EUROPEAN UNION AGENCY
FOR CYBERSECURITY

# ENGINEERING PERSONAL DATA PROTECTION IN EU DATA SPACES

JANUARY 2024

# ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure, and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

## CONTACT

For contacting the authors please use isdp@enisa.europa.eu
For media enquiries about this paper, please use press@enisa.europa.eu

## CONTRIBUTORS

Isabel Barbera, Claude Castelluccia, Giuseppe D'acquisto, Marta Fydrych Gasowska, Marit Hansen, Jaap-Henk Hoepman, Meiko Jensen, Konstantinos Limniotis, Maria Raphael, Marie-Charlotte Roques Bonnet, Fernando Silva, Fatbardh Veseli, Barbara Vieira, Kim Wuyts, Christian Zimmermann, Luis de Salvador Carrasco, Peter Kraus, Teresa Martinez Sanchez and Prokopios Drogkaris.

## EDITORS

Prokopios Drogkaris (ENISA), Javier Gomez Prieto (ENISA)

## ACKNOWLEDGEMENTS

We would like to thank the EC colleagues for their review and valuable comments.

## LEGAL NOTICE

This publication represents the views and interpretations of ENISA, unless stated otherwise. It does not endorse a regulatory obligation of ENISA or of ENISA bodies pursuant to the Regulation (EU) No 2019/881.

ENISA has the right to alter, update or remove the publication or any of its contents. It is intended for information purposes only and it must be accessible free of charge. All references to it or its use as a whole or partially must contain ENISA as its source.

Third-party sources are quoted as appropriate. ENISA is not responsible or liable for the content of the external sources including external websites referenced in this publication.

Neither ENISA nor any person acting on its behalf is responsible for the use that might be made of the information contained in this publication.

ENISA maintains its intellectual property rights in relation to this publication.

## COPYRIGHT NOTICE

ISBN 978-92-9204-650-7, DOI 10.2824/210862

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

The recent EU legislative initiatives promoting data sharing are sectoral and cross-sectoral instruments that aim to make data available by regulating the reuse of publicly and privately held data, including personal data. They also facilitate data sharing by creating of novel intermediaries and sharing environments where the parties involved can pool data and facilities in a trusted and secure way.

Common European data spaces (EU data spaces) are a novel concept introduced in the European strategy for data and elaborated further within the Data Governance Act (DGA). It is envisioned that they will facilitate innovation, economic growth and digital transformation and revolve around creating a framework for data sharing that respects privacy, security and other applicable regulatory considerations while promoting cross-sector collaboration and interoperability.

This report attempts to contextualise the main design principles regarding protection of personal data and demonstrate how to engineer personal data protection through two use cases of an envisioned EU data space in the pharmaceutical domain.

Despite the potential of the EU data spaces, there are still considerations regarding appropriate technical and organisational measures and how to engineer them into practice, both from a data protection and from a cybersecurity point of view. Even if there are already a good number of privacy enhancing technologies that can support us in meeting specific data protection goals, we should not neglect the fact that we are called to address new processing operations, where the roles and responsibilities are not always clearly defined.

# 1. INTRODUCTION

## 1.1 DATA DRIVEN INNOVATION

Data Driven Innovation (DDI) is not an entirely new concept; it evolves around processing large volumes of data towards extracting meaningful insights and creating valuable innovation(s) [1]. Over the last decade we have been discussing on both the opportunities but also the challenges of big data [2] [3] & [4]. In order though to leverage the full analytical power of data and drive impactful advancements, however, we need to move beyond multiple collection and analysis points towards a more collaborative approach, while continuing to respect the protection of personal and business-sensitive/critical data.

The European strategy for data [5] was announced in 2020 and is a 5-year plan that presents a vision for a 'single European data space', which is described as "a genuine single market for data – open to data from across the world – where personal as well as non-personal data, including sensitive business data, are secure and businesses have easy access to high-quality industrial data, boosting growth and creating value". In this context, the European strategy for data recognizes the strategic importance of investing in common EU data spaces as a mechanism to drive growth and innovation in key economic sectors and public interest domains.

**The European strategy for data recognises the strategic importance of investing in EU data spaces as a mechanism to drive growth and innovation.**

## 1.2 COMMON EUROPEAN DATA SPACES

Common European Data Spaces (hereinafter EU data spaces) are a novel concept introduced in the European strategy for data and elaborated further within the Data Governance Act (DGA) [6]. It is envisioned that they will facilitate innovation, economic growth and digital transformation and revolve around creating a framework for data sharing that respects privacy, security and other applicable regulatory considerations while promoting cross-sector collaboration by implementing the following set of measures by:

- fostering access to and the re-use of certain categories of data held by public sector bodies that cannot be made available as open data due to the protection that applies to the data.
- ensuring that data intermediaries will function as trustworthy facilitator of data sharing or pooling in the EU data spaces.
- facilitating data sharing, specifically to enable the use of data across sectors and for certain purposes.

Despite the benefits of EU data spaces, a challenging endeavour is linked to the number of stakeholders with diverging strategic goals and specific data needs that may be challenging to reconcile. Additionally, as the EU data spaces will operate within the framework of existing Union policy and law, it is crucial to identify cross-sector commonalities, common terminology, design and compliance frameworks and articulate proper data engineering tools [7].

## 1.3 DESIGN PRICNIPLES OF EU DATA SPACES

While the DGA provides an overarching horizontal governance framework for EU data spaces, it also highlights the need for them to operate according to other applicable Union policy and law, such as those relating to data protection, cybersecurity, intellectual property, etc., and complying with relevant sectoral-related legislation. One essential element will be the implementation of tools for pooling, accessing, using and sharing data while complying with applicable legislation, allowing data holders to manage access rights and conditions over time.

Such considerations are also highlighted in the recent open finance report [8] of the Expert Group on European financial data space.

Secondly, to be “accessed and used in the most effective and responsible manner possible”, EU data spaces must be designed, set and maintained so that they provide a secured and supervised processing environment. They must also remain technically interoperable with others while ensuring, as needed, commercial or statistical confidentiality, protection of intellectual property rights of third parties and protection of personal data. Therefore, consistent and predictable rules regarding access and re-use of data are key for data holders and data users to comply with Union policy and law.

Lastly, basic and uniform recommendations on conditions for reuse and related technical and organisational measures (TOMs) shall be adopted, notably to help data holders better understand how they should adapt security & confidentiality rules and fine-tune their corporate policies so that they keep complying with EU data protection requirements (GDPR).

## 1.4 INTEROPERABILITY AT THE CORE OF EU DATA SPACES

Data holders are subject to an obligation to promote interoperability. Indeed, providers of data sharing services shall “*facilitate the exchange of the data in the format in which it receives it from the data holder and shall* convert the data into specific formats only to enhance interoperability within and across sectors or if requested by the data user or where mandated by Union law or to ensure harmonisation with international or European data standards”. Such interoperability should be defined first and foremost at the EU level but also taking into consideration technical standards or well-acknowledged specifications.

Intermediaries may facilitate interoperability and the sharing of personal data, assisting data holders with anonymising or pseudonymising personal data, drafting and executing personal data sharing agreements or facilitating the exercise of individuals’ rights.

## 1.5 SCOPE AND OBJECTIVES

This report addresses the design and deployment of EU data spaces from an engineering perspective with an emphasis on the engineering of personal data protection. The main objectives of this report are to contextualise the main design principles regarding protection of personal data and demonstrate how to engineer personal data protection through two use cases of an envisioned EU data space in the pharmaceutical domain.

This work is meant to support policy makers, regulators and data protection practitioners and is performed in the context of ENISA’s tasks under the Cybersecurity Act (CSA) [9] to support Member States on specific cybersecurity aspects of Union policy and law relating to data protection and privacy. This work builds upon the Agency's activities in the area of Data Protection Engineering and is produced in collaboration with the ENISA Ad Hoc Working Group on Data Protection Engineering[1].

## 1.6 STRUCTURE OF THE DOCUMENT

This report is structured in four main sections as follows:

Section 1 provides the context of EU data spaces and highlights key considerations relating to data-driven innovation, design principles of data spaces and interoperability.

[1] European Union Agency for Cybersecurity, ‘Ad-Hoc Working Group on Data Protection Engineering’, ENISA website, https://www.enisa.europa.eu/topics/data-protection/ad-hoc-working-group-on-data-protection-engineering

Section 2 elaborates on specific data protection considerations in EU data spaces. More specifically, it addresses terminology and roles; the input and output problems of privacy as well as the role of data protection engineering in the implementation of data spaces. The section also focuses on other key considerations such as interoperability, accountability, efficiency, impact assessments, privacy-enhancing technologies (PETS) selection and data subject rights.

Section 3 provides data protection engineering insights into an envisioned pharmaceutical data space by illustrating two specific use cases of application. Each use case attempts to highlight how specific pseudonymisation techniques can be deployed by different actors towards providing data users with useful but adequately protected data-sets.

Chapter 4 concludes the repots and summarises the key findings and considerations with regards to engineering personal data protection in EU data spaces.

# 2. DATA PROTECTION CONSIDERATIONS IN EU DATA SPACES

## 2.1 TERMINOLOGY AND ROLES ANALYSIS

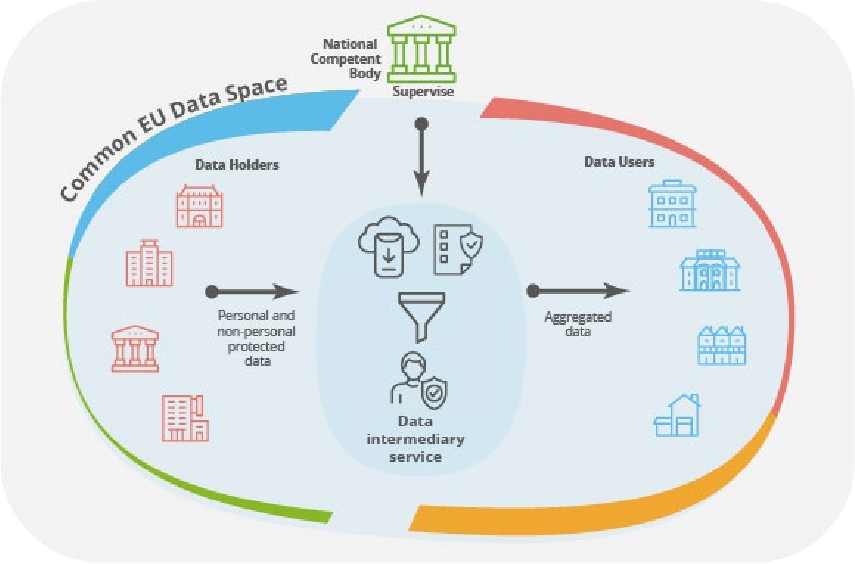
An EU data space, as defined under the DGA, consists of three main actors; the data holder(s), the data intermediary and the data user(s). The DGA provides a definition for each one of them which are briefly presented below:

- **Data Holder** is a legal person who is not a data subject with respect to the specific data in question, but has the right to grant access to or to share certain personal data or non-personal data.
- **Data Intermediary** is an entity that acts as an intermediary between data holders and data users. The Data Intermediary plays a role in facilitating the secure and controlled sharing of data by providing services such as data access.
- **Data User** is a natural or legal person who has lawful access to certain personal or non-personal data and has the right to use that data obtained from the intermediary for commercial or non-commercial purposes.

**The processing operation depends on how the sharing is performed and what the actual role of the data intermediary is.**

An illustrative representation of the interactions between these three actors is provided in Figure 1 below.

**Figure 1: Main EU Data Space Actors**



The DGA applies to "any digital representation of acts, facts or information", including personal data. When shared data includes personal data, there needs to be a mapping in place between the roles (and relevant responsibilities) between the DGA and the GDPR. However, this might not be straightforward, as the processing operation depends on how the sharing is performed and what the actual role of the data intermediary is.

In the context of the DGA and the GDPR, the Data holder is responsible for ensuring the lawful and proper collection, processing, and storage of data, the data intermediary provides services to facilitate controlled data sharing, processing, and storage and the data user receives and uses data for various purposes, such as analysis, research, or other legitimate interests. Even based on these rather generic descriptions, it cannot be safely established who is acting as the controller, whether there is more than one controller, acting as joint controllers, whether there is a processor and whether data users are data recipients. Even if the GDPR sets specific requirements that have to be fulfilled by each role, it is not clear under the generic model which entity alone or which entities collectively “determine the purposes and means of the processing of personal data”, which entity acts “on behalf of the data controller” and if an entity “to which the personal data are disclosed”. As pointed out in the recent publication [10] from the Agencia Española de Protección de Datos (AEPD) “The most important aspect defining a processing operation is its purpose”.

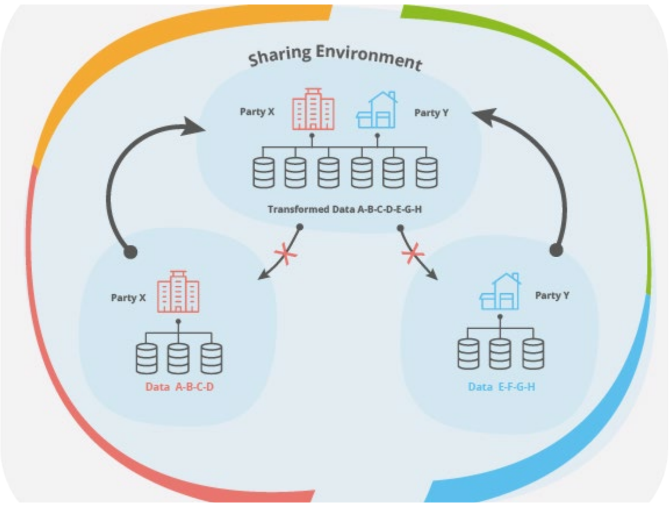
In essence, the DGA and the GDPR create a framework where Data holders, data Intermediaries, and data users work together to ensure the responsible and compliant sharing, processing, and use of data. They must align their practices with the principles and obligations outlined in both legislative acts to protect individuals' rights and privacy while fostering innovation and data-driven initiatives.

## 2.2 INPUT PRIVACY AND OUTPUT PRIVACY PROBLEMS

Before a data sharing process is initiated, we need to consider the possible risks for the data subjects that can emerge during the processing that will be performed by the sharing environment. Two main challenges can be identified, as presented below:

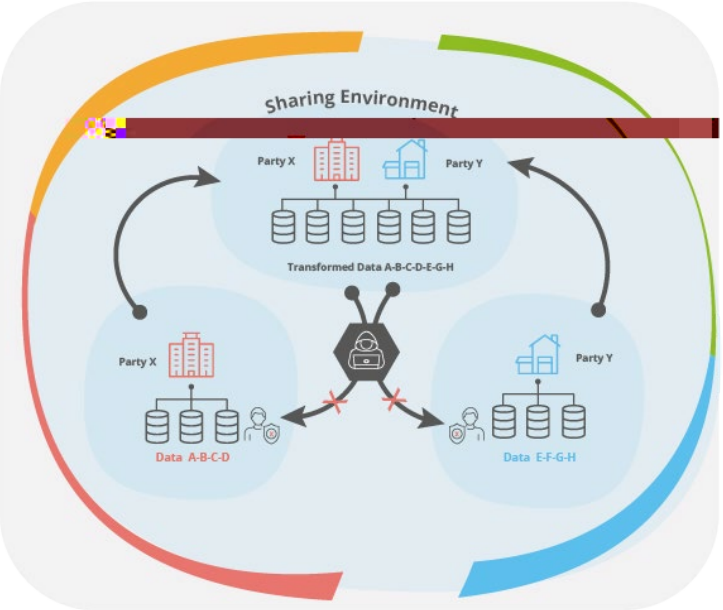
- **The input privacy problem**: The objective is to allow processing to be performed in data that has been shared but at the same time ensure that the sharing environment will not be able to revert back to the initial data, which can lead to the singling out or identification of individuals.

Figure 2: Input Privacy Problem



- **The output privacy problem:** The objective is to prevent the singling out or identification of individuals after the computations performed by the sharing environment have taken place.

Figure 3: Output Privacy Problem



**Data protection engineering offers data controllers a factual option for data sharing while minimizing the risk of information misuse, data breaches etc.**

Both input privacy and output privacy are crucial aspects of data privacy and security in sharing environments such as EU data spaces. Ensuring the protection of personal data from the moment they are collected to the moment the results are shared is an integral element for trust on such sharing frameworks [11]. The means to address these two risks is through deploying relevant data protection engineering building blocks [12] while adhering to the GDPR principles.

## 2.3 THE ROLE OF DATA PROTECTION ENGINEERING

Data protection engineering can be a very important enabler for the deployment of EU data spaces where data sharing opportunities and protection of personal data can coexist fruitfully and not hinder each other. Not addressing the legal and technical data protection requirements constraints inherent in the implementation of EU data spaces may be a blocking factor against the adoption of the data sharing paradigm and may limit the scope of the EU data strategy. This prerequisite is highlighted not only in the DGA itself but also in a recent report published by AEPD [13].

Data protection engineering is not a mere “compliance tool” for the GDPR. In implementing appropriate measures and the necessary safeguards to reinforce data protection principles and enable the exercise of individuals’ rights, data protection engineering offers data controllers a factual option for data sharing while minimising the risk of information misuse, data breaches or other security threats. The development of compelling and convincing use cases for safe and lawful data sharing is one of the most crucial challenges for the successful implementation of EU data spaces. Data protection engineering has the potential to strike a balance between data sharing and data protection. The risk of using new and poorly understood measures can act as a disincentive to adoption. This can be particularly true for emerging technologies, which may not yet have established best practices. Developing standards and building up on existing good practices can reduce the complexity and uncertainty associated with adopting these techniques. This can help increase trust and confidence in data protection engineering tools and promote their widespread adoption.

Another non-trivial aspect is the role of a data intermediary in an EU data space sharing scenario as it might entail a decision-making concerning risk mitigation. Once the potential risks of the planned data processing activity have been identified by the data controller, the question of how to mitigate which of the risks identified becomes eminent. The standard methods for risk mitigation, as documented for example in the Privacy Design Strategies [14], cover e.g., the application of privacy-enhancing technologies (PETs), or the decision to split the processing activities among multiple separate actors. Advanced approaches in this respect are discussed in [12], [15] & [16].

The data intermediary may or may not be part of the group of decision-makers (based on being either a data controller or a data processor), but definitely is one of the entities that need to actually implement the set of PETs chosen. If (advanced) data pseudonymisation (or even anonymisation) is identified as the best means to address a certain data protection impact assessment (DPIA) risk, the task of implementing the application of pseudonymisation in the dataset under consideration must be performed by someone. Of course, it would be possible for the data intermediary to hand over the full, non-pseudonymised dataset to another data processor, which then performs the pseudonymisation, but this design would actually introduce a new risk vector, thus not ideally mitigating the data disclosure risk in question. In the best case, the dataset would be pseudonymised at the data intermediary itself (or at the data storage locations the data intermediary brokers for). However, this approach then would require that the data controllers querying the datasets from the data intermediary would need to tell the data intermediary exactly how to perform the specific pseudonymisation scheme in consideration. Subsequently, the data intermediary would need to instantiate and perform the data pseudonymisation by itself, providing only the pseudonymised dataset to the querying data user.

**Supporting a holistic DPIA with multiple data controllers and data intermediaries requires a holistic approach**

The same need for risk mitigation on demand holds true for all other means of risk mitigation, as discussed above. If it is decided that federated learning is to be utilised as the (privacy-protecting) means to train a machine learning model, that decision must be implemented and coordinated by the data controller, in close collaboration with the data intermediary that provides access to the data storage locations. If a *k*-anonymity or differential privacy scheme is to be utilised to protect the data from disclosure, said technique must be implemented at the site that stores the data.

As can be seen, in order to reasonably implement privacy enhancing technologies in data space scenarios, it is essential for the data intermediary to be able to perform said tasks, i.e. to have implementations ready to deploy these techniques to the datasets in question, and be able to deploy these implementations dynamically on every data sharing scenario concerned – as instructed by the data controller.

## 2.4 DATA PROTECTION IMPACT ASSESSMENTS IN DATA SPACES

Data protection engineering can be helpful in terms of (semi-)automating the necessary information collection and delivery with respect to the DPIAs performed by the data controllers. As a key activity in every DPIA is the elicitation and assessment of risks to the rights and freedoms of the data subjects concerned, the data intermediary may perform such an activity for its own systems and services once and provide the identified risks and correlated relevant information to the data controllers automatically. Thus, the data controllers can incorporate this list of risks in their DPIA.

It is essential to understand that the process of a DPIA requires more than just the concatenation of elicited risk lists of data controller(s) and data processor(s). Additional risks may stem from the constellation of collaborating entities (i.e., data controllers and data processors), and therefore depend on the exact interactions substantiated in the processing. These inter-organisational risks can only be observed when analysing the collaborations and the processing operation as a whole. For example, let us consider the case where encrypted personal data are stored at one data processor in the processing chain and the corresponding encryption key is stored at another data processor. Performing the DPIA for each of these separately may result in low risks. Either, the data are encrypted (and thus protected from adversaries), or the data are not even stored (except for the decryption key). Individually, it is probable that neither risk would score very high in a DPIA assessment.

However, if those two data processors happen to utilise the same data processor for the actual storage, the combination of both instances may become a severe risk to the data processing, as now both the decryption key and the encrypted data are in the hands of the same organisation (and all of its potential hackers that have data access). As can be seen, the choice and constellation of data controllers and processors is a highly relevant aspect of every compositional DPIA, which cannot be prepared statically before a collaboration scenario actually manifests. Therein, these compositional risks are different from the single organisation risks like local power outages or security staff assessment. To coclude, the task of supporting a holistic DPIA with multiple data controllers and data intermediaries is a non-trivial one. It requires close attention during the engineering of the whole processing operation and the deployment of the EU data space.

## 2.5 MAIN ACCOUNTABILITY BUILDING BLOCKS

An additional aspect related to establishing trust lies within the notion of accountability (principle) of the controller. The controller is responsible for, and must be able to demonstrate compliance with the personal data processing principles established in article 5(1) of the GDPR. Thus, it is the obligation of the controller to take the necessary measures in order to comply with the requirements of the GDPR, and be able to demonstrate such compliance at any time, without the need for the supervisory authority to make specific enquiries and requests to assess conformity, while exercising its powers.

Whether the controllers or processors are public or private entities, all data holders willing to promote the re-use of personal data for social and economic good must demonstrate accountability, by means of, as applicable, revamped internal mechanisms (policies, procedures, risk-based assessments, controls, and other measures related to data sharing), data sharing agreements and sensible privacy management programmes (PMPs).

Based on the DGA provisions on EU data spaces, the main building blocks towards achieving accountability are as follows:

**1. Clear-cut identification of responsibilities and obligations for data holders and data users**

Data holders must comply with their legal obligations, such as those under the GDPR (i.e. establish a lawful basis before sharing the personal data with any other party), as much as data users that receive the personal data (i.e. establishing on the basis of which lawful basis they intend to carry out their processing activity). Such obligations can usefully be laid down in an agreement (see principle 9.).

**2. Effective internal governance of personal data sharing**

Efficient handling of responsibilities and obligations deriving from data sharing (i.e. drafting of data sharing agreements, additional technical and organisational measures to adopt & implement) is required.  This governance model should specifically frame instances in which data sharing involves the concerted monitoring of data processing with data processors and sub-processors (or also with intermediaries and other competent third parties).

**3. Cooperative external governance of personal data sharing**

Define how data holders will cooperate with each other (targeted partners) within sectoral bodies and competent authorities but also with European Commission, the European Data Innovation Board (EDIB), and any stakeholder in charge of better framing re-use of data (for instance notably regarding possible handling of data breaches').

**4. Implementation of Data sharing programme**

Define policies, procedures and other measures to ensure that data holders will remain accountable while sharing personal data, effectively mitigating the risks arising from such data sharing.

**5. Design of targeted Data Sharing Accountability tools**

Reduce the risks resulting from personal data sharing, including ad hoc access and data re-use centric security mechanisms and any supplementary due diligence to impose on data holders or users.

**6. Balance security / risk-mitigation objectives and the need for sufficient quality of data to be shared**

Data holders and data users must embed data protection into the design of applications, devices, and systems (i.e. PETS) while making sure such measures do not deprive data users from using qualitative, relevant and fairly reliable data. In practice, data sharing efficiency should be assessed in relation to data security as much as on the front of data quality.

**7. Ethical assessment of envisaged data sharing practices**

Targeted assessments should consider risks tied to data sharing processing (i.e. absence of any illegal, unfair or deceptive practices, or of any intent to share personal data to harm or disadvantage an individual or a group of individuals) versus the benefits of such data sharing processing (i.e. any public interest which would suffer from the absence of re-use of data for public good).

**8. Transparent information sharing between data holder and data user**

The recipient of personal data must carry out a targeted risk assessment in relation to the intended purpose for processing, making it clear to the data holder. The data holder must consider any additional safeguard or control it may wish to impose on the recipient to ensure the security, fairness and confidentiality of data.

**9. Contractual framing of data sharing practices within EU data spaces (access) and from one data space to another (interoperability) through sector-based or targeted data sharing agreements**

Data holders and data users need to consider their specific responsibilities and obligations and frame them clearly, in a way that is appropriate and proportionate to the identified risks (case-by-case analysis). Defining respective responsibilities, setting binding obligations and determining liability framing is essential to create trust. Specifically, agreements might detail the specific qualifications of parties as data holder, data user, data controller, processor, sub-processor, intermediary or third party under the data sharing arrangement. Optionally, due diligence to ensure that all personal data were lawfully collected and transparency information was provided could also be framed by clauses referring to safeguards such as voluntary and transparent limitation of data uses or targeted contractual safeguards

**10. Transparency towards individuals**

Both data holders and data users must ensure that individuals understand how their personal data is being shared, re-used and let them know how they can exercise their rights in practice (i.e. a right to opt out of data sharing, right to deletion). Such “transparency” obligations might depend on whether data sharing is mandated by law / decided by the public sector or results from ad hoc or case-by-case decisions.

Data holders and data users, undertaking these roles under the Data Act similarly assume accountability obligations which may be addressed through the implementation of the above-mentioned building blocks.

## 2.6 EFFICIENT EU DATA SPACES THROUGH SAFEGUARDS AND TRUSTED INTERMEDIARIES

Accountability programmes and handling are not the only way to create a common, consistent and standardised EU backbone for effective and interoperable EU data spaces in practice. Unambiguously, before setting data sharing projects, data holders would benefit from considering whether a data sharing agreement is needed with recipients to fulfil their accountability obligations or mitigate identified risks to individuals -notably specifying the purpose of data sharing, defining security measures and ensuring that each party is clear about its roles and responsibilities, its respective governance obligations and liability provisions. Separately, data intermediaries, whether “designed to help individuals exercise their rights under the General Data Protection Regulation (GDPR)” or to “facilitate the aggregation and exchange of substantial amounts of relevant data” and strengthen the “efficient pooling of data as well as to the facilitation of bilateral data sharing”, will have a crucial part to play, even if their role and obligations are still to be fine-tuned and evidenced in practice. With an essential feature of the intermediary being that they “do not use the data exchanged for any other purpose”, technical and organisational measures will have to be agreed upon and specified in practice.

# 3. HEALTH - PHARMACEUTICAL USE CASES

## 3.1 BACKGROUND

The EU pharmaceutical strategy [17] was announced in 2020 and aimed to address various challenges and opportunities within the pharmaceutical sector to ensure the availability, accessibility, affordability, and sustainability of medicines for EU citizens. Currently, EU pharmaceutical legislation has enabled the authorisation of safe, efficacious and high-quality medicinal products. However, there is also a growing problem of shortages of medicinal products for many EU/EEA countries, as manifested in the explanatory memorandum of the recent EC proposal for a directive on medicinal products for human use [18]. Further to that, there is also a growing need for scientific support and accelerated assessment and authorisation of medicinal products that offer therapeutic advancement in areas of unmet medical needs.

The current use case discusses a pharmaceutical data space as a possible mean to address the availability of pharmaceutical products on the market based on current needs, possible future needs and vigilance in relation to the use of pharmaceutical products. National Health Authorities are seeking to ensure the availability of pharmaceutical products on the market based both current needs and on possible future needs (for example due to possible increases in the number of specific treatments, diseases per geographical region etc). The initial data for such analyses can be obtained from prescription data, pharmaceutical companies offering the products to the market, healthcare providers, research institutions and national regulatory authorities.

## 3.2 PROBLEM(S) DEFINITION

The envisioned pharmaceutical data space aims to support the following analyses for the data user, which in this case is the National Health Authority:

- Availability of pharmaceutical products in the market: This analysis will be performed based on the prescription data from past years, the availability of pharmaceutical products by pharmaceutical companies and indicators from research institutions for possible eminent needs due possible rises of specific diseases. The analysis will be performed at geographic regions level.
- Research and analysis on the efficiency of pharmaceutical products: This analysis will be performed based on prescription data and data on what was the prescribed medication for each medical diagnosis.

As data communicated to the data intermediary may very well include personal data, specific attention has to be paid by the data holders and the data intermediary to the level of protection of these data. Furthermore, there is an important aspect that has to be highlighted with regards to whether the scope of processing these personal data falls within the primary use of the initial collection or they falls under the secondary use of data and the data controllers have to perform an assessment of whether they are compatible with the initial purpose of collection.

## 3.3 USE CASE - AVAILABILITY OF PHARMACEUTICAL PRODUCTS IN THE MARKET

One of the intended uses of the envisioned pharmaceutical data space is to ensure the availability of pharmaceutical products in the market. It is assumed that three main type of data holders exist, which are listed below along with the information that each data holder type shares with the data intermediary. Each data holder stores additional data to the one presented above however only the data listed below are deemed as necessary for the provision of the service to the data user.

- The **National electronic prescription system** that shares information related to pharmaceutical prescriptions;

| Social Security Number | Date of Birth | Gender | Post Code | Prescribed Medication | Dosage | Symptoms | Date of Prescription | Prescription Duration |
|---|---|---|---|---|---|---|---|---|

- The **pharmaceutical companies** that share information on each medication they make available to the market;

| Medication | Description | Quantity Available |
|---|---|---|

- The **healthcare providers** that share information with regard to the medication that is used for diagnosed medical conditions and possible unwanted interactions between different medications.

| Medical Condition | Unwanted Medication Interaction | Interaction Symptoms Indicators |
|---|---|---|

The data user is the National regulatory authority which seeks to collect information on what is the current state of prescriptions for medication combined with the availability of products from pharmaceutical companies and the need for alternative pharmaceutical products in cases where specific combinations are not recommended.

### 3.3.1 Technologies to be used

One of the design goals, in terms of implementing data protection engineering, is that the intermediary should be able to respond to requests from the data user (National regulatory authority), without being able to identify or single out individuals. To achieve this goal, specific masking data protection engineering techniques should be applied by the data holders when sharing the data, as show below.

1. The **National electronic prescription provider** creates an identifier for each record to be shared by replacing specific fields with a deterministically generated pseudonym, based on a key *k* that is known only to the data holder. The same *k* is applied to all records. This could be, for example, a keyed hash function (e.g. a Message Authentication Code - MAC) as described in [19]. In the current scenario the Social Security Number (SSN) can be used as the identifier.
2. The data set shared by the prescription provider cannot be considered as fully pseudonymised as not all data protection risks have been addressed. Re-identification risks still exist due to the so-called quasi-identifiers [20], therefore they need to be properly masked by the data holder. In this respect, techniques such as attribute generalization. In our use case, such quasi-identifiers (and their possible generalizations) are the following:

i) Date of birth: it is replaced with a range of ages (i.e. 50-55)
ii) Postal Code: it is replaced with the first three characters of the postal code; these three numbers should suffice to provide information on the wider area of residence.
iii) Date of Prescription: it is replaced only with the month and the year instead of the full date.

The degree for each of the above generalizations is depended on the level of risk of the resulting output that will not allow re-identification or singling out with respect to an individual [21].

The **pharmaceutical companies** and **the healthcare providers** do not share any personal data and therefore, from a data protection engineering point of view, no masking is required.

### 3.3.2 Considerations

The use case described above envisions a data sharing scenario where the masking and generalisation are **performed by the data holders**, prior to sharing the data sets with the data intermediary. However, apart from the fact that, by these means, the data holders become accountable for implementing strong generalization, this approach poses several challenges also from an implementation point of view.

More precisely, it is essential to establish a mechanism to ensure that the various data holders implement generalisation on the same level; for example, if a data holder chooses a generalised “45-50” field for the age, there should not be another data holder that chooses a different generalised field such as, e.g., “40-50”. Hence, it seems that this distributed nature of the data holders pose some limitations since there is a need for them to “jointly agree” on some parameters.Taking into account that the proper selection of these parameters is highly contingent on each specific dataset, this is not a straightforward task.

**The decisions on how each attribute will be generalised is strongly related on the level of risk of re-identification or singling out an individual.**

## 3.4 USE CASE - RESEARCH AND ANALYSIS ON THE EFFICIENCY OF PHARMACEUTICAL PRODUCTS

An additional possible utilization of the envisioned pharmaceutical data space is to support research and analysis on the efficiency of pharmaceutical products. For simplicity, it is assumed that only the two data holder types exist, similar to the previous use case, which are listed below along with the information that each one shares with the data intermediary. Each data holder stores additional data to those presented; however, only the data listed below are deemed as necessary for the provision of the service to the data user.

- The **National electronic prescription system** that shares information related to pharmaceutical prescriptions;

| Social Security Number | Date of Birth | Gender | Post Code | Prescribed Medication | Dosage | Symptoms | Date of Prescription | Prescription Duration |
|---|---|---|---|---|---|---|---|---|

- The **healthcare providers** that share information with regard to:
    - the medication that is used for diagnosed medical conditions and possible unwanted interactions between different medications.

| Medical Condition | Unwanted Medication Interaction | Interaction Symptoms Indicators |
|---|---|---|

- The medical diagnoses, lab and test results and prescribed medication of treated patients.

| Social Security Number | Medical Diagnoses | Lab and test results | Prescribed Medication |
|---|---|---|---|

The data users are research institutions which seeks to collect information on how effective pharmaceutical products have been in treating specific symptoms and how unwanted medication interactions have affected their effectiveness.

### 3.4.1 Technologies to be used

Within this use case, there are two design goals in terms of implementing data protection engineering. The first goal is that the intermediary should be able to respond to requests from the data users (research institutions), without being able to identify or single out individuals. The second goal is that the data users should also not be able to identify or single out individuals but also not be able to correlate data. To achieve these two goals, specific masking data protection engineering techniques should be applied by the data holders when sharing the data.

1. The **National electronic prescription provider** masks parts of the data set to be shared similar to the previous use case. Once more the Social Security Number (SSN) field is replaced with a deterministically generated pseudonym, based on a key *k* that is known only to the data holder and the quasi-identifiers are replaced with ranges.
   It is again noted that the degree for each of the above generalizations is depended on the level of risk of the resulting output will not allow re-identification or singling out with respect to an individual [16].
2. The **healthcare providers** masks also the SSN part of the data set to be shared with a deterministically generated pseudonym, based on a key *k* that is known only to the data holder.

Since the same field will be pseudonymized by different data holders with a different key, the data intermediary will not be able to corelate the data received from different data holders that refer to the same SSN.

3. The **intermediary** will utilise Polymorphic Encryption and Pseudonymisation (PEP) [22] for the data sets that will be transmitted to the data users and will act as the transcryptor, as discussed also in [11]. Each data set will be assigned different pseudonyms for each data user, thus preventing pseudonym linking across multiple data users. The intermediary here, even if masking already pseudonymised/generalized data, acts as a trusted third-party pseudonymising entity [19].

### 3.4.2 Considerations

The use case described above envisions a data sharing scenario where the masking and generalisation are **performed by the data holders**, prior to sharing the data sets with the data intermediary but additional masking is **performed by the data intermediary.** This operation strengthens the role of the intermediary as trustworthy organisers instead of only brokers of data sharing. However, it also increases the intermederiary's responsibilities and obligations as discussed under section 2.4.

Further to the interoperability of the generalization discussed in the previous use cases, the additional role of the intermediary brings forward the additional responsibilities that have to be fulfilled. Even without analysing whether the intermediary should be considered as a controller or a processor, it has to be able to cope with the needs and rights of the data subjects and data users, keep track of the data sources and data processing tasks and possibly evaluate and update data usage policies at multiple stages throughout the data processing lifecycle.

# 4. CONCLUSIONS

Common European Data Spaces is an emerging concept, outlined by the European Strategy for Data, which aim to foster European activities toward data economy. Data Spaces is an umbrella term corresponding to any ecosystem of possible interactions between public and private sector entities alongside new governance and business processes. These competencies [23] will have to follow a data engineering approach to meet all the requirements and legal obligations.

Data protection engineering is not a mere “compliance tool” for the GDPR. In implementing appropriate measures and the necessary safeguards to reinforce data protection principles and to enable the exercise of individuals’ rights, data protection engineering offers data controllers a practical option for data sharing while minimising the risk of information misuse, data breaches or other security threats [12]. The development of compelling and convincing use cases for safe and lawful data sharing is one of the most crucial challenges for the successful implementation of Common European data spaces.

Building up on the definition of the main actors and the DGA provisions around the EU Data Spaces, the identification of building blocks and requirements represents the starting point for their successful development and deployment. Within the scope of this report, we attempted to provide such set of building blocks with regards to the accountability of the controller(s) and the processor(s). These building blocks are intended to cover applicable, revamped internal mechanisms (policies, procedures, risk-based assessments, technical and organisational controls, and other measures related to data sharing), data sharing agreements and sensible privacy management programs (PMPs).

While attempting to analyse further the embodiment of personal data masking techniques into specific processing operations, this report outlined an envisioned data space in the pharmaceutical area. Through two specific use cases, we showcased the different roles and responsibilities that can be assigned to data intermediaries by the data holders with regards to personal data protection. Given the different data protection goals of each use case, we demonstrated how the intermediary could be actively involved in the pseudonymised data masking process. Even if this approach has already been discussed in a typical data sharing scenario [11], it can be further exploited as it provides new incentives for data sovereignty and data governance considerations. Further to that, we also showcased how specific masking and generalisation techniques can be practically deployed by the data holders and the data intermediary.

Despite the potential of EU data spaces, there are still considerations regarding the appropriate technical and organisational measures and how to engineer them into practise, both from a data protection but also from a cybersecurity point of view. Even if there are already a good number of privacy enhancing technologies that can support us in meeting specific data protection goals, we should not neglect the fact that we are called to address new processing operations, where roles and responsibilities are not always clearly defined.

# 5. REFERENCES

[1] Joint Research Centre, "Emerging approaches for data-driven innovation in Europe," 2022.

[2] ENISA, "Big Data Threat Landscape," 2016.

[3] ENISA, "Big Data Security," 2016.

[4] ENISA, "Privacy by design in big data," 2015.

[5] European Commission, "A European strategy for data," 2020.

[6] European Union, "Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)," 2022.

[7] Joint Research Centre (European Commission), "European data spaces: Scientific insights into data sharing and utilisation at scale," 2023.

[8] Directorate-General for Financial Stability, Financial Services and Capital Markets Union, "Report on open finance," 2022.

[9] European Union, "Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification (Cybersecurity Act)," 2019.

[10] AEPD, "Approach to Data Spaces from GDPR Perspective," 2023.

[11] ENISA, "Engineering Personal Data Sharing," 2023.

[12] ENISA, "Data Protection Engineering: From Theory to Practice," 2022.

[13] AEPD, "Approach to Data Spaces from GDPR Perspective," 2023.

[14] ENISA, "Privacy and Data Protection by Design," 2015.

[15] ENISA, "Deploying Pseudonymisation Techniques," 2022.

[16] ENISA, "Data Pseudonymisation: Advanced Techniques and Use Cases," 2021.

[17] European Commission, "COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Pharmaceutical Strategy for Europe," 2020.

[18] European Commission, "Proposal for a directive of the European Parliasment and of the Council on the Union code relating to medicinal products for human use, and repealing Directive 2001/83/EC and Directive 2009/35/EC," 2023.

[19] ENISA, "Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation," 2019.

[20] L. Sweeney, "k-anonymity: a model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 10, no. 5, p. 557–570, 2002.

[21] B. Chen, D. Kifer and K. LeFevre, "Privacy-Preserving Data Publishing," *Foundations and Trends in Databases,* vol. 2, pp. 1-167, 2009.

[22] M. Hildebrandt, E. Verheul, B. Jacobs, C. Meijer and J. de Ruiter, "Polymorphic Encryption and Pseudonymisation for Personalised Healthcare: A Whitepaper," 2016.

[23] S. Scerri, T. Tuikka, I. de Vallejo and E. Curry, "Common European Data Spaces: Challenges and Opportunities," in *Data Spaces*, Springer, Cham, 2022.

[24] M. Hildebrandt, E. Verheul, B. Jacobs, C. Meijer and J. de Ruiter, "Polymorphic Encryption and Pseudonymisation for Personalised Healthcare: A Whitepaper," 2016.

[25] ENISA, "Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation," 2019.

[illegible] at www.enisa.europa.

[illegible] ENISA is the European Network and Informat