(21.600)	(21.600)	(21.600)	(21.600)
	Costs total	PO3	Total	Total	Costs	Costs total	(3.848.241)	(4.263.442)	(336.369)	(1.227.415)	(674.915)	(674.915)	(674.915)	(674.915)
	Benefits total	PO3	Total	Total	Benefits	Benefits total	-	-	-	-	-	-	-	-
	Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	(3.848.241)	(4.263.442)	(336.369)	(1.227.415)	(674.915)	(674.915)	(674.915)	(674.915)
	Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	-							
Total	Data holders	PO2	Total	Data holders	no.	no.			133	133	133	133	133	133
		PO2	Total	Data holders		OPEX data stew ard FTE			3,5	3,5	3,5	3,5	3,5	3,5
		PO2	Total	Data holders		€k/FTE			(44)	(44)	(44)	(44)	(44)	(44)
		PO2	Total	Data holders	Costs	OPEX	(111.007)	(122.950)	(20.492)	(20.492)	(20.492)	(20.492)	(20.492)	(20.492)
		PO2	Total	Data holders	no.	no.				442	442	442	442	442
		PO2	Total	Data holders		OPEX data norm. FTE				4,0	4,0	4,0	4,0	4,0
		PO2	Total	Data holders		€k/FTE				(44)	(44)	(44)	(44)	(44)
		PO2	Total	Data holders	Costs	OPEX	(357.507)	(390.317)		(78.063)	(78.063)	(78.063)	(78.063)	(78.063)
		PO2	Total	Data holders	no.	no.				442	-	-	-	-
		PO2	Total	Data holders		CAPEX k€				(1.250)	-	-	-	-
		PO2	Total	Data holders	Costs	CAPEX infrastructu	(536.408)	(552.500)		(552.500)	-	-	-	-
	Data re-users	PO2	Total	Data re-users	no.	no.				1.907	1.907	1.907	1.907	1.907
		PO2	Total	Data re-users		OPEX audit procedures FTE				2,0	2,0	2,0	2,0	2,0
		PO2	Total	Data re-users		€k/FTE				(50)	(50)	(50)	(50)	(50)
		PO2	Total	Data re-users	Costs	OPEX	(880.112)	(960.883)		(192.177)	(192.177)	(192.177)	(192.177)	(192.177)
		PO2	Total	Data re-users	no.	no.			1	1	1	1	1	1
		PO2	Total	Data re-users		OPEX data stew ard FTE			3.150	3.150	3.150	3.150	3.150	3.150
		PO2	Total	Data re-users		€k/FTE			(50)	(50)	(50)	(50)	(50)	(50)
		PO2	Total	Data re-users	Costs	OPEX	(852.582)	(944.308)	(157.385)	(157.385)	(157.385)	(157.385)	(157.385)	(157.385)
	National structure	PO2	Total	National structure	no.	no.			27	27	27	27	27	27
		PO2	Total	National structure		OPEX FTE/MS			16	16	16	16	16	16
		PO2	Total	National structure		€k/FTE			(50)	(50)	(50)	(50)	(50)	(50)
		PO2	Total	National structure	Costs	OPEX	(117.011)	(129.600)	(21.600)	(21.600)	(21.600)	(21.600)	(21.600)	(21.600)
	Costs total	PO2	Total	Total	Costs	Costs total	(2.802.958)	(3.100.557)	(199.476)	(1.022.216)	(469.716)	(469.716)	(469.716)	(469.716)
	Benefits total	PO2	Total	Total	Benefits	Benefits total	-	-	-	-	-	-	-	-
	Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	(2.802.958)	(3.100.557)	(199.476)	(1.022.216)	(469.716)	(469.716)	(469.716)	(469.716)
	Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	-							

4.2.1.3 Measures supporting citizen empowerment (‘human-centric data economy’)

The figure below presents the input summary for the cost-benefit analysis for Measures supporting citizen empowerment (‘human-centric data economy’) a) “Smart home appliances”.



Input & Summary

Input	Unit		Value	Source/estimate
Number of stakeholders				
data holders	total no. EU27 in 2018		3 100	Eurostat (home appliances)/Dealroom (fitness)
data re-users	total no. EU27 in 2018		25 074	Eurostat (repair shops)/Dealroom (health industry-SaaS, marketplace+eCommerce)
data intermediaries	total no. EU27 in 2023		100	Estimate
other (data companies)	total no. EU27 in 2018		677 160	Estimate based in EU Data Monitoring Tool
Number of stakeholders (50+ employed)				
data holders			337	
data re-users (excludes enterprises with less than 50 employed persons)			16	used to estimate by costs (reciprocity clause case)
Stakeholders affected (holders)				
PO1	total no. EU27 in 2023		310	Assumption based on Eurostat/Dealroom data and sector
PO2	total no. EU27 in 2024		3 100	Estimated value based on Eurostat/Dealroom data
PO3	total no. EU27 in 2025		3 100	Estimated value based on Eurostat/Dealroom data
Stakeholders affected (re-users)				
PO1	total no. EU27 in 2023		25 066	Assumption based on Eurostat/Dealroom data and sector
PO2	total no. EU27 in 2024		25 066	Estimated value based on Eurostat/Dealroom data
PO3	total no. EU27 in 2025		25 074	Estimated value based on Eurostat/Dealroom data
Benefits affected data holders - OEM (cost savings/efficiency gains)				
PO1	% of OPEX p.a		-	
PO2	% of OPEX p.a		-	
PO3	% of OPEX p.a		-	
Benefits affected data re-users (cost savings/efficiency gains)				
PO1	% of OPEX p.a		-	
PO2	% of OPEX p.a		-	
PO3	% of OPEX p.a		-	
Benefits customers (cost/savings/efficiency gains)				
PO1	direct benefits		189 000	
PO2	direct benefits		1 890 000	
PO3	direct benefits		1 890 000	
Benefits affected data holders - OEM (additional revenue)				
PO1	additional revenues		25 665	
PO2	additional revenues		12 832	
PO3	additional revenues		-	
Benefits affected data re-users (additional revenue)				
PO1	additional revenues		2 375	
PO2	additional revenues		1 188	
PO3	additional revenues		2 375	
CAPEX per company (one time costs) - OEM				
PO1	CAPEX (EUR)		95 000	API + security, etc.
PO2	CAPEX (EUR)		230 000	
PO3	CAPEX (EUR)		775 000	
CAPEX per company (one time costs) - re-users				
PO1	CAPEX (EUR)		71 250	
PO2	CAPEX (EUR)		172 500	
PO3	CAPEX (EUR)		-	
OPEX per company on average for one year - OEM				
PO1	OPEX total 2023 EUR		59 246	Estimates
PO2	OPEX total 2023 EUR		73 219	Estimates
PO3	OPEX total 2023 EUR		107 192	Excludes costs with reciprocity clause
OPEX per company on average for one year - re-users				
PO1	OPEX total 2023 EUR		59 246	Estimates
PO2	OPEX total 2023 EUR		73 219	Estimates
PO3	OPEX total 2023 EUR		-	Excludes costs with reciprocity clause
Costs (implementation of PO) - OEM				
PO1	Implementation (2023) in EUR		154 246	estimate only for 2023 (includes both CAPEX and OPEX)
PO2	Implementation (2023) in EUR		303 219	estimate only for 2023 (includes both CAPEX and OPEX)
PO3	Implementation (2023) in EUR		882 192	estimate only for 2023 (includes both CAPEX and OPEX)
Costs (implementation of PO) - re-users				
PO1	Implementation (2023) in EUR		130 496	estimate only for 2023 (includes both CAPEX and OPEX)
PO2	Implementation (2023) in EUR		245 719	estimate only for 2023 (includes both CAPEX and OPEX)
PO3	Implementation (2023) in EUR		-	estimate only for 2023 (includes both CAPEX and OPEX)
General				
Social Discount Rate		%	*3%*	
re-users (e.g. repair shops) costs factor (PO3)			-	
costs' depreciation (15% per year)			*(15%)*	
Annual growth estimates of cost and benefits for smart home appliances and fitness trackers (used in the CBA sheet)				
Market growth growth (APPLiA)			*23%*	growth rate of smart home appliances' users
annual efficiency gains growth (smart appliances)			*19%*	growth rate of smart home revenues
Market growth (fitness)			*21%*	growth rate of fitness market
annual efficiency gains growth (fitness)			*2%*	growth rate efficiency gains for fitness trackers

Results (€)

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
European Commission	PO1	-	(149 753.3)	(149 753.3)	-
Data holders	PO1	55 367 404.3	(107 570 574.0)	(52 203 169.8)	0.5
Data re-users	PO1	414 293 422.5	(5 183 097.6)	409 110 324.9	79.9
Customers	PO1	840 354.0	-	840 354.0	n/a
Total	PO1	470 501 180.8	(112 903 424.9)	357 597 755.9	4.2

Benefits/Costs PO2 - Data re-u	PO2	Benefits	Costs	NPV	BCR
European Commission	PO2	-	(294 387.2)	(294 387.2)	-
Data holders	PO2	276 837 021.4	(1 781 757 752.1)	(1 504 920 730.7)	0.2
Data re-users	PO2	208 299 391.8	(8 276 169.0)	200 023 222.8	25.2
Customers	PO2	8 655 646.6	-	8 655 646.6	n/a
Total	PO2	487 472 981.0	(1 676 295 913.6)	(1 188 822 932.6)	0.3

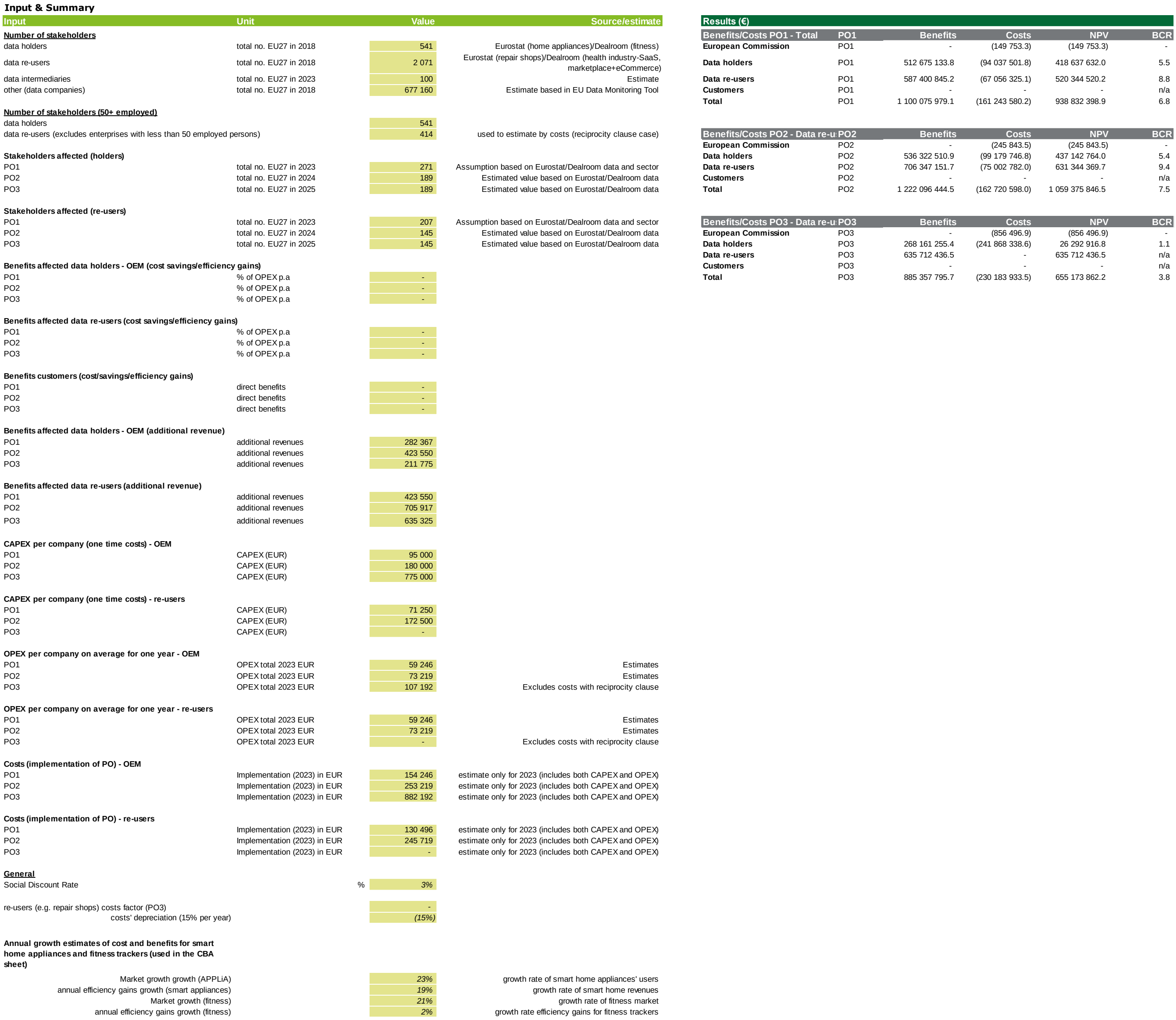
Benefits/Costs PO3 - Data re-u	PO3	Benefits	Costs	NPV	BCR
European Commission	PO3	-	(856 496.9)	(856 496.9)	-
Data holders	PO3	-	(3 967 152 643.6)	(3 967 152 643.6)	-
Data re-users	PO3	426 858 416.8	-	426 858 416.8	n/a
Customers	PO3	8 655 646.6	-	8 655 646.6	n/a
Total	PO3	422 829 187.7	(3 762 311 806.3)	(3 339 482 618.6)	0.1

The figure below presents the cost-benefit analysis for Measures supporting citizen empowerment (‘human-centric data economy’) a) “Smart home appliances”.

Cost-Benefit Analysis

€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Benefits	PO1	Total	Data holders	no.	no.	n/a	n/a	-	310	310	310	310	310
	PO1	Total	Data holders	efficiency gains	efficiency gains	-	-	-	0	0	0	0	0
	PO1	Total	Data holders	additional revenue	additional revenue	178 604.5	203 299.4	-	25 665	31 615	38 945	47 975	59 099
	PO1	Total	Data holders	Benefits	Benefits	**55 367 404.3**	**63 022 818.1**	**-**	**7 956 000**	**9 800 687**	**12 073 084**	**14 872 363**	**18 320 685**
	PO1	Total	Data re-users	no.	no.	n/a	n/a	-	25 066	25 066	25 066	25 066	25 066
	PO1	Total	Data re-users	efficiency gains	efficiency gains	-	-	-	0	0	0	0	0
	PO1	Total	Data re-users	additional revenue	additional revenue	16 528.1	18 813.4	-	2 375	2 926	3 604	4 440	5 469
	PO1	Total	Data re-users	Benefits	Benefits	**414 293 423**	**471 576 000**	**-**	**59 531 750**	**73 334 845**	**90 338 341**	**111 284 285**	**137 086 779**
	PO1	Total	Customers	Benefits	Benefits	**840 354**	**945 000**	**-**	**189 000**	**189 000**	**189 000**	**189 000**	**189 000**
Costs	PO1	Total	Data holders	Costs	Implementation	(107 570 574.0)	(115 929 799.8)	(47 816 229.0)	-18 366 229	-15 611 295	-13 269 600	-11 279 160	-9 587 286
	PO1	Total	Data re-users	Costs	Implementation	(5 183 097.6)	(5 603 473.5)	(2 087 934.4)	-947 934	-805 744	-684 883	-582 150	-494 828
	PO1	Total	European Commission	Costs	Implementation	(149 753.3)	(154 245.9)	(154 245.9)	-	-	-	-	-
Costs total	PO1	Total	Total	Costs	Costs total	**(112 903 424.9)**	**(121 687 519.3)**	**(50 058 409.3)**	**(19 314 163.4)**	**(16 417 038.9)**	**(13 954 483.1)**	**(11 861 310.6)**	**(10 082 114.0)**
Benefits total	PO1	Total	Total	Benefits	Benefits total	**470 501 180.8**	**535 543 817.9**	**-**	**67 676 750.0**	**83 324 531.9**	**102 600 425.2**	**126 345 647.8**	**155 596 463.1**
Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	**357 597 755.9**	**413 856 298.7**	**(50 058 409.3)**	**48 362 586.6**	**66 907 493.0**	**88 645 942.2**	**114 484 337.2**	**145 514 349.0**
Benefit/Cost-ratio	PO1	Total	Benefit/Cost-ratio	BCR	BCR	**4.17**							
Benefits	PO2	Total	Data holders	no.	no.	n/a	n/a	-	3 100	3 100	3 100	3 100	3 100
	PO2	Total	Data holders	efficiency gains %	efficiency gains	-	-	-	0	0	0	0	0
	PO2	Total	Data holders	OPEX	additional revenue	89 302	101 649.7	-	12 832	15 808	19 473	23 988	29 549
	PO2	Total	Data holders	Benefits	Benefits	276 837 021	315 114 090.7	-	**39 780 000**	**49 003 433**	**60 365 422**	**74 361 813**	**91 603 423**
	PO2	Total	Data re-users	no.	no.				25 066	25 066	25 066	25 066	25 066
	PO2	Total	Data re-users	efficiency gains %	efficiency gains				0	0	0	0	0
	PO2	Total	Data re-users	OPEX	additional revenue	8 310	9 176.5		1 188	1 463	1 773	2 149	2 604
	PO2	Total	Data re-users	Benefits	Benefits	208 299 392	230 017 821.5		**29 765 875**	**36 667 423**	**44 440 916**	**53 862 390**	**65 281 217**
	PO2	Total	Customers	Benefits	Benefits	**8 655 647**	**9 450 000.0**		**1 890 000**	**1 890 000**	**1 890 000**	**1 890 000**	**1 890 000**
Costs	PO2	Total	European Commission	Costs	Implementation	(294 387.2)	(303 218.9)	(303 218.9)	-	-	-	-	-
	PO2	Total	Data holders	Costs	Implementation	(1 781 757 752.1)	(1 781 757 752.1)	(939 978 435.0)	-226 978 435	-192 931 670	-163 991 919	-139 393 131	-118 484 162
	PO2	Total	Data re-users	Costs	Implementation	(8 276 169.0)	(8 276 169.0)	(3 931 501.6)	-1 171 502	-995 776	-846 410	-719 448	-611 531
Costs total	PO2	Total	Total	Costs	Costs total	**(1 676 295 913.6)**	**(1 790 337 140.0)**	**(944 213 155.5)**	**(228 149 936.6)**	**(193 927 446.1)**	**(164 838 329.2)**	**(140 112 579.8)**	**(119 095 692.8)**
Benefits total	PO2	Total	Total	Benefits	Benefits total	**487 472 981.0**	**554 581 912.2**	**-**	**71 435 875.0**	**87 560 855.9**	**106 696 338.0**	**130 114 203.5**	**158 774 639.8**
Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	**(1 188 822 932.6)**	**(1 235 755 227.8)**	**(944 213 155.5)**	**(156 714 061.6)**	**(106 366 590.2)**	**(58 141 991.2)**	**(9 998 376.3)**	**39 678 946.9**
Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	**0.29**							
Benefits	PO3	Total	Data holders	no.	no.	n/a	n/a	-	3 100	3 100	3 100	3 100	3 100
	PO3	Total	Data holders	efficiency gains %	efficiency gains	-	-	-	0	0	0	0	0
	PO3	Total	Data holders	OPEX	additional revenue	-	-	-	0	0	0	0	0
	PO3	Total	Data holders	Benefits	Benefits	-	-	-	**0**	**0**	**0**	**0**	**0**
	PO3	Total	Data re-users	no.	no.				25 074	25 074	25 074	25 074	25 074
	PO3	Total	Data re-users	efficiency gains %	efficiency gains	-	-		0	0	0	0	0
	PO3	Total	Data re-users	OPEX	additional revenue	17 024	18 813.4		2 375	2 926	3 604	4 440	5 469
	PO3	Total	Data re-users	Benefits	Benefits	426 858 417	471 726 506.8		**59 550 750**	**73 358 251**	**90 367 173**	**111 319 802**	**137 130 531**
	PO3	Total	Customers	Benefits	Benefits	**8 655 647**	**9 450 000.0**		**1 890 000**	**1 890 000**	**1 890 000**	**1 890 000**	**1 890 000**
Costs	PO3	Total	European Commission	Costs	Implementation	(856 496.9)	(882 191.8)	(882 191.8)	-	-	-	-	-
	PO3	Total	Data holders	Costs	Implementation	(3 967 152 643.6)	(3 967 152 643.6)	(2 734 794 580.0)	-332 294 580	-282 450 393	-240 082 834	-204 070 409	-173 459 848
	PO3	Total	Data re-users	Costs	Implementation	-	-	-	0	0	0	0	0
Costs total	PO3	Total	Total	Costs	Costs total	**(3 762 311 806.3)**	**(3 968 034 835.4)**	**(2 735 676 771.8)**	**(332 294 580.0)**	**(282 450 393.0)**	**(240 082 834.1)**	**(204 070 408.9)**	**(173 459 847.6)**
Benefits total	PO3	Total	Total	Benefits	Benefits total	**422 829 187.7**	**481 176 506.8**	**-**	**61 440 750.0**	**75 248 250.6**	**92 257 173.0**	**113 209 802.4**	**139 020 530.8**
Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	**(3 339 482 618.6)**	**(3 486 858 328.6)**	**(2 735 676 771.8)**	**(270 853 830.0)**	**(207 202 142.4)**	**(147 825 661.0)**	**(90 860 606.6)**	**(34 439 316.8)**
Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	**0.11**							

The figure below presents the input summary for the cost-benefit analysis for Measures supporting citizen empowerment (‘human-centric data economy’) b) “Fitness tracker”.



Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders	total no. EU27 in 2018	541	Eurostat (home appliances)/Dealroom (fitness)
data re-users	total no. EU27 in 2018	2 071	Eurostat (repair shops)/Dealroom (health industry-SaaS, marketplace+eCommerce)
data intermediaries	total no. EU27 in 2023	100	Estimate
other (data companies)	total no. EU27 in 2018	677 160	Estimate based in EU Data Monitoring Tool
Number of stakeholders (50+ employed)			
data holders		541	
data re-users (excludes enterprises with less than 50 employed persons)		414	used to estimate by costs (reciprocity clause case)
Stakeholders affected (holders)			
PO1	total no. EU27 in 2023	271	Assumption based on Eurostat/Dealroom data and sector
PO2	total no. EU27 in 2024	189	Estimated value based on Eurostat/Dealroom data
PO3	total no. EU27 in 2025	189	Estimated value based on Eurostat/Dealroom data
Stakeholders affected (re-users)			
PO1	total no. EU27 in 2023	207	Assumption based on Eurostat/Dealroom data and sector
PO2	total no. EU27 in 2024	145	Estimated value based on Eurostat/Dealroom data
PO3	total no. EU27 in 2025	145	Estimated value based on Eurostat/Dealroom data
Benefits affected data holders - OEM (cost savings/efficiency gains)			
PO1	% of OPEX p.a	-	
PO2	% of OPEX p.a	-	
PO3	% of OPEX p.a	-	
Benefits affected data re-users (cost savings/efficiency gains)			
PO1	% of OPEX p.a	-	
PO2	% of OPEX p.a	-	
PO3	% of OPEX p.a	-	
Benefits customers (cost/savings/efficiency gains)			
PO1	direct benefits	-	
PO2	direct benefits	-	
PO3	direct benefits	-	
Benefits affected data holders - OEM (additional revenue)			
PO1	additional revenues	282 367	
PO2	additional revenues	423 550	
PO3	additional revenues	211 775	
Benefits affected data re-users (additional revenue)			
PO1	additional revenues	423 550	
PO2	additional revenues	705 917	
PO3	additional revenues	635 325	
CAPEX per company (one time costs) - OEM			
PO1	CAPEX (EUR)	95 000	
PO2	CAPEX (EUR)	180 000	
PO3	CAPEX (EUR)	775 000	
CAPEX per company (one time costs) - re-users			
PO1	CAPEX (EUR)	71 250	
PO2	CAPEX (EUR)	172 500	
PO3	CAPEX (EUR)	-	
OPEX per company on average for one year - OEM			
PO1	OPEX total 2023 EUR	59 246	Estimates
PO2	OPEX total 2023 EUR	73 219	Estimates
PO3	OPEX total 2023 EUR	107 192	Excludes costs with reciprocity clause
OPEX per company on average for one year - re-users			
PO1	OPEX total 2023 EUR	59 246	Estimates
PO2	OPEX total 2023 EUR	73 219	Estimates
PO3	OPEX total 2023 EUR	-	Excludes costs with reciprocity clause
Costs (implementation of PO) - OEM			
PO1	Implementation (2023) in EUR	154 246	estimate only for 2023 (includes both CAPEX and OPEX)
PO2	Implementation (2023) in EUR	253 219	estimate only for 2023 (includes both CAPEX and OPEX)
PO3	Implementation (2023) in EUR	882 192	estimate only for 2023 (includes both CAPEX and OPEX)
Costs (implementation of PO) - re-users			
PO1	Implementation (2023) in EUR	130 496	estimate only for 2023 (includes both CAPEX and OPEX)
PO2	Implementation (2023) in EUR	245 719	estimate only for 2023 (includes both CAPEX and OPEX)
PO3	Implementation (2023) in EUR	-	estimate only for 2023 (includes both CAPEX and OPEX)
General			
Social Discount Rate	%	*3%*	
re-users (e.g. repair shops) costs factor (PO3)		-	
costs' depreciation (15% per year)		*(15%)*	
Annual growth estimates of cost and benefits for smart home appliances and fitness trackers (used in the CBA sheet)			
Market growth growth (APPLiA)		*23%*	growth rate of smart home appliances' users
annual efficiency gains growth (smart appliances)		*19%*	growth rate of smart home revenues
Market growth (fitness)		*21%*	growth rate of fitness market
annual efficiency gains growth (fitness)		*2%*	growth rate efficiency gains for fitness trackers

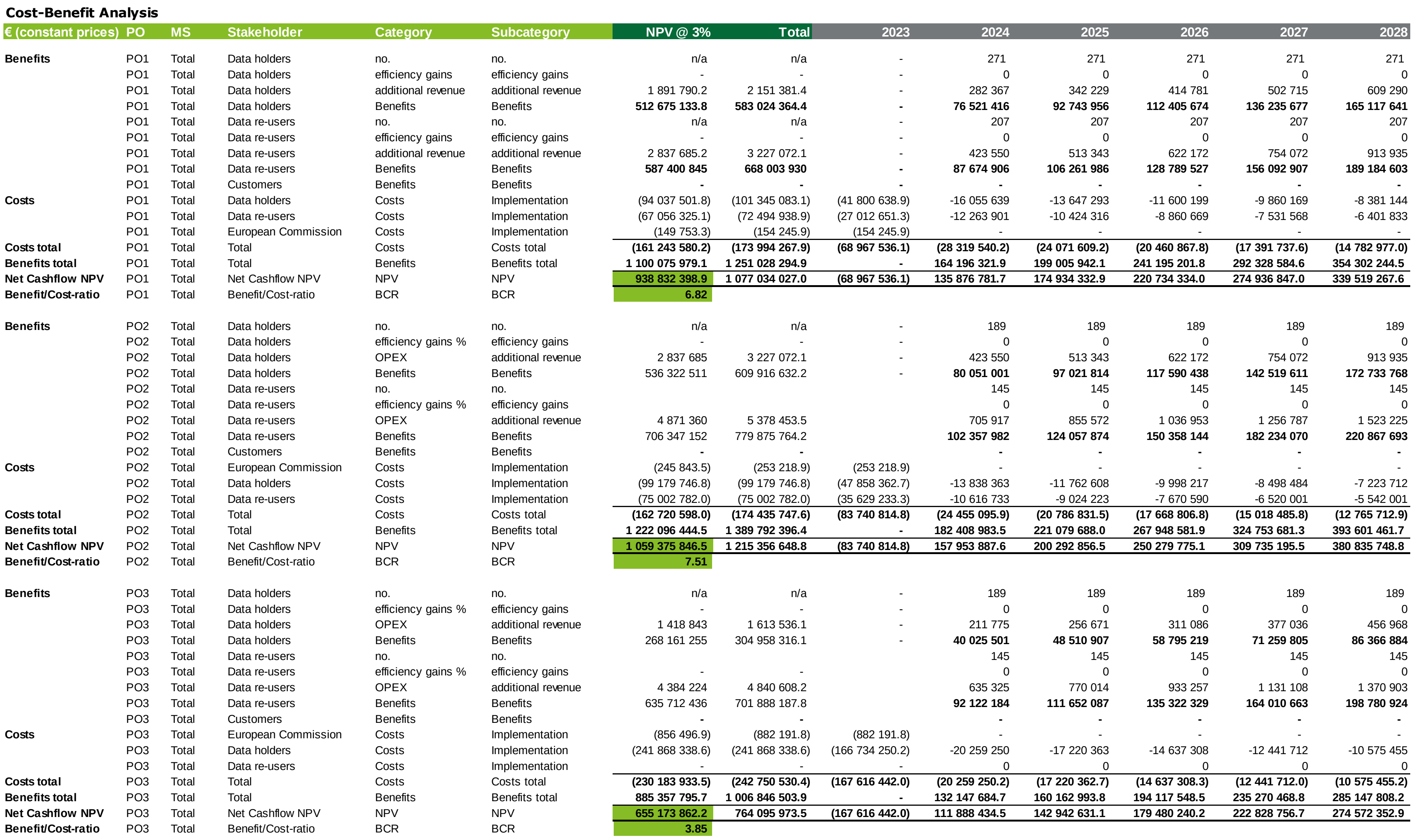
Results (€)

Benefits/Costs PO1 - Total	PO1	Benefits	Costs	NPV	BCR
European Commission	PO1	-	(149 753.3)	(149 753.3)	-
Data holders	PO1	512 675 133.8	(94 037 501.8)	418 637 632.0	5.5
Data re-users	PO1	587 400 845.2	(67 056 325.1)	520 344 520.2	8.8
Customers	PO1	-	-	-	n/a
Total	PO1	1 100 075 979.1	(161 243 580.2)	938 832 398.9	6.8

Benefits/Costs PO2 - Data re-u	PO2	Benefits	Costs	NPV	BCR
European Commission	PO2	-	(245 843.5)	(245 843.5)	-
Data holders	PO2	536 322 510.9	(99 179 746.8)	437 142 764.0	5.4
Data re-users	PO2	706 347 151.7	(75 002 782.0)	631 344 369.7	9.4
Customers	PO2	-	-	-	n/a
Total	PO2	1 222 096 444.5	(162 720 598.0)	1 059 375 846.5	7.5

Benefits/Costs PO3 - Data re-u	PO3	Benefits	Costs	NPV	BCR
European Commission	PO3	-	(856 496.9)	(856 496.9)	-
Data holders	PO3	268 161 255.4	(241 868 338.6)	26 292 916.8	1.1
Data re-users	PO3	635 712 436.5	-	635 712 436.5	n/a
Customers	PO3	-	-	-	n/a
Total	PO3	885 357 795.7	(230 183 933.5)	655 173 862.2	3.8

The figure below presents the cost-benefit analysis for Measures supporting citizen empowerment (‘human-centric data economy’) b) “Fitness tracker”.

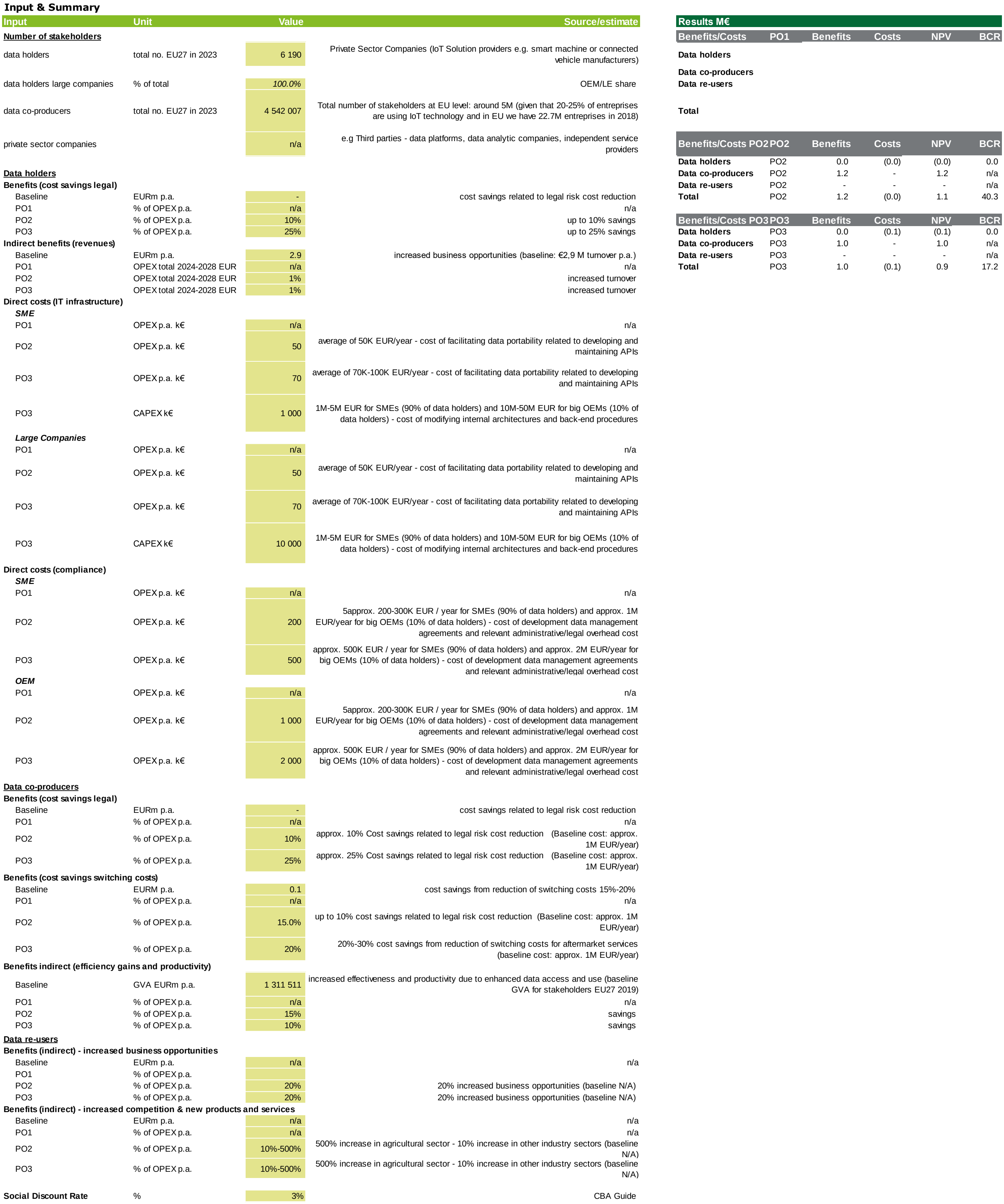


Cost-Benefit Analysis

€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Benefits	PO1	Total	Data holders	no.	no.	n/a	n/a	-	271	271	271	271	271
	PO1	Total	Data holders	efficiency gains	efficiency gains	-	-	-	0	0	0	0	0
	PO1	Total	Data holders	additional revenue	additional revenue	1 891 790.2	2 151 381.4	-	282 367	342 229	414 781	502 715	609 290
	PO1	Total	Data holders	Benefits	Benefits	512 675 133.8	583 024 364.4	-	76 521 416	92 743 956	112 405 674	136 235 677	165 117 641
	PO1	Total	Data re-users	no.	no.	n/a	n/a	-	207	207	207	207	207
	PO1	Total	Data re-users	efficiency gains	efficiency gains	-	-	-	0	0	0	0	0
	PO1	Total	Data re-users	additional revenue	additional revenue	2 837 685.2	3 227 072.1	-	423 550	513 343	622 172	754 072	913 935
	PO1	Total	Data re-users	Benefits	Benefits	587 400 845	668 003 930	-	87 674 906	106 261 986	128 789 527	156 092 907	189 184 603
	PO1	Total	Customers	Benefits	Benefits	-	-	-	-	-	-	-	-
Costs	PO1	Total	Data holders	Costs	Implementation	(94 037 501.8)	(101 345 083.1)	(41 800 638.9)	-16 055 639	-13 647 293	-11 600 199	-9 860 169	-8 381 144
	PO1	Total	Data re-users	Costs	Implementation	(67 056 325.1)	(72 494 938.9)	(27 012 651.3)	-12 263 901	-10 424 316	-8 860 669	-7 531 568	-6 401 833
	PO1	Total	European Commission	Costs	Implementation	(149 753.3)	(154 245.9)	(154 245.9)	-	-	-	-	-
Costs total	PO1	Total	Total	Costs	Costs total	(161 243 580.2)	(173 994 267.9)	(68 967 536.1)	(28 319 540.2)	(24 071 609.2)	(20 460 867.8)	(17 391 737.6)	(14 782 977.0)
Benefits total	PO1	Total	Total	Benefits	Benefits total	1 100 075 979.1	1 251 028 294.9	-	164 196 321.9	199 005 942.1	241 195 201.8	292 328 584.6	354 302 244.5
Net Cashflow NPV	PO1	Total	Net Cashflow NPV	NPV	NPV	938 832 398.9	1 077 034 027.0	(68 967 536.1)	135 876 781.7	174 934 332.9	220 734 334.0	274 936 847.0	339 519 267.6
Benefit/Cost-ratio	PO1	Total	Benefit/Cost-ratio	BCR	BCR	6.82							
Benefits	PO2	Total	Data holders	no.	no.	n/a	n/a	-	189	189	189	189	189
	PO2	Total	Data holders	efficiency gains %	efficiency gains	-	-	-	0	0	0	0	0
	PO2	Total	Data holders	OPEX	additional revenue	2 837 685	3 227 072.1	-	423 550	513 343	622 172	754 072	913 935
	PO2	Total	Data holders	Benefits	Benefits	536 322 511	609 916 632.2	-	80 051 001	97 021 814	117 590 438	142 519 611	172 733 768
	PO2	Total	Data re-users	no.	no.				145	145	145	145	145
	PO2	Total	Data re-users	efficiency gains %	efficiency gains				0	0	0	0	0
	PO2	Total	Data re-users	OPEX	additional revenue	4 871 360	5 378 453.5		705 917	855 572	1 036 953	1 256 787	1 523 225
	PO2	Total	Data re-users	Benefits	Benefits	706 347 152	779 875 764.2		102 357 982	124 057 874	150 358 144	182 234 070	220 867 693
	PO2	Total	Customers	Benefits	Benefits	-	-		-	-	-	-	-
Costs	PO2	Total	European Commission	Costs	Implementation	(245 843.5)	(253 218.9)	(253 218.9)	-	-	-	-	-
	PO2	Total	Data holders	Costs	Implementation	(99 179 746.8)	(99 179 746.8)	(47 858 362.7)	-13 838 363	-11 762 608	-9 998 217	-8 498 484	-7 223 712
	PO2	Total	Data re-users	Costs	Implementation	(75 002 782.0)	(75 002 782.0)	(35 629 233.3)	-10 616 733	-9 024 223	-7 670 590	-6 520 001	-5 542 001
Costs total	PO2	Total	Total	Costs	Costs total	(162 720 598.0)	(174 435 747.6)	(83 740 814.8)	(24 455 095.9)	(20 786 831.5)	(17 668 806.8)	(15 018 485.8)	(12 765 712.9)
Benefits total	PO2	Total	Total	Benefits	Benefits total	1 222 096 444.5	1 389 792 396.4	-	182 408 983.5	221 079 688.0	267 948 581.9	324 753 681.3	393 601 461.7
Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	1 059 375 846.5	1 215 356 648.8	(83 740 814.8)	157 953 887.6	200 292 856.5	250 279 775.1	309 735 195.5	380 835 748.8
Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	7.51							
Benefits	PO3	Total	Data holders	no.	no.	n/a	n/a	-	189	189	189	189	189
	PO3	Total	Data holders	efficiency gains %	efficiency gains	-	-	-	0	0	0	0	0
	PO3	Total	Data holders	OPEX	additional revenue	1 418 843	1 613 536.1	-	211 775	256 671	311 086	377 036	456 968
	PO3	Total	Data holders	Benefits	Benefits	268 161 255	304 958 316.1	-	40 025 501	48 510 907	58 795 219	71 259 805	86 366 884
	PO3	Total	Data re-users	no.	no.				145	145	145	145	145
	PO3	Total	Data re-users	efficiency gains %	efficiency gains	-	-		0	0	0	0	0
	PO3	Total	Data re-users	OPEX	additional revenue	4 384 224	4 840 608.2		635 325	770 014	933 257	1 131 108	1 370 903
	PO3	Total	Data re-users	Benefits	Benefits	635 712 436	701 888 187.8		92 122 184	111 652 087	135 322 329	164 010 663	198 780 924
	PO3	Total	Customers	Benefits	Benefits	-	-		-	-	-	-	-
Costs	PO3	Total	European Commission	Costs	Implementation	(856 496.9)	(882 191.8)	(882 191.8)	-	-	-	-	-
	PO3	Total	Data holders	Costs	Implementation	(241 868 338.6)	(241 868 338.6)	(166 734 250.2)	-20 259 250	-17 220 363	-14 637 308	-12 441 712	-10 575 455
	PO3	Total	Data re-users	Costs	Implementation	-	-	-	0	0	0	0	0
Costs total	PO3	Total	Total	Costs	Costs total	(230 183 933.5)	(242 750 530.4)	(167 616 442.0)	(20 259 250.2)	(17 220 362.7)	(14 637 308.3)	(12 441 712.0)	(10 575 455.2)
Benefits total	PO3	Total	Total	Benefits	Benefits total	885 357 795.7	1 006 846 503.9	-	132 147 684.7	160 162 993.8	194 117 548.5	235 270 468.8	285 147 808.2
Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	655 173 862.2	764 095 973.5	(167 616 442.0)	111 888 434.5	142 942 631.1	179 480 240.2	222 828 756.7	274 572 352.9
Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	3.85							

4.2.1.4 Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing

The figure below presents the input summary for the cost-benefit analysis for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.



Input & Summary

Input	Unit	Value	Source/estimate
Number of stakeholders			
data holders	total no. EU27 in 2023	6 190	Private Sector Companies (IoT Solution providers e.g. smart machine or connected vehicle manufacturers)
data holders large companies	% of total	*100.0%*	OEM/LE share
data co-producers	total no. EU27 in 2023	4 542 007	Total number of stakeholders at EU level: around 5M (given that 20-25% of entreprises are using IoT technology and in EU we have 22.7M entreprises in 2018)
private sector companies		n/a	e.g Third parties - data platforms, data analytic companies, independent service providers
Data holders			
Benefits (cost savings legal)			
Baseline	EURm p.a.	-	cost savings related to legal risk cost reduction
PO1	% of OPEX p.a.	n/a	n/a
PO2	% of OPEX p.a.	10%	up to 10% savings
PO3	% of OPEX p.a.	25%	up to 25% savings
Indirect benefits (revenues)			
Baseline	EURm p.a.	2.9	increased business opportunities (baseline: €2,9 M turnover p.a.)
PO1	OPEX total 2024-2028 EUR	n/a	n/a
PO2	OPEX total 2024-2028 EUR	1%	increased turnover
PO3	OPEX total 2024-2028 EUR	1%	increased turnover
Direct costs (IT infrastructure)			
SME			
PO1	OPEX p.a. k€	n/a	n/a
PO2	OPEX p.a. k€	50	average of 50K EUR/year - cost of facilitating data portability related to developing and maintaining APIs
PO3	OPEX p.a. k€	70	average of 70K-100K EUR/year - cost of facilitating data portability related to developing and maintaining APIs
PO3	CAPEX k€	1 000	1M-5M EUR for SMEs (90% of data holders) and 10M-50M EUR for big OEMs (10% of data holders) - cost of modifying internal architectures and back-end procedures
Large Companies			
PO1	OPEX p.a. k€	n/a	n/a
PO2	OPEX p.a. k€	50	average of 50K EUR/year - cost of facilitating data portability related to developing and maintaining APIs
PO3	OPEX p.a. k€	70	average of 70K-100K EUR/year - cost of facilitating data portability related to developing and maintaining APIs
PO3	CAPEX k€	10 000	1M-5M EUR for SMEs (90% of data holders) and 10M-50M EUR for big OEMs (10% of data holders) - cost of modifying internal architectures and back-end procedures
Direct costs (compliance)			
SME			
PO1	OPEX p.a. k€	n/a	n/a
PO2	OPEX p.a. k€	200	5approx. 200-300K EUR / year for SMEs (90% of data holders) and approx. 1M EUR/year for big OEMs (10% of data holders) - cost of development data management agreements and relevant administrative/legal overhead cost
PO3	OPEX p.a. k€	500	approx. 500K EUR / year for SMEs (90% of data holders) and approx. 2M EUR/year for big OEMs (10% of data holders) - cost of development data management agreements and relevant administrative/legal overhead cost
OEM			
PO1	OPEX p.a. k€	n/a	n/a
PO2	OPEX p.a. k€	1 000	5approx. 200-300K EUR / year for SMEs (90% of data holders) and approx. 1M EUR/year for big OEMs (10% of data holders) - cost of development data management agreements and relevant administrative/legal overhead cost
PO3	OPEX p.a. k€	2 000	approx. 500K EUR / year for SMEs (90% of data holders) and approx. 2M EUR/year for big OEMs (10% of data holders) - cost of development data management agreements and relevant administrative/legal overhead cost
Data co-producers			
Benefits (cost savings legal)			
Baseline	EURm p.a.	-	cost savings related to legal risk cost reduction
PO1	% of OPEX p.a.	n/a	n/a
PO2	% of OPEX p.a.	10%	approx. 10% Cost savings related to legal risk cost reduction (Baseline cost: approx. 1M EUR/year)
PO3	% of OPEX p.a.	25%	approx. 25% Cost savings related to legal risk cost reduction (Baseline cost: approx. 1M EUR/year)
Benefits (cost savings switching costs)			
Baseline	EURM p.a.	0.1	cost savings from reduction of switching costs 15%-20%
PO1	% of OPEX p.a.	n/a	n/a
PO2	% of OPEX p.a.	15.0%	up to 10% cost savings related to legal risk cost reduction (Baseline cost: approx. 1M EUR/year)
PO3	% of OPEX p.a.	20%	20%-30% cost savings from reduction of switching costs for aftermarket services (baseline cost: approx. 1M EUR/year)
Benefits indirect (efficiency gains and productivity)			
Baseline	GVA EURm p.a.	1 311 511	increased effectiveness and productivity due to enhanced data access and use (baseline GVA for stakeholders EU27 2019)
PO1	% of OPEX p.a.	n/a	n/a
PO2	% of OPEX p.a.	15%	savings
PO3	% of OPEX p.a.	10%	savings
Data re-users			
Benefits (indirect) - increased business opportunities			
Baseline	EURm p.a.	n/a	n/a
PO1	% of OPEX p.a.		
PO2	% of OPEX p.a.	20%	20% increased business opportunities (baseline N/A)
PO3	% of OPEX p.a.	20%	20% increased business opportunities (baseline N/A)
Benefits (indirect) - increased competition & new products and services			
Baseline	EURm p.a.	n/a	n/a
PO1	% of OPEX p.a.	n/a	n/a
PO2	% of OPEX p.a.	10%-500%	500% increase in agricultural sector - 10% increase in other industry sectors (baseline N/A)
PO3	% of OPEX p.a.	10%-500%	500% increase in agricultural sector - 10% increase in other industry sectors (baseline N/A)
Social Discount Rate	%	3%	CBA Guide

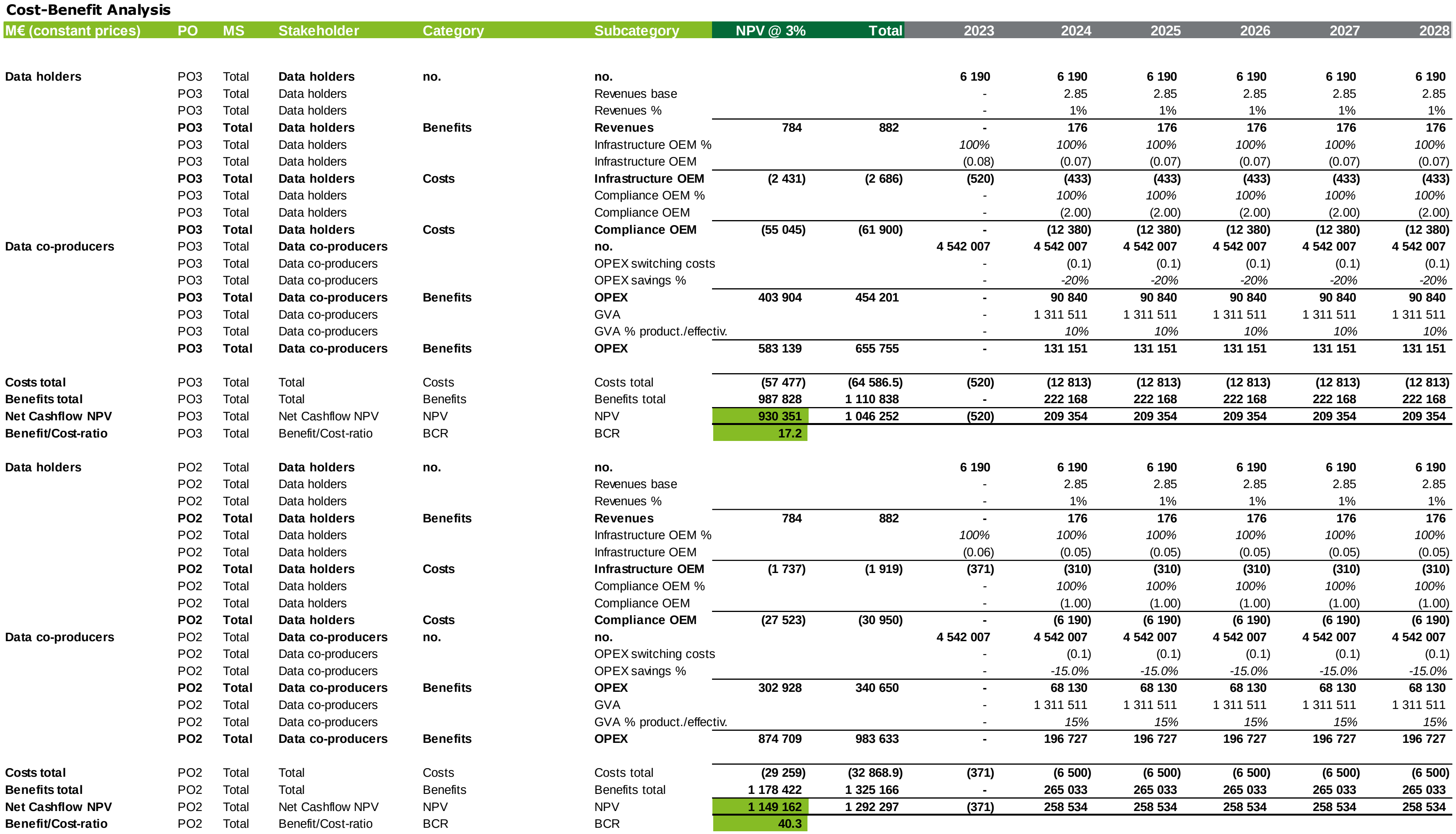
Results M€

Benefits/Costs	PO1	Benefits	Costs	NPV	BCR
Data holders					
Data co-producers					
Data re-users					
Total					

Benefits/Costs PO2	PO2	Benefits	Costs	NPV	BCR
Data holders	PO2	0.0	(0.0)	(0.0)	0.0
Data co-producers	PO2	1.2	-	1.2	n/a
Data re-users	PO2	-	-	-	n/a
Total	PO2	1.2	(0.0)	1.1	40.3

Benefits/Costs PO3	PO3	Benefits	Costs	NPV	BCR
Data holders	PO3	0.0	(0.1)	(0.1)	0.0
Data co-producers	PO3	1.0	-	1.0	n/a
Data re-users	PO3	-	-	-	n/a
Total	PO3	1.0	(0.1)	0.9	17.2

The figure below presents the cost-benefit analysis for Measures clarifying and potentially further developing rights on co-generated data and business-to-business data sharing.



Cost-Benefit Analysis

M€ (constant prices)	PO	MS	Stakeholder	Category	Subcategory	NPV @ 3%	Total	2023	2024	2025	2026	2027	2028
Data holders	PO3	Total	Data holders	no.	no.			6 190	6 190	6 190	6 190	6 190	6 190
	PO3	Total	Data holders		Revenues base			-	2.85	2.85	2.85	2.85	2.85
	PO3	Total	Data holders		Revenues %			-	1%	1%	1%	1%	1%
	PO3	Total	Data holders	Benefits	Revenues	784	882	-	176	176	176	176	176
	PO3	Total	Data holders		Infrastructure OEM %			100%	100%	100%	100%	100%	100%
	PO3	Total	Data holders		Infrastructure OEM			(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	PO3	Total	Data holders	Costs	Infrastructure OEM	(2 431)	(2 686)	(520)	(433)	(433)	(433)	(433)	(433)
	PO3	Total	Data holders		Compliance OEM %			-	100%	100%	100%	100%	100%
	PO3	Total	Data holders		Compliance OEM			-	(2.00)	(2.00)	(2.00)	(2.00)	(2.00)
	PO3	Total	Data holders	Costs	Compliance OEM	(55 045)	(61 900)	-	(12 380)	(12 380)	(12 380)	(12 380)	(12 380)
Data co-producers	PO3	Total	Data co-producers		no.			4 542 007	4 542 007	4 542 007	4 542 007	4 542 007	4 542 007
	PO3	Total	Data co-producers		OPEX switching costs			-	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	PO3	Total	Data co-producers		OPEX savings %			-	-20%	-20%	-20%	-20%	-20%
	PO3	Total	Data co-producers	Benefits	OPEX	403 904	454 201	-	90 840	90 840	90 840	90 840	90 840
	PO3	Total	Data co-producers		GVA			-	1 311 511	1 311 511	1 311 511	1 311 511	1 311 511
	PO3	Total	Data co-producers		GVA % product./effectiv.			-	10%	10%	10%	10%	10%
	PO3	Total	Data co-producers	Benefits	OPEX	583 139	655 755	-	131 151	131 151	131 151	131 151	131 151
Costs total	PO3	Total	Total	Costs	Costs total	(57 477)	(64 586.5)	(520)	(12 813)	(12 813)	(12 813)	(12 813)	(12 813)
Benefits total	PO3	Total	Total	Benefits	Benefits total	987 828	1 110 838	-	222 168	222 168	222 168	222 168	222 168
Net Cashflow NPV	PO3	Total	Net Cashflow NPV	NPV	NPV	930 351	1 046 252	(520)	209 354	209 354	209 354	209 354	209 354
Benefit/Cost-ratio	PO3	Total	Benefit/Cost-ratio	BCR	BCR	17.2							
Data holders	PO2	Total	Data holders	no.	no.			6 190	6 190	6 190	6 190	6 190	6 190
	PO2	Total	Data holders		Revenues base			-	2.85	2.85	2.85	2.85	2.85
	PO2	Total	Data holders		Revenues %			-	1%	1%	1%	1%	1%
	PO2	Total	Data holders	Benefits	Revenues	784	882	-	176	176	176	176	176
	PO2	Total	Data holders		Infrastructure OEM %			100%	100%	100%	100%	100%	100%
	PO2	Total	Data holders		Infrastructure OEM			(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
	PO2	Total	Data holders	Costs	Infrastructure OEM	(1 737)	(1 919)	(371)	(310)	(310)	(310)	(310)	(310)
	PO2	Total	Data holders		Compliance OEM %			-	100%	100%	100%	100%	100%
	PO2	Total	Data holders		Compliance OEM			-	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
	PO2	Total	Data holders	Costs	Compliance OEM	(27 523)	(30 950)	-	(6 190)	(6 190)	(6 190)	(6 190)	(6 190)
Data co-producers	PO2	Total	Data co-producers	no.	no.			4 542 007	4 542 007	4 542 007	4 542 007	4 542 007	4 542 007
	PO2	Total	Data co-producers		OPEX switching costs			-	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	PO2	Total	Data co-producers		OPEX savings %			-	-15.0%	-15.0%	-15.0%	-15.0%	-15.0%
	PO2	Total	Data co-producers	Benefits	OPEX	302 928	340 650	-	68 130	68 130	68 130	68 130	68 130
	PO2	Total	Data co-producers		GVA			-	1 311 511	1 311 511	1 311 511	1 311 511	1 311 511
	PO2	Total	Data co-producers		GVA % product./effectiv.			-	15%	15%	15%	15%	15%
	PO2	Total	Data co-producers	Benefits	OPEX	874 709	983 633	-	196 727	196 727	196 727	196 727	196 727
Costs total	PO2	Total	Total	Costs	Costs total	(29 259)	(32 868.9)	(371)	(6 500)	(6 500)	(6 500)	(6 500)	(6 500)
Benefits total	PO2	Total	Total	Benefits	Benefits total	1 178 422	1 325 166	-	265 033	265 033	265 033	265 033	265 033
Net Cashflow NPV	PO2	Total	Net Cashflow NPV	NPV	NPV	1 149 162	1 292 297	(371)	258 534	258 534	258 534	258 534	258 534
Benefit/Cost-ratio	PO2	Total	Benefit/Cost-ratio	BCR	BCR	40.3							

4.2.2 Business-to-Government data sharing for the public interest | Legal analysis

4.2.2.1 Context – requirements for a legal framework

As has been explained in the preceding sections of this report, the central objective is to enable and promote business-to-government data sharing under fair, reliable and transparent terms. More precisely, the desired outcome is a framework that enables access to and use of (big) data sources held by private companies, if that data is potentially valuable for innovative government uses and for better policymaking, in a flexible manner and in compliance with European data protection law.

These objectives don't necessarily depend on the creation of a legislative framework. Nonetheless, it is important to examine the legal context in which the B2G discussion takes place, and to highlight some of the main legal barriers that need to be overcome. This legal analysis aims to pinpoint to what extent non-legislative intervention could be adequate in addressing the main problems, and also to identify relevant legal precursors - i.e. potentially influential existing legal frameworks that shape the scope of B2G data sharing in the absence of legislative intervention, or that could provide useful models to follow if legislative intervention would be chosen in the future. The final result should be a summary legal description of the viability, potential conflicts, and approaches for a B2G data sharing framework.

4.2.2.1.1 Essential legal topics

When examining B2G data sharing from a legal perspective, it is important to have a clear understanding of the principal legal challenges that need to be resolved. These stem on the one hand from the need to ensure an unambiguous, clear, proportional and effective legal framework to scope such data sharing requirements; and on the other hand from the need to ensure the compatibility of such a new legal framework with other legal initiatives, either in the sense that there no conflict with other legal frameworks, or in the sense that conflicts can be cleanly resolved by establishing which legislative framework takes precedence. The latter point (the issue of precedence) is particularly important when assessing the viability of nonregulatory intervention, since nonregulatory intervention is in principle incapable of overruling legal requirements imposed by binding law.

By way of a simple example: if data sharing is e.g. impossible in a particular case because it would violate the GDPR's constraints in relation to purpose limitation (as will be explained further below), then that problem cannot be resolved by promoting best practices for data sharing. Such cases (if they exist) would need to be solved through legislative intervention that establishes the intent of the legislator in case of conflicts.

The analysis in earlier sections of the report identified some of the main challenges that also need to be addressed by any future legal framework addressing B2G data sharing (whether regulatory or nonregulatory in nature). Firstly, a clear **definition of B2G data sharing** is required, that recognizes that the fundamental requirement and interest in B2G data sharing lies in ensuring **access and usage rights** for defined entities in the public sector – as opposed to a more constrained interpretation of 'sharing' that might focus on providing copies of data to certain designated entities. Copying is not the sole manner of implementing data sharing procedures, nor is it the optimal strategy for data sharing.

Secondly, the **prerequisites and constraints of both access and usage rights** must be integrated in a legal framework. These include a **demarcation of the public interests** that a public sector body can invoke in order to justify B2G data sharing , a clear **scoping** of the data to be shared, a **definition of the usage rights** of the public sector body (including any purpose limitations and permitted onwards sharing with third parties), the **modalities** of data sharing (whether sharing occurs on an ad hoc basis or continuously; whether an intermediary should be involved), and any **remuneration or compensation** for data sharing. Furthermore, the **quality** (or in a broader sense fitness for purpose) of the data could be regulated, and any

safeguards to be implemented and respected (including by way of non-exhaustive examples any transparency obligations; any impact assessments that validate the likelihood of utility of data sharing and identify ethical challenges; any data minimization obligations that ensure that only the data which is strictly needed for the public interest task is shared; fundamental rights protection including notably data protection compliance; and/or the involvement of data stewards in organisations who are subject to B2G data sharing duties).

4.2.2.1.2 Influential legal precursors

In order to identify and assess what these legal challenges imply in practice, and how they have been addressed in related legal initiatives, it is useful to examine a series of prior legal frameworks at the EU level that impact B2G data sharing possibilities. Data protection law is arguably the most significant influence; but other relevant lessons can be learned from legislation that protects data against certain re-uses (such as the Database Directive or the Trade Secrets Directive); or that focuses on facilitating certain data exchanges (the Open Data Directive and the Free Flow of Data Directive), that already sustains certain B2G data sharing (the Regulation on European statistics and on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities), and on data affecting data sharing modalities such as the emerging Data Governance Act.

In the sections below, we will examine these initiatives briefly, in order to assess what their main legal characteristics and, how they approach data sharing (directly or indirectly, and what relevant lessons could be learned.

4.2.2.1.3 General Data Protection Regulation (EU) 2016/679 (GDPR)[494]

4.2.2.1.3.1 Key characteristics

The GDPR constitutes the EU's principal legal framework in relation to the processing of personal data. It harmonizes the European internal market rules in relation to data processing, by stating the main legal requirements to ensure the lawfulness of processing, as well as introducing a series of fundamental principles for personal data processing (such as transparency, data minimization, proportionality and purpose limitation), harmonizing data subject rights, and introducing a governance framework for compliance based on (among other points) national data protection authorities. Arguably the GDPR's central innovation compared to prior data protection frameworks is the accountability principle, i.e. the obligation for data processing entities to not only comply with the GDPR, but to maintain documentary evidence to show compliance with data protection law.

Data sharing would qualify as processing of personal data if the data to be shared includes personal data. Both accessing the data (or copying it) and any subsequent use of the data would be considered as data processing in the sense of the GDPR.

4.2.2.1.3.2 Relevance to data sharing – access and use

The GDPR is a central enabling and constraining framework in relation to data sharing, including in a B2G context. It is enabling in the sense that it provides a common legal framework across the European Union, thus ensuring that personal data can be exchanged on relatively even terms across the European Union. It is however also constraining, in the sense that it introduces principles and requirements that limit what can lawfully be done with personal data. Specifically, in relation to B2G data sharing, any such data sharing relating to personal data should:

- Have a clear **legal basis**. Potential legal bases are enumerated in the GDPR itself. Data processed by businesses are likely to be based on consent, necessity for performance of a contract or legitimate

[494] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

interest. Barring some more exceptional circumstances[495], the making available of data by a private company to a public sector body requesting the data would probably require a legal basis of the necessity of processing for the performance of a task carried out in the public interest or in the exercise of official authority vested in that body, or compliance with a legal obligation. This, however, implies that a specific legal framework exists in EU or Member State law that establishes and constrains this task or authority, which is a strong argument in favour of a regulatory intervention.

- Be **transparent**, implying that data subjects are informed in an appropriate and accessible manner on the B2G data sharing activities and subsequent use of the data.
- Be **limited to specific purposes**, implying that the purposes of sharing and subsequent use must occur within a legal framework that clearly scopes and demarcates permissible usage of the data.
- Be as **minimal** as possible, taking into account the purposes of the processing; this is an argument in favour of data **access (rather than copying**), and in favour of **targeted data requests (rather than blanket requests** targeting an entire database or data environment), as well as further safeguards such as pseudonymisation and anonymisation.
- Be **accurate and up to date**, implying that clarity should be established on the provided data's factual accuracy and periodicity of updates (if any).
- Respect **storage limitations**, implying that data is not kept in a form permitting identification of the data subjects for a period of time that's longer than required to achieve the stated purposes. This argues in favour of **data deletion** by a requesting public sector body once the requested data has achieved its purpose, **and/or anonymization** requirements in relation to such.
- Be processed in a way that safeguards **integrity and confidentiality**, implying that high level security obligations should apply to any data sharing mechanisms and any subsequent usages.

While the GDPR applies only to personal data and not to any B2G data sharing, the concept of personal data is broad, and therefore will reasonably apply to many B2G data sharing cases. Where data sets include inextricably mixed personal and non-personal data, the GDPR should be presumed to apply[496]. Moreover, the principles above are useful in any context where the data to be shared may be sensitive or confidential, since there too the original data holders would reasonably insist on similar legal constraints to be applied.

4.2.2.1.3.3 Lessons learned and useful inputs

The principles set out above are generally useful as core legal requirements for establishing a legal framework for B2G data sharing, especially when this data comprises or includes personal data. The principles of purpose limitation, minimization and transparency will be critical to ensure trust in a B2G data sharing framework. Moreover, the legal basis requirement highlights that the introduction of a broadly scoped B2G framework that should allow public sector bodies to demand certain personal data to be shared with them will typically require regulatory intervention. In the absence of specific legislation authorizing certain public sector bodies to demand such data to be shared with them, public sector bodies will lack the legal mandate to structurally legitimize the processing of personal data in many use cases.

Several other elements of the GDPR can provide useful building blocks for the creation of a legal framework for B2G data sharing. Potentially relevant elements include:

- The mandatory appointment of a data protection officer (DPO) for certain types of legal entities. This is a **useful model on which a duty to appoint a data steward** can be based. The justification of such a legal obligation is comparable, in the sense that both a DPO and a data steward would be principally charged with providing independent counsel in relation to how data should be treated. Moreover, the example of the GDPR shows that the duty could be scoped by only targeting certain legal entities that

[495] Lawfulness could also be based on the consent of the persons concerned, which would however not be practical in most situations; or on the necessity of processing in order to protect the vital interests of the data subject or of another natural person, which would be useful e.g. in case of pandemics, but would not cover all relevant B2G data sharing cases.

[496] See the Article 2(2) of the Free Flow of Data Regulation

engage in data relevant data processing activities (with relevance being determined by the nature of the data controller or of the processing activities in the case of the GDPR; and by the potential obligation to share data in the case of a B2G data sharing framework).

- The **accountability principle** and the need to assess risk, including in some cases via a formal **data protection impact assessment (DPIA).** While DPIAs are not always required under the GDPR, they are mandatory for processing activities that contain certain risk factors. Moreover, records must be kept of data processing activities, thus facilitating both transparency and accountability.
- The **governance and supervision framework**, based on national supervisory bodies, who supervise compliance with the legal framework, issue guidance, and address complaints (among other tasks).

While these requirements were created specifically to protect the fundamental right to data protection as enshrined (i.a.) in the European Charter of Fundamental Rights, some of this approach may be transposable to a B2G data sharing context where businesses may have legitimate concerns on how their key assets – their data – will be protected against abuses.

4.2.2.1.4 European Statistics Regulation (EC) No 223/2009497

4.2.2.1.4.1 Key characteristics

The European Statistical System (ESS) Regulation provides the central legal framework for the development, production and dissemination of European statistics that complies with the principles of professional independence, impartiality, objectivity, reliability, statistical confidentiality and cost effectiveness. It governs both the activities of Eurostat at the EU level, and also affects the national statistical institutes (NSIs) appointed by the Member States, who act as the national contact points for Eurostat.

The ESS is defined in the Regulation as a partnership between Eurostat, NSIs and other national authorities responsible in each Member State for the development, production and dissemination of European statistics. The Regulation also establishes the ESS Committee composed of NSI representatives and chaired by Eurostat. It works with the Commission on, among other things, the European statistical programme; issues concerning statistical confidentiality; and the further development of a European statistics Code of Practice[498]. This Code acts as the cornerstone of the common quality framework of the ESS, and defines 16 principles covering the institutional environment, statistical processes and statistical outputs.

Under the ESS, NSIs can receive a legal mandate to collect confidential data, which may be used (in principle) exclusively for statistical purposes. Confidential data may be transmitted by an authority of the ESS to another authority as long as this act is recognised as being necessary to the development, production or dissemination of European statistics. Other uses can be permitted under regulated circumstances (e.g. access by researchers for scientific purposes).

4.2.2.1.4.2 Relevance to data sharing – access and use

The relevance of the European Statistics Regulation for B2G sharing is readily apparent, since it is one of the instances where private sector data can be accessed and used for a specifically defined public good – namely the development, production and dissemination of official statistics. While the Regulation is therefore focused on that one specific purpose, some of the elements of the European Statistics Regulation are arguably more broadly applicable, including notably:

- Its approach to describing **permissible use** (notably for statistical purposes, and including exception regimes in case of consent for other use and for scientific research);

[497] See https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009R0223
[498] See https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7

- The introduction of requirements in relation to **professional independence** for the authorities involved in developing, producing and disseminating statistics and ensuring that they are free from any pressures from political or interest groups, or from EU or national authorities.
- The role of **national authorities (NSI), Eurostat, and complementary governance bodies**[499] in the execution of the European Statistics Regulation, and in enabling transmissions of confidential data when this is necessary to achieve the goals of the Regulation.
- The **reliance on non-legislative norms** (the European statistics Code of Practice and the various Member State Commitments on Confidence in Statistics[500]) to set out further details that ensure the quality of underlying procedures.

It is notable that the ESS relies on Member States to collect data and compile statistics for national and EU purposes, and that an annual work programme identifies new priorities and actions.

4.2.2.1.4.3 Lessons learned and useful inputs

The ESS Regulation is a useful model for legislation at the EU level that creates a legal basis for data access and data use for a specific public good. While it of course focuses principally on statistics, some of the elements described above – notably in relation to professional independence, the quality framework in the Code of Practice, and the transmission of data between authorities – are likely to be useful inputs in a broader legal framework for B2G data sharing that would cover other types of public interest usage. Moreover, for B2G data sharing – as with statistical data collection – it is likely that EU level cooperation and aggregation may be necessary in some cases to achieve the desired public good. On that point too, the European Statistics Regulation is a useful input, given the interactions between the NSIs and Eurostat.

4.2.2.1.5 The Free Flow of Data Regulation (EU) 2018/1807

4.2.2.1.5.1 Key characteristics

The Free Flow of Data (FFD) Regulation creates an EU level framework for the free flow of non-personal data in the European Union. It aims to ensure that every organisation is able to store and process such non-personal data anywhere in the European Union, and ensures access for public authorities to such data, also when it is located in another Member State or when it is stored or processed in the cloud. Finally, it also contains rudimentary principles for switching of cloud service providers for professional users.

4.2.2.1.5.2 Relevance to data sharing – access and use

The relevance of the FFD Regulation to B2G data sharing may not be readily apparent. However, two elements are potentially relevant:

- Firstly, the inclusion of a general principle that **requests by national competent authorities to request, or obtain, access to data** for the performance of their official duties in accordance with Union or national law **may not be refused on the basis that the data are processed in another Member State**, in combination with a cooperation duty and cooperation procedure between competent authorities in different Member States when such cross border access requests are impeded.
- Secondly, the recognition that data sets can be composed of both personal and non-personal data, in combination with separately published **guidance in relation to such mixed data sets**[501]. The guidance essentially notes that, when data sets can be split into non-personal and personal data, the former should observe the rules of the FFD and the latter the rules of the GDPR; and that data sets where such separation is not possible should adhere unreservedly to the rules of the GDPR.

[499] See https://ec.europa.eu/eurostat/web/ess/about-us/ess-gov-bodies
[500] See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0516
[501] See https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets

Both of these inputs (cross border competence and cooperation, and precedence of the fundamental right to data protection) can be usefully integrated into a future legal framework for B2G data sharing.

4.2.2.1.5.3 Lessons learned and useful inputs

The FFD Regulation is principally relevant and useful in the context of B2G data sharing as an instrument that focuses on cross border cooperation in ensuring the effectiveness of data access requests, while also affirming the primacy of data protection rules.

4.2.2.1.6 Database Directive 96/9/EC[502]

4.2.2.1.6.1 Key characteristics

The Directive on the legal protection of databases aimed to harmonise protection of databases, stimulate investment in them and safeguard the balance between the rights and interests of database producers and users. It created a harmonized legal framework for the treatment of databases under copyright law in the EU and establishes a *sui generis* right for the makers of databases, irrespective of whether they qualify for copyright.

The Directive also harmonises exceptions to the copyright regime, as well as to the *sui generis* right. These include use or extraction of the database for the purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved; or use or extraction and/or re-utilization of the database for the purposes of public security of for the purposes of an administrative or judicial procedure.

4.2.2.1.6.2 Relevance to data sharing – access and use

The relevance of the Database Directive might seem significant, given that access and use requests in the context of B2G data sharing could include databases that could qualify for copyright protection or (more likely) *sui generis* protection.

However, the impact is not necessarily substantial. The Directive clarified that copyright protection applies only to databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation. In the absence of such personal originality, no copyright applies to the selection or arrangement criteria of the database. Given that much of the interest in B2G data sharing focuses on factual, systematic, objective and maximally comprehensive databases, copyright usually will not apply.

The *sui generis* rights regime allows the maker of a database to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of a database. That prevention right is also qualified by some basic legally defined rights for lawful users of databases, including the right to extract and/or re-utilize insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever. Due to this limitation to *insubstantial* parts of the contents, the Directive would not address access or usage needs that inherently require substantial (or even comprehensive) extraction and re-utilisation of database contents, which may be needed in a B2G data sharing context.

Similarly, as the 2018 evaluation of the Directive highlighted[503], a series of rulings from the European Court of Justice in 2004[504] have clarified the scope of the *sui generis* right, noting that the right does not apply to databases that are the by-products of the main activity of an organisation. As a result, it could be argued that *sui generis* protections would not apply broadly to machine-generated data and IoT devices, since (and to the extent that) such data is principally a by-product of a device's principal functionality.

[502] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009

[503] See https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection

[504] Fixtures Marketing Ltd v. Oy Veikkaus Ab (C-46/02, 9/11/2004), Fixtures Marketing Ltd v. Svenska Spel Ab (C-338/02, 9/11/2004), British Horseracing Board Ltd v. William Hill (C-203/02, 9/11/2004), and Fixtures Marketing Ltd v. OPAP (C-444/02, 9/11/2004)

However, as noted above, there is no universal consensus on whether the ECJ's rulings affect machine-generated data and IoT devices in such a clear and systematic way. Furthermore, even if the inapplicability of *sui generis* protections for machine generated data was universally accepted, the exclusion would still only apply to the extent that co-generated data is generated as an ancillary by-product of a main product or service that wasn't the subject of a substantial separate investment. This however leaves a gap in situations where data is a key component or even the main outcome of using a specific product or service for which the database-maker invested substantially for obtaining, verifying or presenting the data. In those cases, *sui generis* protection could arguably still apply. Thus, some ambiguity remains.

4.2.2.1.6.3 Lessons learned and useful inputs

While the Directive is unlikely to have a structural and systemic impact on B2G data sharing, it is none the less possible that some of the targeted data is subject to copyright and/or *sui generis* protection, and that not all anticipated B2G use cases would qualify as a case of lawful use as regulated by the Directive.

4.2.2.1.7 The Trade Secrets Directive (EU) 2016/943

4.2.2.1.7.1 Key characteristics

The Trade Secrets Directive[505] was adopted in 2016 and aimed to harmonize national laws in the Member States in relation to the unlawful acquisition, disclosure and use of trade secrets. It establishes a definition of trade secrets, as constituting information which meets all of the following requirements:

(a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;

(b) it has commercial value because it is secret;

(c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret;

It also defines a legal framework for the lawful and unlawful acquisition, use and disclosure of trade secrets, and clarifies that reverse engineering and parallel innovation must remain possible.

4.2.2.1.7.2 Relevance to data sharing – access and use

The Trade Secrets Directive contains provisions that ensure that the protections against unlawful use of trade secrets cannot be invoked to stop access, use or disclosure of trade secrets when this is necessary for the purpose of protecting a legitimate interest recognised by Union or national law. Specifically, article 3.2 of the Directive notes that "*The acquisition, use or disclosure of a trade secret shall be considered lawful to the extent that such acquisition, use or disclosure is required or allowed by Union or national law*" – thus ensuring that B2G data sharing of a trade secret would not be considered unlawful under the Directive provided that an appropriate EU or national law is available.

Moreover, article 5 requires Member States to ensure that the application of the measures, procedures and remedies of this Directive cannot be relied upon when "*the alleged acquisition, use or disclosure of the trade secret was carried out in any of the following cases: (a) for exercising the right to freedom of expression and information as set out in the Charter, including respect for the freedom and pluralism of the media; (b) for revealing misconduct, wrongdoing or illegal activity, provided that the respondent acted for the purpose of protecting the general public interest; (c) disclosure by workers to their representatives as part of the legitimate exercise by those representatives of their functions in accordance with Union or national law, provided that such disclosure was necessary for that exercise; (d) for the purpose of protecting a legitimate interest recognised by Union or national law*".

[505] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L0943

Recital (20) adds that that "*the protection of trade secrets should not extend to cases in which disclosure of a trade secret serves the public interest, insofar as directly relevant misconduct, wrongdoing or illegal activity is revealed*" (principally in relation to whistleblowing activity); and recital (21) adds that the "*measures, procedures and remedies intended to protect trade secrets […] not jeopardise or undermine fundamental rights and freedoms or the public interest, such as public safety, consumer protection, public health and environmental protection, and should be without prejudice to the mobility of workers. In this respect, the measures, procedures and remedies provided for in this Directive are aimed at ensuring that competent judicial authorities take into account factors such as the value of a trade secret, the seriousness of the conduct resulting in the unlawful acquisition, use or disclosure of the trade secret and the impact of such conduct. It should also be ensured that the competent judicial authorities have the discretion to weigh up the interests of the parties to the legal proceedings, as well as the interests of third parties including, where appropriate, consumers*".

In other words, in certain cases, the Directive recognizes the primacy of public interest (and allows Member States to introduce legislation on this point) in certain information over the private interest in its possible qualification as a trade secret. For that reason, the Directive does not result in barriers to the possibility of creating a legal framework (regulatory or non-regulatory) that compels certain forms of B2G data sharing, even when the targeted data qualifies as a trade secret in the sense of the Directive. Indeed, as the references in articles 3.2 and 5 show, the Directive explicitly acknowledges the possibility of specific provisions under national or Union law that permit trade secret sharing, without invalidating trade secret protection in general.

4.2.2.1.7.3 Lessons learned and useful inputs

The analysis above shows that the Trade Secrets Directive explicitly contains room for legislative intervention, both at the national and EU level, to allow the acquisition, use or disclosure of a trade secret for the purpose of protecting a legitimate interest recognised by Union or national law. In this way, the exceptions to the application of the Directive are arguably a **useful starting point for the potential scoping of permitted B2G data access and usage requests**, since this entails situations that require adequate protection of the "fundamental rights and freedoms or the public interest, such as public safety, consumer protection, public health and environmental protection".

In addition, the Directive also to some extent enables and facilitates data sharing, due to its focus on creating a safe environment for securely sharing of confidential, in a comparable way as could be seen e.g. in the European Statistics Regulation. This emphasis on the creation of a trustworthy framework for data exchanges is at any rate a prerequisite for effective B2G data sharing, at least in situations where businesses must have appropriate assurances that their data will be protected against access and use by other actors than the public authorities that originally requested it.

4.2.2.1.8 Open Data Directive (EU) 2019/1024 506(PSI Directive)

4.2.2.1.8.1 Key characteristics

The Open Data Directive is the most current recasting and amendment of prior directives regarding the re-use of public sector information (PSI)[507], which collectively aim to harmonise the rules across the Member States under which commercial and non-commercial use of certain information held by public sector bodies or by some public undertakings should be permitted, along with rules on transparency, charging, non-discrimination, exclusivity, and redress mechanisms.

The Open Data Directive explicitly aims to "promote the use of open data and stimulate innovation in products and services". It also explicitly introduces the concept of high-value datasets, defined as data that is associated with important benefits for the society and economy when reused. High-value data sets are

[506] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024
[507] See https://digital-strategy.ec.europa.eu/en/policies/open-data-0

subject to a separate set of rules ensuring their availability free of charge, in machine readable formats, provided via Application Programming Interfaces (APIs) and, where relevant, as bulk download. The Commission can adopt a list of specific high-value datasets by way of an implementing act, following an impact assessment.

4.2.2.1.8.2 Relevance to data sharing – access and use

Somewhat oversimplifying the issue, the Open Data Directive could be examined as the flipside of the B2G data sharing topic, notably G2B and G2C data sharing. Interests, objectives and therefore legal procedures and limitations are as a result quite different. Notably, the Directive takes the baseline position (subject to exceptions) that PSI data should be freely available for commercial and non-commercial use by any stakeholder, subject to marginal cost payments at most. Such relatively unconstrained re-use is not envisaged by a B2G data sharing framework. None the less, some considerations of the Open Data Directive could be usefully applied in a B2G context as well.

Principally:

- The focus on **high value datasets** could possibly be relevant as a starting point for describing data sets for which a significant public interest exists that warrants the application of a B2G data sharing framework. While static definitions always include the risk of introducing an incorrect focus, the approach can ensure that the scope of access and usage rights of public sector bodies are predictable.
- The Open Data Directive rightly recognizes the **importance of dynamic data access and APIs,** focusing on flexible access and usage rights, rather than focusing on static one-off data transfers. This approach is undoubtedly of interest too, with the same caveats on technical and financial feasibility applying – i.e. in the same way that smaller public sector bodies might struggle to create and maintain APIs and dynamic data services, some private sector businesses would likely face the same issue, and comparable solutions could be explored.
- Finally, it is worth noting that the Directive includes specific provisions on **research data**. Member States are required to adopt national policies and relevant actions aiming at **making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles**. The Directive allows Member States to take into account "concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests" (Article 10.1 of the Directive), in accordance with the principle of 'as open as possible, as closed as necessary'. Research data generally is required to be re-usable for commercial or non-commercial purposes insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. This should act as an enabler for some forms of B2G data access and re-use, notably when the data qualifies as publicly funded research data.

4.2.2.1.8.3 Lessons learned and useful inputs

The similarities between a theoretical future B2G framework and the G2B framework created by the Open Data Directive should not be overstated, since the scoping and objectives are quite different: the former should contain appropriate constraints to ensure that *some* data can be claimed for specific public interest purposes, whereas the latter requires public sector bodies (whose activities are largely publicly funded) to allow unconstrained re-uses wherever feasible. The Open Data Directive model can therefore not be readily or easily transplanted to a different policy context such as B2G data sharing.

None the less, the framework for publicly funded data could usefully be expanded upon, as could the facilitating of data access and re-use for research purposes by public sector bodies. The latter could conceivably include policy research (policy evaluation, preparation and forecasting) and compliance assessment, which would already comprise many critical use cases for B2G data sharing. Furthermore, the legal approach favouring dynamic data access via APIs as a priority is useful, as are the fallback scenarios for cases where such dynamic data access is not technically or economically viable.

Finally, prior PSI legislation as well as the current Open Data Directive recognized the need to establish a remuneration scheme in at least some instances, that takes into account the efforts required of the data holder to make data available. In the G2B context, the framework has focused on permitting only marginal costs as a default rule, thereby pushing charging options for public sector bodies gradually towards zero in many cases. Exceptions however exist that take into account the investments that may have been (or remain) required for some G2B exchanges. The general model and cost elements can certainly be built upon in a B2G data sharing framework. On these points, the approach of the Open Data Directive is a useful precursor.

In addition, the interaction between G2B and B2G data sharing should not be overlooked. The outcome of B2G data sharing should of course not be that the shared data is thereafter treated as regular PSI, which is therefore available for access and re-use by other private actors (a "B2G2B" data sharing chain). This is an issue that would need to be addressed by any future B2G framework.

4.2.2.1.9 Emerging: Proposal for a Data Governance Act

4.2.2.1.9.1 Key characteristics

The Data Governance Act was proposed in November 2020 and aims to provide a legal underpinning for responsible data sharing, especially in sectors where data sharing could be a driving force for the European data economy and for society in general. The proposed Act aims to make more data available and facilitate data sharing across sectors, including by directly creating new governance mechanisms that can support responsible sharing. It aims to do so through four pillars of the proposal:

- Firstly, it facilitates the reuse of certain public sector data that cannot be made available as open data under current law, due to e.g. intellectual property rights or privacy concerns. The reuse of such data would now be encouraged under the proposal, although specific mechanisms are defined in this case (including supervised data processing in a privacy-preserving manner), to ensure that the confidentiality of the data remains safeguarded. Member States must transparently communicate the conditions that apply to such reuse.
- Secondly, it creates a new role in the data economy: so called data sharing service providers that will function as data intermediaries, responsible for trustworthy data sharing or pooling.
- Thirdly, it creates a legal framework for data altruism, allowing citizens and businesses to make their data more freely available for the benefit of society, under the auspices of so-called data altruism organisations. Such organisations may choose to undergo a prior registration in the EU if they meet the applicable requirements (including that they must be independent, and may not be for profit), thereafter being permitted to refer to themselves as "data altruism organisation recognised in the Union". Furthermore, they must periodically report on their activities.
- And fourthly, it creates a framework for cooperation, supervision, and enforcement of these rules, including both national competent authorities and a new European Data Innovation Board. The latter is charged with advising and assisting the Commission in developing a consistent practice of public sector bodies, competent bodies and competent authorities, advising on the prioritisation of cross-sector standards, advising on interoperability of data, and facilitating the cooperation between national competent authorities through capacity-building and the exchange of information.

4.2.2.1.9.2 Relevance to data sharing – access and use

While B2G data sharing is not the explicit focus of the proposed Data Governance Act, it contains several components that could be usefully built upon by a future legal framework that aims to facilitate B2G data sharing. These include notably the following:

- The new re-use rules contain explicit provisions in relation to the use of so-called "**pre-processed data**", where such pre-processing by public sector bodies aims to anonymize or pseudonymise personal data or delete commercially confidential information before allowing it to be re-used by third parties, and to the use of **secure processing environments** (a legally defined concept) when this is required to safeguard the interests in the data. These elements can of course be useful to protect private sector information as well in B2G cases.
- The new framework for **providers of data sharing services** (including the legal and procedural safeguards to ensure their independence and trustworthiness, and the quality of their services) could be used as an input for the creation of any intermediaries that would intervene as a trusted third party in making data accessible (including by pre-processing it where necessary or by providing dynamic data services) to public sector bodies.

4.2.2.1.9.3 Lessons learned and useful inputs

Since the Data Governance Act is still in a proposal stage, it would be imprudent to rely too strongly on its potential impacts. None the less, the issues mentioned above could certainly prove to be beneficial in order to build an efficient and trustworthy B2G data sharing ecosystem.

4.2.2.1.10 Conclusions on viability, potential conflicts, and approaches for a B2G data sharing framework

The legal analysis shows that there a **substantial number of useful legal precursors** at the EU level that can provide some of the inputs for creating a legal framework to enable and sustain B2G data sharing. However, it is also clear that **none of these already provides a suitable and comprehensive answer for B2G data sharing**, including in terms of defined access rights, modalities (including the role of intermediaries, dynamic data sharing, data stewards and security), justifications of data access and use and purpose limitation linked to the public interest, or safeguards for the legitimate interests of targeted companies.

Moreover, the analysis would also suggest that these barriers are **difficult or impossible to overcome at the EU level through purely non-regulatory action**. Notably, data protection rights (and conceivably some intellectual property rights) would act as a barrier to a predictable B2G data sharing environment. For personal data, a key challenge is the lack of a dependable legal basis for data claims in the absence of legislative protections, especially in situations where data sharing cannot be entirely justified by appealing to the protection of vital interests.

Such a framework would be capable of addressing the legal issues identified above and would not seem to trigger specific consistency concerns.

4.2.3 Measures supporting citizen empowerment ('human-centric data economy'): Case Studies

The following section includes a presentation of the case studies carried out for Measures supporting citizen empowerment ('human-centric data economy'), namely Open Banking LTD, Withings, Green Button initiative and Reciprocity clause in Australia Consumer Data Rights.

The case studies bring an overview of best practices within area where data portability is used, to understand better the context and learn from other sectors' experience. One of these cases is the Open Banking LTD, which brings forward the experience of the banking sector with open standards and data portability. In the same context, the experience of the Australia Consumers Data Rights is another valuable source for understanding the data portability effects for both businesses and customers. At the same time, Withings case study presents the current developments concerning data portability and access in the field of fitness trackers and wearables, highlighting the potential of the sector. The Green Button initiative is a best practice of non-regulatory intervention within an area where technological interoperability might not be sufficiently mature for a regulatory intervention. It shows how an environment can be further enhanced to improve data portability perception and acceptance.

4.2.3.1 Open Banking Implementation Entity (OBIE)

4.2.3.1.1 Introduction

Open Banking was set up by the Competition and Markets Authority (CMA) on behalf of the UK Government in 2016. It was designed to bring more competition and innovation to financial services.

In the same year, CMA establishes the Open Banking Implementation Entity (OBIE), a company set up by to deliver Open Banking.

Since January 2018, when the EU’s Second Payment Services Directive (PSD2) came into effect, Open Banking enables customers and small and medium-sized businesses to share their current account information securely with other third-party providers.

4.2.3.1.2 Governance

The company is governed by the Competition and Markets Authority.

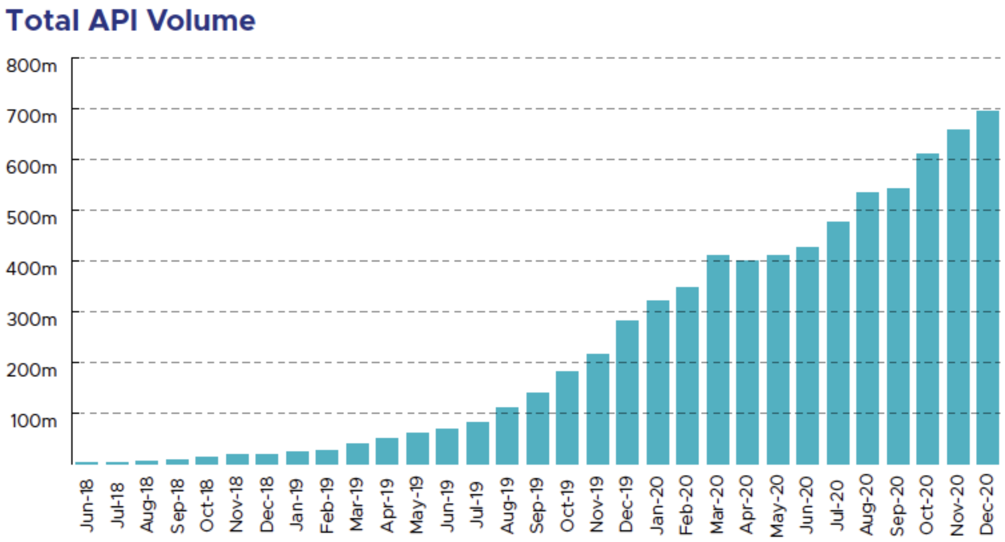
4.2.3.1.3 Operations

OBIE works with the UK’s largest banks and building societies as well as challenger banks, financial technology companies, third party providers and consumer groups. Its main role is to:

- Design the specifications for the Application Programme Interfaces (APIs) that banks and building societies use to securely provide Open Banking
- Support regulated third party providers and banks and building societies to use the Open Banking standards
- Create security and messaging standards
- Manage the Open Banking Directory which allows regulated participants like banks, building societies and third-party providers to enrol in Open Banking
- Produce guidelines for participants in the Open Banking ecosystem
- Set out the process for managing disputes and complaints

Currently, there are over three million active users of open banking-enabled products. The ecosystem is thriving: 301 firms are active in the market, with another 450 in the pipeline.[508] Since 2018, the total APIs volume (measured through calls per month) shows significant increase. In 2021, OBIE hosts 104 apps that supports consumers, businesses and offers technical support (21 apps).

[508] Open Banking Annual Report 2020.



*Source: OBIE data

The OBIE supported multiple initiatives to encourage user adoption and grow the open banking market during 2020. It sponsored the Nesta Open Up 2020 programme, a £1.5m prize challenge to promote open banking solutions. In 2020, it began working on the Consumer Evaluation Framework (CEF), a framework through which the success of the CMA’s remedies can be assessed.

The OBIE provides a broad range of critical services and support to the ecosystem participants, including promotion of open banking, customer and stakeholder engagement to enhance customer adoption, provision of technical support and operational expertise to ecosystem participants, encouraging and facilitating new entrants into the ecosystem.

4.2.3.1.4 Financing

Open Banking LTD is a non-profit organisation, funded by the UK’s nine largest banks and building societies: Allied Irish Bank, Bank of Ireland, Barclays, Danske, HSBC, Lloyds Banking Group, Nationwide, RBS Group and Santander.

4.2.3.2 Withings

4.2.3.2.1 Introduction

Withings is a French consumer electronics private company, established in 2008, in Issy-les-Moulineaux, France. In 2016, the company was purchased by Finnish company Nokia, becoming a division of Nokia known as Nokia Health. It kept its own brand until 2017, when was replaced by Nokia brand. After two years from the acquisition, the company regains its independence again, functioning currently under the Withings name.

4.2.3.2.2 Governance

Withings is a privately held company.

4.2.3.2.3 Operations

Since 2009, when they released the first connected body scale on the market, the Withings brand has grown to become synonymous with integrating innovative and meaningful measures into easy-to-use devices designed to empower people to make the right decisions for their health.

Today, the ecosystem of connected health devices and apps includes a range of smart scales designed to help fulfill fitness and weight goals, a family of stylish activity trackers and hybrid smartwatches, an advanced sleep-tracking mat, and medically accurate devices for easy and effective blood pressure and temperature monitoring. The devices sync automatically with the free Health Mate app, where people can track progress, get advice, and share data with their doctors.

The Withings developer portal allows developers to create applications that take advantage of the Withings devices and the data they record. Developers have the ability to access data stored in the Health Mate app for users who gave them prior consent, including weight, body fat, activity, sleep, blood pressure and heart rate, ECG, PWV and more, and to integrate them into their services. The API is public, and it only requires registration of the application. API apps are limited to 5000 active users and have a request limit of 120 requests per minute for a registered partner (calling a service for a first user, followed by one for a second user, will increase the request counter by 2 requests). If the developer needs more than 5000 users or more than 120 requests per minute (needs a Corporate Service Level Agreement and a larger number of requests per minute), he is redirected towards the Enterprise plan section.

The Withings developer platform is organised into three types of services based on specific use cases.

Product	Details
Data API	The Data API is a complete Open API that provides service **to retrieve user health and wellness data and other useful device information.** The Data API uses the OAuth authentication protocol to authenticate safely requests between users and the application developed. Users can give access to their Withings data without giving away their password.
Device & Logistics API	Device & Logistics API enables partners to: 1. Deliver pre-activated and pre-configured devices to program members or patients, removing the friction of device setup and account creation 2. Use the Dropshipment API to deliver devices directly from Withings to program members or patients. This API is reserved to partners using Withings Cellular Data Hub, Withings SDK, or want to access the Dropshipment API services.
Custom Solutions	Withings is also able to provide access to advanced technologies, resources and support to help developers build specific workflows using health data: 1. Withings SDK is a native iOS/Android development Kit allowing partners to setup Withings devices into their own mobile app. 2. Infrastructure hosted on Withings HIPAA & HDS Cloud. 3. Isolated Program Member account dissociated from Withings consumer account.

4. Withings support and recommendations for developer's custom health workflow.

5. Corporate SLA on Withings APIs.

Additionally, Withings has an in-house research body, dedicated to accelerating the connected health revolution through a combination of in-house research and academic partnerships. Using real-time data, it tracks the extent to which key risk factors for heart disease are linked to lifestyle, such as sedentary behavior, overweight and obesity, and high blood pressure, and what steps can be taken to reduce risks.

Withings products have been involved in numerous clinical trials, such as in the Advanced stage Breast cancer and Lifestyle Exercise (ABLE) Trial. The trial aimed to assess the feasibility of a physical activity intervention in women with metastatic breast cancer and to explore the effects of physical activity on functional, psychological, and clinical parameters.[509] Another study that used Withings data aimed to determine whether wearable activity trackers could provide information regarding users' adherence to home confinement policies because of their capacity for seamless and continuous monitoring of individuals' natural activity patterns regardless of their location.[510]

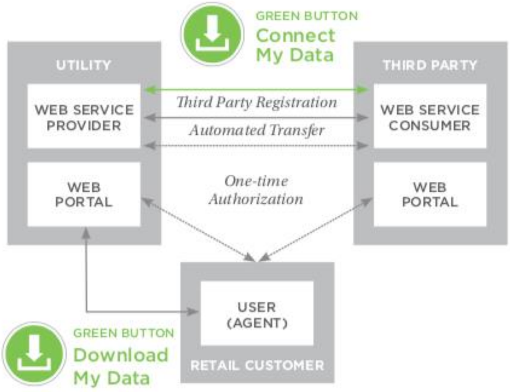
4.2.3.2.4 Financing

Withings is a privately held company.

4.2.3.3 Green Button initiative

4.2.3.3.1 Introduction

The US Green Button initiative aims to facilitate customers' access and reuse to energy consumption data. It is a technical standard and a label that energy providers can display on their website. By clicking on it, the consumer has easy access to the data in standardized format and two options: data are available for **download** in standard XML format or for ongoing **connection** with third parties through standardized API.



The key data shared are:

- Readings

[509] Delrieu L, Pialoux V, Pérol O, Morelle M, Martin A, Friedenreich C, Febvey-Combes O, Pérol D, Belladame E, Clémençon M, Roitmann E, Dufresne A, Bachelot T, Heudel PE, Touillaud M, Trédan O, Fervers B, "Feasibility and Health Benefits of an Individualized Physical Activity Intervention in Women With Metastatic Breast Cancer: Intervention Study," JMIR Mhealth Uhealth 2020;8(1):e12306; https://mhealth.jmir.org/2020/1/e12306

[510] Pépin JL, Bruno RM, Yang RY, Vercamer V, Jouhaud P, Escourrou P, Boutouyrie P, "Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis," J Med Internet Res 2020;22(6):e19787; http://www.jmir.org/2020/6/e19787/

- Interval data
- Summary Information
- Power Quality Metrics

4.2.3.3.2 Governance

This industry led initiative has been spearheaded in 2012 by the White House under President Obama. The coordination was held by the National Institute for Standard and Technology which led the development of the standard. Later, a Green Button Alliance was created in 2015, bringing together utilities and data re-users (service providers). The Alliance is the body which supports adoption, certification, marketing and training.

4.2.3.3.3 Operations

The initiative has been widely adopted: by 2015, 150 utilities had joined, representing 100 million users. For example, one Californian provider (Pacific Gas and Electric) reports 120.000 customers using the option, and 100 third parties registered to receive data.

Some data have shown that it has led to reduce energy consumption between 6 and 18%. Moreover, it has extended from energy to gas and water and become mandatory in some states. It is being extended to other countries such as Korea and Canada.

However, some challenges have also appeared, notably on usability. While formally compliant, several companies delay the provision of data or increase friction. The Green Button initiative managers report too limited oversight and registry of implementation, and too few mandatory parts. For this reason, there are calls for a stricter monitoring regime, including start to finish end users experience to avoid deliberate friction by data holders.

In a recent presentation, the Green Button Alliance shared interesting lessons learnt:

What worked	**What didn't**
• Use of "off-the-shelf" standards and technologies • No need to invent transport, security, or authorization methods • Seamless integration with existing utility systems • Open forum to discuss changes to the standard • OpenADE.org — our anyone-welcome, technical task force (open automated data exchange — a pre- "Green Button" moniker, still used today) • OpenADE ideas brought to NAESB and to IETF • Few barriers to implementation • GitHub for examples • No memberships required	• Waiting to form a trade group • From the initial ideas to the forming of GBA: five years had past • Lack of coordination between efforts (standard, support, go-to, testing, websites) • Use of logos and terms ("Green Button") without oversight/compliance • No registry of implementations • Too-few mandatory parts • Establishing minimum-implementation requirements that don't meet the needs of the industry or the consumers • No requirements to meet the latest standards

- No licensing fees
- ESPI standard available for low cost to anyone
- Separation of usage data from personal data
- Parallel data streams
- Security and GDPR adherence
- Community acceptance
- Non-voting governmental participation
- No lobbying by GBA
- Single place for all resources:
- GBA provides a community (Slack, GitHub, Zoom)
- GBA provides technical education
- GBA provides compliance testing
- GBA provides support of standardization enhancements

4.2.3.3.4 Financing

The initiative is entirely self-funded by industry.

4.2.3.4 Reciprocity clause in Australia Consumer Data Rights

Australia introduced its Consumer Data Right (CDR) on 4 February 2020. The provisions are to be deployed on a sector by sector basis, starting from banking, followed by energy and telecom, and ultimately to the whole economy. The Open Banking initiative was accordingly launched in June 2020.

The purpose of reciprocity is to allow a fair competitive market where all players have similar obligations, and in particular to avoid the risk of "platform envelopment." It acts as a non-regulatory incentive for financial institutions to fully participate. At the same time, there are risks that excessive reciprocity clauses will act as a deterrent for companies to join the scheme.

This disposition is implemented at sector level. This sectorial approach to reciprocity is defined not by the type of entity, but by the type of service delivered. It applies only to the delivery of the financial products listed in the table below. Concretely, a personal loan provider (phase 2d in the table) requesting to be registered as accredited person is subject to reciprocity for the data concerning loans, but a budgeting app provider is not.[511]

Phase 1 products		**Phase 2 products**	**Phase 3 products**
(a) a savings account (b) a call account (c) a term deposit (d) a current account (e) a cheque account	(a) a residential home loan (b) a home loan for an investment property (c) a mortgage offset account	(a) business finance (b) a loan for an investment (c) a line of credit (personal)	

[511] Author's reflection to be checked with potential interviewees: it appears therefore that a "Big Tech" company could still join the scheme for providing non-financial services, gain access to consumers data while not having to provide data in return until reciprocity is deployed economy-wide.

(f) a debit card account (g) a transaction account (h) a personal basic account (i) a GST or tax account (j) a personal credit or charge card account (k) a business credit or charge card account	(d) a personal loan	(d) a line of credit (business) (e) an overdraft (personal) (f) an overdraft (business) (g) asset finance (including leases) (h) a cash management account (i) a farm management account (j) a pensioner deeming account (k) a retirement savings account (l) a trust account (m) a foreign currency account (n) a consumer lease.

Any entity (also companies outside of finance) wishing to benefit from the CDR data portability regime should is subject to a process of accreditation. The Australian Competition and Consumer Commission (ACCC) is charged with the accreditation process.

The reciprocity applies to "equivalent data", that is, data that fall under the scope of the Open Banking initiative and that are generated with respect to a product in the scope (see table above). For example, if Amazon requested a CDR accreditation for its lending service, it would only have to share data generated through this service, not other consumer data generated through its marketplace. The ACCC has to determine what exactly constitutes equivalent data. The definition of the "equivalent data" scope is an important factor in determining the likelihood of ADR companies to join the scheme.

The reciprocity applies to "equivalent data", that is, data that fall under the scope of the Open Banking initiative and that are generated with respect to a product in the scope (see table above). For example, if Amazon requested a CDR accreditation for its lending service, it would only have to share data generated through this service, not other consumer data generated through its marketplace. The ACCC has to determine what exactly constitutes equivalent data. The definition of the "equivalent data" scope is an important factor in determining the likelihood of ADR companies to join the scheme.

The governance of the scheme is peculiar. The initiative is led by the treasury department of the Australian government, which designs the high-level provisions and strategy. At the technical level, the initiative is supported by a dedicated Data Standard Body (DSB). This is composed of external specialist contractors, not civil servants, and sits under the national research body CSIRO as part of Data61, its specialist data science structure.[512] The consumer data standardisation effort is therefore managed independently from other standardisation efforts and bodies. The culture of DSB is deeply permeated in technology, digital and

[512] Commonwealth Scientific and Industrial Research Organisation.

open source, as shown by their activity on GitHub and their systematic usage of blogs as a communication tool.[513]

Uniquely, beside standardisation of data models, APIs and interoperability, the DSB is also in charge of developing Consumer Experience (CX) standards, to encourage the adoption of CDR by consumer across the economy.[514] The CX standards focus on the consent model – to make sure consumer fully understand and can exercise their consent for data portability across entities.

In terms of adoption, the initiative is still at an early stage. The large banks had to participate by regulation, and 90 smaller banks have also joined. In terms of accredited re-users, three are fully live, while other 10 have started the accreditation process. Participants have reported the process to be quite cumbersome and it is therefore under review to lower the barriers to entry.

The just-published "Inquiry into future directions for the Consumer Data Rights" carried out for the Australian government recommends the expansion of reciprocity from a sector-by-sector basis to a cross sector. It also recommends exemptions for small business.[515]

[513] https://consumerdatastandards.gov.au/ and https://consumerdatastandardsaustralia.github.io/

[514] https://consumerdatastandards.gov.au/

[515] Scott Farrell et al., Inquiry into Future Directions for the Consumer Data Rights, Australian Government 2021.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

KK-01-21-277-EN-N

Publications Office
of the European Union

doi: 10.2759/759296

ISBN 978-92-76-41909-9