Open Statistical Data Interoperability Framework

www.cef-interstat.eu

D1.1: Report on the maturity of Linked Open Statistical Data in Europe

Co-financed by the Connecting Europe Facility of the European Union

Project full title
INTERSTAT - Open Statistical Data Interoperability Framework

Grant Agreement No.
INEA/CEF/ICT/A2019/2063524

Project Document Number
Deliverable 1.1 (Activity 1)

Project Document Delivery Date
22.03.2021 (v2.0)

Deliverable Type and Security
Report – Public

Authors
ISTAT: Adele Maria Bianco, Paolo Francescangeli, Marina Gandolfo, Giuseppina Ruocco, Monica Scannapieco, Carlo Vaccari

Contributors
INSEE: Franck Cotton

Reviewers
ENG: Martino Maggio, Francesco Arigliano
Stéphane Dufour and Davide Taibi (External Expert Advisory Board)

Co-financed by the Connecting Europe Facility of the European Union

Table of Contents

Co-financed by the Connecting Europe Facility of the European Union

List of figures

List of tables

Co-financed by the Connecting Europe Facility of the European Union

Executive Summary

Early in the 1990s, the Open Data movement spread, pushing for full availability of open licensed data to all end users. More recently, awareness of the need to share not only data, but also their meaning, has matured, introducing a semantic layer that allows for the integration of data definitions, and this led to the development of the Linked Open Data (LOD).

LOD is structured data interlinked with other data, so that they can be retrieved through semantic queries. LOD are available not only for human readers, because they share information in a way readable by computers.

In recent years in many European countries, National Statistical Institutes (NSIs) are using Linked Open Statistical Data (LOSD) to disseminate data. Many experts consider LOSD the future standard for the dissemination of Statistical Data.

In this document, you can find the description of the maturity of Linked Open Statistical Data in Europe.

We start introducing the basic concepts, laws and directives composing the legal framework for LOSD. Then we analyse the main initiatives and projects carried out in Europe on the use of LOD in Official Statistics. In the final chapters, we try to give the elements for the definition of a LOSD strategy and some guidelines to overcome legal and procedural barriers.

Note: Introduction section is mandatory and shall always have the structure as in this template. Please note that there are explanatory notes with the various examples in the following sections.

About this document

The maturity of the Linked Opens Statistical Data usage in Europe is analysed.

Intended audience

The document should be useful for people who need to understand the recent development in the field of Linked Open Statistical Data. It can be useful both for NSIs managers, for statistician and for IT staff. In addition, users of Statistical portal can find it useful to understand the framework in which LOSD have been developed and implemented in European countries.

Reading recommendations

This document is divided into five sections:

Section 1 provides an overview of the project and the intended outcome of the deliverable.

Section 2 highlights the principles, rules and policies that have a greater impact on Open Statistical Data, and more in general on data dissemination. The analysis starts from the statistical principles stated in the Fundamental Principles of Official Statistics (FPOS) and embedded in the European Statistics Code of Practice (ES CoP). Then we examine the Generic Law on Official Statistics and the Open Data Directive, to describe the latest changes introduced in the legal framework to regulate and support interoperability. In addition, the reference to official statistical standards allows addressing the different aspects of Open Statistical Data from a broader perspective. This chapter relates to Strategy and Policy dimension and provides a general overview of legal barriers, issues and solutions.

Section 3 summarizes the main initiatives launched to improve skills, build capabilities, and or infrastructures. We describe each initiative in terms of general context, main objectives and key outcomes.

Co-financed by the Connecting Europe Facility of the European Union

Section 4 deals with LOSD maturity assessment, based on the analysis of the state of play in some ESS countries. Moreover, the inventory of drivers and stakeholders involved in LOSD dissemination and use facilitates the understanding of the joint ESS LOD strategy proposed for improving interoperability. In the end, we provide an overview of the main insights gained from the latest projects with the analysis of joint activities with high priority.

Section 5 summarizes the legal and procedural barriers highlighted in the previous sections and provides a set of architectural guidelines to deal with internal issues preventing LOSD publishing and turn challenges in opportunities.

Co-financed by the Connecting Europe Facility of the European Union

1 Legal framework and reference standards

1.1 LOSD key concepts

The modernization program within the ESS Vision 2020 has launched several initiatives to improve data accessibility and reliability, and to promote a proactive commitment with different types of users, to meet their requirements and expectations changing over time. In accordance with their institutional role of official statistics providers, many National Statistical Institutes (NSIs) have improved data access and discoverability by enhancing the semantic operability of disseminated results. In the last years, the number of NSIs releasing official statistics in open/public format has increased, thus allowing the linkage with other sources and providing data provenance information.

The term Open[1] Data refers to information that can be freely accessed, used, and reused by anyone for any purpose also for commercial use, subject only to the attribution specification . The main features of Open Data are availability and accessibility, achieved by reducing legal or technological constraints. More precisely, Open Data are published in a machine-readable format and can be navigated and downloaded via web and under an open license. The drive to increase Open Data dissemination has several implications, such as the use of neutral data formats (e.g., CSV and RDF) instead of proprietary ones, or the difficult application of standard methods for disclosure control due to the open licence.

Linked Open Data (LOD) results from the combination of Open Data and Linked Data technologies. Linked Data refers to a set of design principles that allow sharing machine-readable interlinked data between different organization and on the Web. Linked Data centers on the following principles [1] recommended by Tim Berners-Lee, who created the Web and launched the Linked Data project:

1. Use URIs[2] to name things
2. Use HTTP URIs so that things can be referred to and looked up ("dereferenced") by people and user agents
3. When someone looks up a URI, provide useful information, using the open Web standards such as RDF, SPARQL
4. Include links to other related things using their URIs when publishing on the Web.

These principles underline the use of open standards and a semantic approach to fulfil users' needs and facilitate linked data search. A great amount of Open Data corresponds to statistical information related to different domains, such as economy, environment, and socio-demography In recent times, many NSIs and public authorities have started to disseminate official statistics adopting the Linked Open Data framework, thus creating Linked Open Statistical Data (LOSD). The benefits of LOSD in terms of data discoverability and enrichment have underlined the need for a semantic data and metadata standardization to increase interoperability. European countries launched many initiatives,

[1] The Full Open Definition provided by the Open Knowledge Foundation is available at the following link https://opendefinition.org/od/2.1/en/

[2]Wikipedia defines URI as "A Uniform Resource Identifier (URI) is a string of characters that unambiguously identifies a particular resource". https://en.wikipedia.org/wiki/Uniform_Resource_Identifier

Co-financed by the Connecting Europe Facility of the European Union

engaging several stakeholders, in order to assess the state of play of LOSD at national and European level, and identify tasks and capabilities that allow delivering Open Data within the official statistical system. One of the main outcomes of these initiatives is a joint ESS LOD strategy, based on the combination of the following five dimensions:

- Strategy and Policy
- People and Capacities
- Data and Metadata
- Governance
- Technology and Infrastructure (analysed more in detail in deliverable D.1.2).

The next chapters explore several aspects related to the first four dimensions, to identify the barriers that prevent the dissemination of statistical output according to the LOSD paradigm.

1.2 Legal framework principles

The legal basis plays a crucial role in safeguarding relevance, accuracy and trust in official statistics. It also supports the cross-border harmonization of statistical processes, methodologies and tools, thus increasing the comparability of survey results between different countries. The statistical legal framework is built upon principles and regulations that foster the production of quality official statistics, and more in general, the modernisation of the statistical system. The UNECE countries adopted the Fundamental Principles of Official Statistics (FPOS) [2] in 1992, to provide reliable official statistics affecting the decision making of several stakeholders.

In brief, the principles and their statements are the following:

- Principle 1. Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.
- Principle 2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.
- Principle 3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.
- Principle 4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.
- Principle 5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.
- Principle 6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.
- Principle 7. The laws, regulations and measures under which the statistical systems operate are to be made public.

Co-financed by the Connecting Europe Facility of the European Union

- Principle 8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.
- Principle 9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.
- Principle 10. Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

1.2.1 The European Statistics Code of Practice

The European Statistics Code of Practice (ES CoP) [3] is the foundation of the Common Quality Framework [4], establishing minimum level of quality standards to improve the quality of official statistics and build trust in statistical authorities. It is based on Principles covering three fundamental aspects, not only the statistical outputs but also institutional environment, and statistical processes. The aim is to strengthen the credibility of official statistics and to guarantee an institutional framework to ensure professional independence, compliance with European standards and statistical principles, harmonisation and comparability of data as well as equal access to all users. The Code is consistent and follows the UN Fundamental Principles of Official statistics. It has been developed and adopted in May 2005 as a Recommendation of the Commission in the Communication on the independence, integrity and accountability of the national and Community statistical authorities.

As a whole, the European Statistical System operates efficiently and satisfies largely the requirements of independence, integrity and accountability. The massive quantity and variety of statistics produced and disseminated by the national and the EU statistical authorities within that System complies with strict requirements on quality and reliability.

The Code of Practice on European Statistics represents a self-regulatory instrument containing standards for the independence of the national and EU statistical authorities, providing a further guarantee for the good functioning of the European Statistical System (ESS) and the production of high quality and reliable statistics. In 2011 and 2017, the ES CoP has been revised considering the modernisation process of European statistics, which drives to several innovations like the use of other sources, the investment in open data, and the major role of National Statistical Organisations (NSOs) as stated in the European statistical Law [5].

Based on a common European definition of quality, the ES CoP contains 16 principles related to the development, production, and dissemination of European official statistics. A set of indicators provides an assessment of the compliance with the principles. As a whole, the principles focus on three main aspects: institutional environment, statistical processes and statistical output.

Within the statistical output, the attention to open data is clearly expressed in principle 15 Accessibility and Clarity and in particular in the indicator 15.2, in which is stated that Dissemination services use modern information and communication technology, methods, platforms and open data standards.

Rounds of Peer Reviews monitor the implementation of ES CoP strategic steps by the members of the ESS. After the two rounds in 2006-2008 and in the 2013-2015, a new round will take place in the 2021-

Co-financed by the Connecting Europe Facility of the European Union

2023. The Peer reviewers develop final reports that analyse the main innovations. The future round will make evident several innovations also on open data.

1.2.2 LOSD and the Generic statistical Law

The "Generic Law on Official Statistics" (GLOS) [6] is another milestone of the legal framework, developed by UNECE, Eurostat and EFTA, supporting the development of a solid legal basis to harmonise the activities of the National Statistical System within the countries of Eastern Europe, Caucasus and Central Asia. The regulation, endorsed by the Conference of European Statisticians (CES), is composed by 38 articles that can be adapted to the national context to a certain extent, due to the compliance of the law with the FPOS and the European Statistics Code of Practice. In 2016, in order to help member countries to align their statistical national frameworks with GLOS, the CES Bureau established a Task Force to provide guidance and harmonize the national legal frameworks. The common elements of statistical legislation identified in the document [7] produced by the TF relate to:

1. Objective and scope of the Law
2. Main principles and definitions of official statistics
3. Organization of the National Statistical System
4. Statistical advisory council and other advisory bodies
5. Coordination of the National Statistical System and statistical programmes
6. Data collection
7. Statistical confidentiality
8. Quality of official statistics
9. Dissemination
10. Statistical services
11. International cooperation
12. Infringements
13. Relationship to other legislation.

In addition, an entire chapter of the guidance document is dedicated to the “Emergence issues related to statistical legislation”. The dissemination of official statistics as Open Data is one of the several subjects analysed in this chapter. After the description of Open Data and LOD features and a general assessment of the state of play, the document provides a set of legal considerations to start clearing the relationship between LOD and official statistics:

1. In order to develop the full potential of official statistics as open data, NSIs need to guarantee the safe data access, storage and sharing, within a governance framework built for privacy and confidentiality protection
2. Applying the principle of symmetry, the increase of statistics available as open data and linked open data for business purposes, should correspond to an increase of private data for statistical analysis
3. From a legal perspective, collected data for statistical purposes can be used only for producing statistics. With respect to confidentiality constraints and privacy protection, NSIs may provide “open statistics” instead of the data itself
4. It could be required to identify specific areas where official principles and standards need to be adapted, although statistical concepts and classifications developed for the statistical production must take into account statistical needs only

Co-financed by the Connecting Europe Facility of the European Union

5. Even if most of official statistics are already open by definition, NSIs could collaborate with several stakeholders, to develop common data standards and data governance frameworks, thus benefiting from emerging innovations and technologies
6. The update of the statistical legislation involves the revision of other related legislation, particularly the regulations dealing with open data, access to data, data rights, etc. During the revision of the Statistical Law, a consultation with the stakeholders involved in open data initiatives would be useful
7. In order to perform data linking, reuse and provide LOD, NSI_s need to cooperate with other members of data communities, defining roles and responsibilities in the project agreements. Among the cooperation activities, NSIs could share their expertise with other authorities planning to increase the dissemination of Open Data.

In addition to the previous analysis, the TF has compared the Open Data principles and the principles applied to official statistics (mainly FPOS), highlighting the differences between them. The following table reports the results of the comparison.

Open data principles	**Official statistics principles**
1. Data must be complete. All public data are made available as open data	Official statistics are compiled on priority data needs of society
2. Data must be primary. Data are published as collected at the source, with the finest possible level of granularity, and not in aggregate or modified forms	Primary data collected for official statistics can only be used for statistical purposes and thus published as statistics only
3. Data must be timely. Data are made available as quickly as necessary to preserve the value of the data	Official statistics need to consider the trade-off between accuracy and timeliness
4. Data must be accessible. Data are available to the widest range of users for the widest range of purposes	Official statistics provide equal and simultaneous access to statistics for all users
5. Data must be machine-processable. Data are structured in a manner that they can be processed in an automated way	Official statistics apply internationally agreed classifications, definitions, and methods to ensure wide comparability
6. Access must be non-discriminatory. Data are available to anyone, with no registration requirement	Official statistics are available to all users preventing disclosure of individual data
7. Data formats must be non-proprietary. Data are available in a format over which no entity has exclusive control	Official statistics can be used freely with reference to the producer of the statistics as the source of information.
8. Data must be license-free. Data are not subject to anyone's copyright, patent, trademark or trade secrets regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statues	Official statistics should be available free of restrictions by referring to the statistical office as the source. However, statistical data are subject to strict confidentiality rules that influence their availability.
9. Permanence refers to the capability of finding information over time	Official statistics are produced regularly over time

Co-financed by the Connecting Europe Facility of the European Union

10. Usage costs. One of the greatest barriers to access to ostensibly public available information is the cost imposed on the public access	Official statistics cannot always be made available at disaggregated level due to confidentiality constraints, privacy protection and all possible non-confidential combinations are too costly to produce.

Table 1 - Open data principles vs Official statistics principles (Source: Guidance on common elements of statistical legislation)

1.2.3 Open Data Directive

Data are an essential resource for decision-making process, for measuring economic growth, job creation and social progress. The Charter of Fundamental Rights of the European Union (Charter) [8] states that everyone has the right to freedom of expression, including the freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. The Public sector in Member States collects, analyzes, produces, reproduces, and disseminates a wide range of information in many areas of activity, such as social, political, economic, legal, geographical, environmental, meteorological, seismic, touristic, business, patent-related and educational areas. Public sector bodies are increasingly making their documents available for reuse in a proactive manner, by ensuring online discoverability and actual availability of documents. The availability of associated metadata in an open format becomes essential to ensure interoperability, reuse and accessibility. In this respect, the NSO_s of the countries have a leading role on the development, production and dissemination of huge amounts of data and related metadata to be used as information for decisions. Their coordinating role in the National Statistical System represents a value added in coordinating other sources and their access for producing and integrating information.

This information is a valuable resource for the digital economy: not only it is used as input for creating data-driven applications and services, but it also ensures greater efficiency in providing public and private services and making more informed decisions to stimulate economic growth and to promote social commitment. Allowing the reuse of documents held by a public sector body adds value for the benefit of reusers, end users and society in general, by promoting transparency and accountability, and allowing the public sector body to improve the quality of the information collected and the performance of its tasks. The European Union is well aware of it and in recent years developed the European Data Strategy [9], which aims to create a single European data space, a truly single market where personal and non-personal data, including confidential data, are secure. This will facilitate access for businesses and public authorities to an almost infinite amount of high-quality data, to stimulate growth and create value. Relevant to the achievements of this common space, it is also the promotion of open data and the reuse of public data available in the space.

The European Union has been committed in the development of a legal framework for open data and the reuse of data held by the public sector since 2003 (Directive on the reuse of public sector information - Public Sector Information, PSI). The main aim was to promote and facilitate the reuse of public sector information by harmonizing the basic conditions for such reuse and removing the main obstacles to it in the internal market. European Commission substantially amended in 2013 the directive introducing the obligation to allow the reuse of generally accessible public data, the introduction of a standard charging rule that limits tariffs to marginal costs for the reproduction, and the obligation for public bodies to ensure greater transparency.

Co-financed by the Connecting Europe Facility of the European Union

After 5 years, the European Commission, in consideration of the profound changes that have occurred and the main challenges essentially related to the progress of digital technologies, felt again the need for a revision of the framework legislation on open data and the reuse of public information. In 2018 the European Commission developed an important package of measures "Towards a common European data space" [10], as a fundamental step towards the creation of a common data space in the EU. The aim is to contribute to the strengthening of the EU data economy through initiatives to increase the amount of public sector data made available for reuse. The review should allow for easier and faster access to data held by the public sector, an increase in the supply of data to support the economy and a modern data-driven society and the promotion of the use of open data in the EU.

Within this approach, the EU, the Publications Office of the European Union and the European Commission developed an initiative of the European Data Portal (EDP) aiming at increasing the impact of open data by making it easy to find and reuse by everyone. In the Portal the metadata of Public Sector Information (PSI) – data held by the public sector – are available on data portals across European countries and EU institutions. The EDP enables users to discover the data, by making its metadata searchable in all EU supported languages.

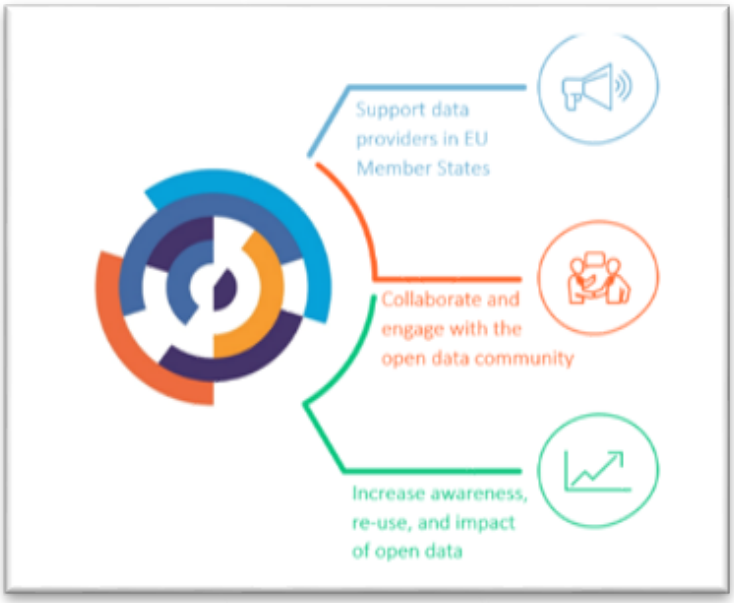


Figure 1: European Data Portal objectives[3]

The work done on the revision of the Directive PSI made possible the adoption and publication on June 2019 of the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the reuse of public sector information [11]. The new Directive allows for an easier and faster access to data held by the public sector, an increase in the supply of data to support the economy, a modern data-driven society and the promotion of the use of open data in the EU. The title of the Directive reflects the shift towards a fully free/open reuse.

The main innovations envisaged by the Open Data Directive include several relevant points.

[3] Source: https://www.europeandataportal.eu/

Co-financed by the Connecting Europe Facility of the European Union

The inclusion of a specific article concerning the list of High Value Datasets (HVD); they are documents, the reuse of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs. The HVD should be free of charge, disseminated in machine-readable formats through Application Programming Interfaces (API).

In order to support the reuse of documents having a particular high value for the economy and society, the Directive defines following list of thematic categories of high-value datasets:

1. Geospatial
2. Earth observation and environment
3. Meteorological
4. Statistics
5. Companies and company ownership
6. Mobility.

The thematic categories could cover, inter alia, postcodes, national and local maps (geospatial), energy consumption and satellite images (earth observation and environment), in situ data from instruments and weather forecasts (meteorological), demographic and economic indicators (statistics), business registers and registration identifiers (companies and company ownership), road signs and inland waterways (mobility). The amendment of the list of thematic categories of high-value datasets by adding further thematic categories is delegated to the Commission. The Commission carries out appropriate consultations during its preparatory work, including the involvement of the expert level.

The Directive reinforces the principle of free access to public sector content; free reuse becomes a principle. Some exceptions are narrowly defined besides stronger transparency. It is foreseen, for example, that a list of all public bodies that charge above marginal costs of dissemination has to be made public. The scope of the Directive is also extended, under certain conditions, to public service companies in the transport and public utility services sector. Greater availability of data in real time using API interfaces is foreseen in order to encourage the development of innovative products and services (such as applications for mobility) by businesses and, above all, start-ups.The extension of the scope of the Directive is also to publicly funded research data. The Member States have to adopt open access policies to support availability of research data. The Directive imposes new transparency and review requirements to public-private agreements that may lead to a situation in which the range of potential re-users would be severely restricted preventing the data lock-in.

The Directive includes practical arrangements to facilitate reuse. In fact, Member States shall:

- Encourage the availability of documents according to the principle of ‘Open by design and by default’
- Facilitate metadata aggregation at Union level, promote data preservation and simplify access to documents
- Ensure that practical arrangements are in place that help re-users in their search for documents available for reuse
- Facilitate the long-term availability for reuse of public sector information, in accordance with the applicable preservation policies.

Finally, the principles, rules and standards introduced by the Open Data would reduce market entry barriers, increase the availability of data, and foster a more transparent process for the establishment of public–private data arrangements.

Co-financed by the Connecting Europe Facility of the European Union

1.2.4 LOSD and legal issues

In order to complete the overview of the legal framework regulating LOSD, further legal aspects must be considered, such as Privacy and Data Protection, Copyright and Licensing, Liability and Competition. These aspects have been examined in depth in two thematic networks in a project on Legal Aspects of (reusing) Public Sector Information (LAPSI) [12]. A study [13] performed within this initiative has analysed the following legal issues which are relevant also for the official statistics domain.

Privacy and Data Protection

The increase of Open Data initiatives has highlighted the need of further analysis and tools to protect personal information that could be inferred by cross-referencing data from different sources. As a result, Open Data publishers have to balance the benefit of Open Data and the privacy protection of the data subjects [14]. These assumption leads to an extension of privacy definition, including several factors, such as harms due to privacy breach, data sensitivity, social and cultural norms, confidentiality and the public interest. Although official statistics relate to aggregate data and do not report individual information, some cross-references between different sets of data, or data linkage could reveal some attributes about an individual, making their identity inferable. For this reason, the analysis of privacy issue for LOSD must consider personal data protection. Regarding confidentiality , the ES CoP states in Principle 5 that statistical confidentiality "is guaranteed in law". Penalties are applied in case of deliberate breaches of statistical confidentiality. In order to protect data security and integrity throughout the statistical process, specific staff trainings and other measures (regulatory, administrative, technical and organizational) are adopted, according to best practices, international standards, European and national regulations.

Concerning data protection, the General Data Protection Regulation (GDPR) [14] establishes some limitations related to the purpose collection of personal data and their cross-border transfer. More specifically, Article 5.1.b, states that personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". In relation to personal data transfer across borders, Article 46.1, specifies that "a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available".

Data protection refers to data processing management and the data controller is the person responsible for processing. According to GDPR (art.4, comma 2), data processing "means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction". During personal data gathering, the controller must specify the purpose of data collection and the time limit for data storage. The controller must collect only the data needed for the task and is not allowed to keep the data after the end of the task. A study about Open data and privacy, carried out by the European data portal [16] has identified six conditions that legitimate the controller to process personal data, and at least one of them must occur for data processing. Once the controller has completed the task, gathered data can be used for another task, only if one of the specified conditions persists.

Copyright and Licensing

Co-financed by the Connecting Europe Facility of the European Union

In addition to privacy concern, copyright and database rights may have an impact on the statistical Open Data strategy adopted by NSIs. More precisely, the copyright regulation entitles the creator of a work to have the exclusive control of every reproduction and change. The access to data and its reuse can be regulated also by database law that grants to data producers the exclusive right to monitor the extraction and reuse of the whole set of data, or a consistent part of it. In general, the use of data may be established by the copyright law, by the dataset terms of use, or by the provisions included in the agreements with the different stakeholders involved in data processing. The potential different scenarios may increase the complexity of linking open statistical data and underlines the need of common guidelines. Therefore, when linking open statistical data, it would be relevant to track the entity corresponding to the data holder who may limit the exercise of copyright rights. Finally, some copyright exceptions and limitations are embedded in the regulation, and subject to the implementation of the European Union Copyright Directive (EUCD) in each Member State.

The website terms of use and reuse, defined in private law and contract regulations, complements the copyright legislation and the dataset terms of use. The licenses used in Open Data platforms may vary from open license to restrictive terms of use, limiting data availability to personal use. The main objective of open licenses is to remove legal constraints, for enabling data use and reuse. Nevertheless, open licenses do not guarantee full interoperability, due to the incompatibilities that may occur when two or more data sources have different license conditions. In order to increase Open Data dissemination, several types of open licenses have been created, such as Creative Commons (CC), Open Knowledge Foundation (OKFN). Aiming at dealing with specific requirements, not covered by standard licenses, several public authorities have created ad-hoc licenses, thus increasing heterogeneity and the burden on users that will have to read and analyse in detail the terms of use. Although open licenses increase data access and availability, they may limit the reuse with hidden or explicit legal constraints, or they could contain incompatible conditions. The addition of compatibility clauses in the terms of use allow to re-license derived data under the same license, or under compatible licenses.

Liability

The liability issue is connected to the need of keeping control over disseminated data, to avoid any modification or misuse that could damage or cause loss to someone. Public authorities try to prevent potential inappropriate use of data by limiting their access, or by inserting restrictive clauses and liability waivers in the license conditions. Therefore, some licenses clearly state that data reuse shall not deceive third parties, or misuse the information provided, the data source or its last update. Despite that, in some cases, the licenses forbid any data alteration, without specifying whether they relate to some tasks such as the combination of datasets, or raw data modification. In general, to prevent the liability risk, a set of guidelines and the specification of the tasks compliant with the legal requirements of data reuse would thus defining the boundaries of legitimate data use.

Competition

The dissemination of open statistical data may also increase the competition with private data producers that make a profit by releasing statistical products. This overlapping is particularly relevant in some sectors where private companies have invested huge resources to create large datasets, such as address registers or road databases. In this case, the availability of open statistical data may benefit new potential users at the expense of the private stakeholders who could invoke the competition regulation to avoid the reduction of their business. On the other hand, public authorities may consider the reuse of Open Data for commercial purposes a misappropriation of common goods. The debate is still in progress and stems from the social and cultural environment before turning into an additional legal barrier to data opening. The trade-off between commercial purposes and the public interest

Co-financed by the Connecting Europe Facility of the European Union

depends on the role of public entities in the information market, and the objective of their activities for producing public data.

1.2.5 The European data strategy

The European Data Strategy, launched in February 2020 by the Commission [17], announced the following initiatives:

1) A **legislative proposal for a governance framework for common European data spaces, has been** adopted on November 25th, 2020
2) An **implementing Act on high-value datasets**, planned for adoption during the first quarter of 2021
3) The Commission in the third quarter of 2021 should adopt a **Data Governance Act** proposal.

In the following, we highlight the main aspects of these initiatives, according to their relevance to INTERSTAT's work.

Legislative proposal for a governance framework for common European data spaces

Common European data spaces are concrete arrangements in which data sharing and/or data pooling can happen across countries. They are composed of a secure IT environment for processing of data by an open number of organisations as well as a set of rules of legislative, administrative and contractual nature that determine the rights of access to and processing of the data. Data are made available on a voluntary basis and can be reused against remuneration or free, depending on the decision of data holder. Common European data spaces should be implemented in strategic economic sectors and domains of public interest (such as manufacturing, green deal, mobility, health, financial, energy, agriculture or public administration) and with the technical tools and infrastructures necessary to use and exchange data and appropriate governance mechanisms in place. Common European data spaces try to foster reuse of public sector data not covered by the Open Data Directive.

Statistical offices can play a key role in the emergence and functioning of common European data spaces, in particular when the pooled data are used for public interest purposes. In addition, they can support the development of common technical standards with regard to the description, quality and interoperability of the data, which will be required to facilitate data sharing across sectors and domains. In this respect, the work carried out within INTERSTAT project can support these specific functions of the statistical offices.

Implementing Act on high-value datasets

The Open Data Directive (ODD) should improve reuse of governments' data by the private sector. It introduces the concept of high-value datasets (see Article 2, list item 10 of the ODD) that are associated with high commercial potential. As set out in chapter V of the ODD, a specific list of high-value datasets will be adopted by the European Commission by means of an implementing Act, together with conditions and arrangements for access and reuse, e.g. terms applicable to reuse, formats of data and metadata and technical arrangements for dissemination. The specific high-value datasets shall be:

- Available free of charge
- Machine readable
- Provided via Application Programme Interfaces (APIs); and
- Made available for bulk download.

Co-financed by the Connecting Europe Facility of the European Union

Statistics is explicitly listed (in Annex I to the ODD) as one of the thematic categories in scope of the high-value dataset concept. To guide the identification of the HVDs in the area of statistics a specific activity was carried out and a list of high-value datasets for statistics was produced with the identified statistical areas of Demography, Poverty and Inequality, National Accounts, Labour Market, Prices, Regional Statistics, Government Financial Statistics, Business statistics (including tourism), Health statistics. The definitive list of high-value datasets by category will be defined in an implementing Act.

Data Governance Act proposal

The Data Governance Act proposal (DGA) should aim to foster access to data held by private sector entities, when these data are of public interest. From the Statistical Offices' perspective, within specific task forces, a work will continue to focus on this subject and particularly on issues related to the legislative framework and engagement with stakeholders with a view to promoting the interests of official statistics in this area.

The European Commission adopted the Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) in November 2020 and it has been submitted to the European Parliament (EP) and the Council of the EU. The proposal will be analysed, negotiated and amended by the European Parliament and the EU Council trough the competent Working Party of the Council dealing with this important dossier, mainly the Working Party on Telecommunications and Information Society (TELECOM), in order to arrive at a compromised final text. The final aim of the activities would be to create a legislative framework that ensures that Statistical Offices can use appropriate data sources in a reliable manner, including new data sources for the production of official statistics.

This Act represents the first of a series of measures announced by the Commission as part of the European data strategy. The DGA is part of the European Data Strategy (for a single data market) and complements the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the reuse of public sector information (Open Data Directive). The data subject of the DGA are not, in fact, "Open Data" as they do not fall within a specific legislation or because they are protected (by statistical or commercial confidentiality, intellectual property, GDPR).

The DGA aims at promoting data availability by strengthening trust in data intermediaries and data sharing mechanisms across the EU through:

- A mechanism to reuse certain categories of data held by the public sector (protected data, including aggregated data covered by statistical confidentiality)
- The possibility of using personal data with the help of data intermediaries, with a role of guarantee and reliability in the dynamics connected to the circulation of data
- A regulatory framework for the notification and supervision of the provision of data sharing services

A mechanism to facilitate the data altruism, related to data made available for reuse for the common good, such as for scientific research or for public services benefit The way to enhance the creation of a harmonized framework for governance on common data spaces and facilitating data sharing runs also by strengthening the trust in the intermediaries to be used in the different data spaces. In this regard, the official statistics can play key role with reference to common data spaces in strategic sectors such as public administration, health, mobility, agriculture, energy or in support of the Green Deal.

Co-financed by the Connecting Europe Facility of the European Union

Statistical Offices are recognised trustworthy data organisations with a clear legal mandate to collect and access information from multiple data sources for European statistical purposes. They have a longstanding experience in handling large amounts of data for the common good and with a particular attention to the preservation of confidentiality. They are natural partners when it comes to building trust in data and building trust between those actors who generate or hold the data and those who want to use them. Therefore, the experiences of NSO_s could be considered, to allow them to take a role as data intermediaries and where feasible other roles of the future governance framework[4].

NSO_s and Other National Authorities (ONA) responsible for the production of European statistics, which are part of the ESS, will be able to contribute as both data producers and users, but also potentially as data intermediaries. There is also a keen interest in data altruism and interoperability standards, as well as in access to data held by private individuals. In general, the ESS adheres to a vision in which data must be more widely accessible and usable within the Union and between different sectors: the benefit for official statistics will translate into a benefit for the whole society and the economy. The project is therefore to be considered of primary importance and the involvement of the NSI_s in the consultative process will be relevant. In light of the above, references to statistics should be explicitly included in the proposal on governance framework, reflecting the role played by the ESS partners.

Furthermore, it will be important also to ensure the participation of representatives of NSO_s in the governance and in discussions with stakeholders. The European Data Innovation Board is the independent body that will ensure consistency and uniformity in the application of the regulation; it would provide support and assistance to the European Commission and facilitate the standardization of European data governance, statistics included. In any case this Proposal is still subject to discussion in the EU Institutions (EU COUNCIL and European Parliament), in the co-decision procedure which has started its process during the German Presidency and will continue during the Portuguese Presidency and forward.

1.3 European Interoperability Framework principles

The interoperability principles relate to relevant behavioural aspects behind interoperability actions, providing guidance about values and concepts driving European Public Services implementation. More precisely, Public Administrations should apply the 12 EIF [18] principles during service design and development. We can divide the principles into four subsets:

- Principle setting the context for joint European interoperability actions (No 1)
- Main interoperability principles (Nos 2 to 5)
- Principles focusing on user needs and expectations (Nos 6 to 9)
- Foundation principles fostering cooperation among public authorities (Nos 10 to 12).

[4] For a discussion about data stewardship and the role of NSOs, see the UNECE document: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_25-1906075E.pdf

Co-financed by the Connecting Europe Facility of the European Union

EIF Principle 1 - Subsidiarity and Proportionality

According to subsidiarity principle, at EU level decisions are taken as closely as possible to citizens, while for the proportionality principle, EU actions cover what is required to accomplish the objectives of the Treaties.

EIF Principle 2 - Openness

In the domain of interoperable public services, openness mainly concerns data specifications and software solutions.

EIF Principle 3 - Transparency

According to this framework, transparency aims at improving citizens understanding of administrative environment, and decision-making. Furthermore, this principle underlines the relevance of interfaces to access internal IT systems, and the personal data protection, ensured by the compliance to the current regulations.

EIF Principle 4 - Reusability

The reuse of implemented solutions, as well as of available information and data, is an interoperability enhancer that allows improving efficiency.

EIF Principle 5 - Technological neutrality and data portability

Technological neutrality refers to the reduction of technological dependencies, by focusing on functional requirements, thus limiting the impact of technological changes. Data portability concerns the ability to transfer and share data among cross-border applications and systems.

EIF Principle 6 - User-centricity

The decision-making related to public services provision and delivery should consider users' needs. Therefore, the design of a new service should be driven by a multi-channel service delivery, a single contact point for all users, a regular analysis and implementation of users' feedback.

EIF Principle 7 - Inclusion and accessibility

Inclusion refers to the opportunity for everyone to access European public services, benefiting of new technologies advantages. Accessibility aims at ensuring that also disadvantaged groups can use public services like any other citizens.

EIF Principle 8 - Security and privacy

In order to interact with citizens, public administrations must put in place a secure and reliable environment that guarantees the privacy, confidentiality, and integrity of collected information, as required by the legal framework.

EIF Principle 9 - Multilingualism

Multilingualism concerns the balance between the ability of public administrations to deliver services in all official EU languages and users' expectations to be served in their own language(s) or their

Co-financed by the Connecting Europe Facility of the European Union

preferred one(s). To deal with this issue, European public services should be delivered in the languages of the potential final users.

EIF Principle 10 - Administrative simplification

Administrative simplification allows reducing administrative burden, by modernizing and streamlining administrative processes, or removing them based on the produced public value.

EIF Principle 11 - Preservation of information

This principle refers to enabling data and information accessibility as long as required, taking into account other related issues, such as reliability and integrity, security and privacy, preservation of electronic signatures or seals.

EIF Principle 12 - Assessment of Effectiveness and Efficiency

In order to ensure effectiveness and efficiency, several methods (e.g., return on investment, transparency, or users' satisfaction) allow to assess the value of European public services and choose among different available solutions.

Co-financed by the Connecting Europe Facility of the European Union

1.4 LOSD and Official statistical standards

The adoption of standards for official statistics is recommended to support statistical production management, and reuse of developed solutions and methods. The reference framework to describe and design the statistical process is the Generic Statistical Business Process Model (GSBPM) [19]. More precisely, GSBPM allows identifying the several steps of the statistical business process and the connection between them. Each step may require more than one iteration because GSBPM is a flexible model whose steps order may change according to the context. The GSBPM structure, depicted in the Figure 3, has the following three main levels:

- Level 0, the whole statistical business process
- Level 1, corresponding to the eight phases of the statistical business process
- Level 2, specifying the sub-processes performed in each phase.

Above the specific phases referring to the statistical production, the GSBPM identifies a set of overarching processes, grouping several activities performed throughout the eight phases, such as Quality Management, Data Management and Metadata Management. GSBPM documentation contains an explicit reference to Open Data in the description of the sub-processes 3.3, 4.3 and 7.2, namely in the "Reuse or build dissemination components", "Run collection" and "Produce dissemination products". Concerning data collection, the benefits of Open Data format are relevant mainly for gathering administrative, geographical, or environmental data.

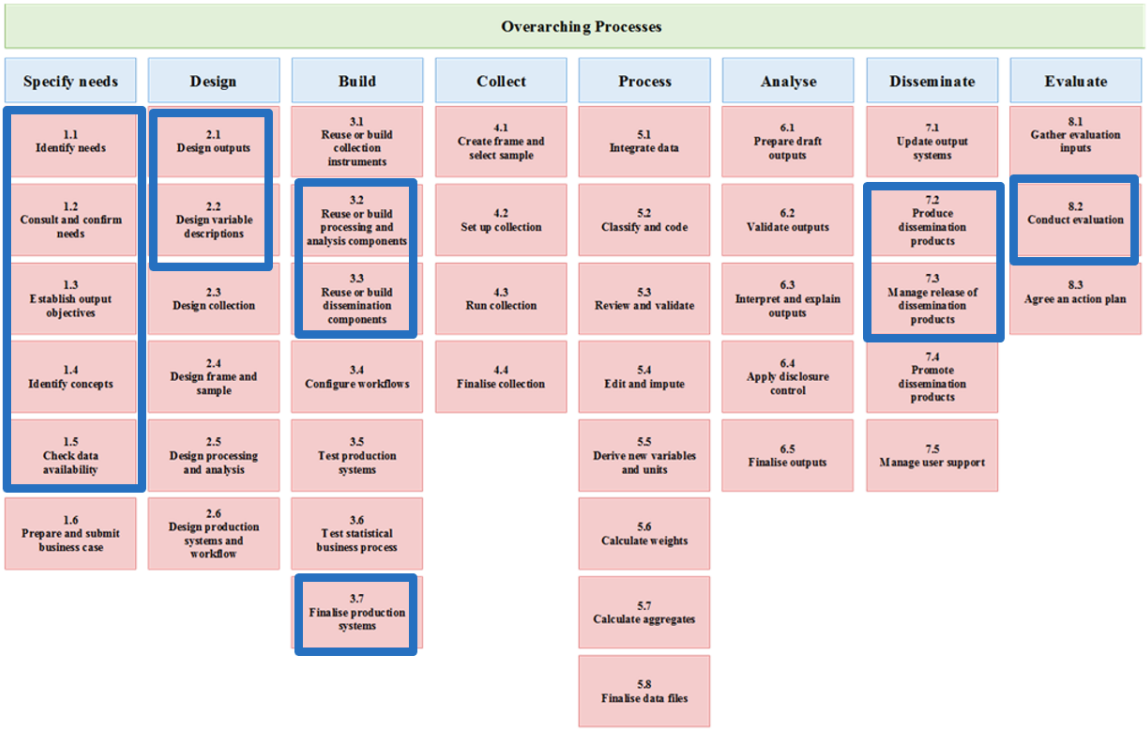


Figure 2 - Level 1 and 2 of the GSBPM[5]

[5] Source: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

Co-financed by the Connecting Europe Facility of the European Union

We will explore the highlighted GSBPM phases more in detail in Section 5.2. Although the release of official statistics relates mainly to the dissemination phase, some steps for publishing LOSD can be executed in the previous phases, enhancing process interoperability.

The Generic Statistical Information Model (GSIM) [20] provides a standard description of the information objects used as inputs and outputs in each step of the statistical process. GSIM allows standardizing data and metadata structures, enabling the reuse of existing tools and application components. GSIM includes around 130 information objects, belonging to five main groups. Each subset is described by well-defined meta-ontologies. GSIM concepts allow modelling the information background, fostering the definition of knowledge objects and their usage in the statistical process. Moreover, several international experts have cooperated to link these metamodels and the result is the association between GSBPM sub processes and phases and GSIM concepts. In order to support technical implementation, GSIM has been mapped also with the reference standard models for data and metadata interchange and description, such as Statistical Data and Metadata eXchange (SDMX) [21] and Data Documentation Initiative (DDI) [22]. Technical standards are used by LOSD mapping tools, to link the actual data and publish them executing a standardized pipeline.

LOSD pipeline has three end points for data and tools integration:

1) Metamodel Ontology describing the knowledge object of interest
2) Actual data repository where the data to be published are stored
3) Application end point for distributed information system usage.

In LOSD pipeline, the conceptual level is integrated into the physical one, instantiating the concepts described by high level modelling regardless of technical details, so that applications can use available information to run their modules.

1.5 Principles of semantic harmonization

Knowledge can be defined, represented, and used in so many ways. A common framework is needed to access the same objects of knowledge in a consistent and standard way, regardless of the underlying formalism defining it. Semantic harmonization frameworks can deal with the variability of representation of knowledge objects, and with the specifics of their semantic definitions. The following nine principles have been developed for the design of a Semantic Harmonization System dealing with medical data [23]. Nevertheless, these principles can be applied to the statistical domain also, as they address the several aspects of concepts and definitions to analyse for knowledge objects harmonization.

1. Separate technical from semantic harmonization

As knowledge can be separated in form and contents, data can be stored over a very wide range of formats and devices. On one hand, technical harmonization allows to access data regardless of the specific implemented tools, so that it can be transformed into a single format, or commonly accessed by a single query system, called endpoint. On the other hand, semantic harmonization deals with a common conceptual representation of the knowledge objects.

Co-financed by the Connecting Europe Facility of the European Union

2. Distribute ownership of local and global knowledge objects

Knowledge about data domain is different from the knowledge in the field of research. Experts in both fields can share knowledge but domain experts define data semantics, while researchers define process and application semantics. Actually, variables used for data analysis are the global representations of knowledge.

3. Separate vocabulary from structure

Data integration is achieved using common vocabularies or taxonomies. There are also specifics about naming conventions and standards for values and measures. As suggested for technical and semantic harmonization, specifying a vocabulary to reference knowledge definitions is different from building the entire semantic structure of knowledge objects. More precisely, knowledge representation should be based on general structures and linked to reference vocabularies for domain terms and dependencies. Consequently, the knowledge representation structure is not affected by reference vocabularies changes and enables the introduction of new vocabularies.

4. Standard Vocabularies Reuse

Ontology can be extended and inherited whenever possible. So, there is no need to duplicate existing knowledge objects and their relations. It is recommended to reuse or reference existing ontologies, vocabularies, and taxonomies whenever possible to maintain semantic congruence and vocabulary stability. The task of extending or creating a new vocabulary should be separated from that of semantic harmonization.

5. Use declarative mappings

Domain experts should describe the relationship between different knowledge objects by declarative rules, established according to the following criteria:

- Mapping rules should be stated in a formal executable language to avoid ambiguities
- Mapping rules must be simple to be understood and receive feedback from other domain experts.

6. Code isolation

Each rule for knowledge specification should have a limited scope, referring to a single concept, and should be managed as a stand-alone object, so that objects changes do not interfere each other. This limits the scope of the objects, but also minimizes the influence between objects and the impact of changing code, with added efficiency and ease of maintenance.

7. Enable integrated security and provenance

A knowledge representation system should keep track of any access to local knowledge objects. Provenance and security issues are particularly relevant, as such kind of systems are more likely to access several data sources with different ownership and to be accessed by other systems out of own control. Data access may be limited in time, or have other restrictions, depending on the type of data users, variables, observed units, granted authorisations. In order to simplify data access management, in addition to users and variables classification, knowledge objects should allow identifying for every derived value the related data source that should be deducible also from the mapping rules.

Co-financed by the Connecting Europe Facility of the European Union

8. Separate WHAT from HOW

Data mapping to a common global knowledge object related to different data sources requires an assessment of the coherence between variables. More precisely, the evaluation of variables compatibility relates to protocol compatibility and involves both, the measured phenomena (what), and the method applied to measure it (how). As an example, when linking resident population from several data sources, some data may not be linked directly, as the calculation method of each data cube may be different for every source, and some data may need recalculation or adjustments. In order to make all the data congruent to a single standard, the knowledge object representation should have some rules to perform or explain the process execution.

9. Balance generic versus specific descriptions

Knowledge systems can be very generic or very detailed, and this has both drawbacks and benefits. A generic system can meet multiple requirements but can lose in detailed specifics. A very detailed system may bring forth duplication, redundancy, and interoperability problems. Consequently, it is necessary to balance both aspects, coping with general knowledge objects for capturing expected types of data and metadata, and customized data structures for storing specific domain information.

Co-financed by the Connecting Europe Facility of the European Union

2 Relevant initiatives and maturity assessment

2.1 The European Data Portal

In order to foster access and reuse of public sector information, the European Commission funded the European Data Portal (EDP) since 2015, within the Connecting Europe Facility (CEF) programme. The Portal is a repository that harvests available metadata on public data and from geospatial portals, covering all EU Member States, EFTA countries, and other countries involved in the EU's neighbourhood policy. The European Data Portal [24] is a single access point for public data, published at national or regional level, or related to a specific domain. Cross-border data comparability is achieved by using a common format for metadata descriptions, more precisely the specifications provided by the DCAT Application Profile for data portals in Europe (DCAT-AP) [25].

The quality of metadata from data providers and data portals is assessed by The Metadata Quality Assessment (MQA) [26] and published using W3C Data Quality Vocabulary (DQV) [27]. MQA is a tool that provides a set of quality indicators and some rating metrics, according to the following dimensions: Findability, Accessibility, Interoperability, and Reusability, derived from the FAIR Guiding Principles for scientific data management and stewardship [28].

The Portal can be accessed by a human-readable website [29], or by a machine-readable API, enabling the metadata search, creation, change or deletion. In relation to data access and use, the following sections have been implemented for:

- Searching datasets, through harvested metadata from different countries grouped in several categories, according to the revision of the DCAT Application Profile. These subsets have also been mapped against the Eurovoc, the EU's multilingual thesaurus.
- Providing Data, for an understanding of Open Data from the provider perspective. Furthermore, this section contains also some instructions to allow the EDP to harvest data from a portal.
- Using Data, reporting some details about Open Data use and its economic benefits.

In addition, the Portal offers several types of supporting resources, such as:

- 16 eLearning modules, available in six languages, describing Open Data basic concepts (e.g.: licensing, data formats, guidance to benefit from open data)
- A collection of use cases providing several examples on Open Data reuse in different domains
- An active presence on social media
- An Open Data news feed to promote Open Data initiatives and updates
- A page for the dissemination of worldwide Open Data events
- Fortnightly Data Talks, jointly with the Support Centre for Data Sharing (SCDS)
- A Country Insights page, to share at a national level news, datasets, use cases, events, and blog posts

Co-financed by the Connecting Europe Facility of the European Union

- Several reports about the effect of Open Data publication and reuse, and 18 analytical reports examining current open data issues
- An annual survey to assess the Open Data Maturity of the Member States.

At the moment, the Portal provides translations of metadata descriptions in all EU official languages, and access to over one million datasets from 36 countries, organised in 81 catalogues. The infrastructure implemented and its functionalities supports the development of data applications and products, enhancing the content of several Open Data portals, including the European Union's Open Data Portal [31].

2.1.1 The European Data Portal maturity measurement

Among the several initiatives to promote Open data access and reuse, every year the European Data Portal carries out to assess the maturity level of Open data in Europe. The results of this analysis are presented in yearly reports and shown in the open data maturity dashboard[6], a tool that allows visualizing and comparing open data maturity level of European countries.

The main steps of the landscaping exercise are:

- Step 1: Methodology and questionnaire updates
- Step 2: Coordination and support to national teams to fill out the questionnaire
- Step3: Data analysis and validation in cooperation with the national teams
- Step 4: Additional desk research to complement the results
- Step 5: Results and findings dissemination, through an in-depth report and country factsheets
- Step 6: Visualisation of the results on the European Data Portal dashboard
- Step 7: Release of an analytical report and webinars planning, to display the top 3 countries with the highest maturity level.

The dimensions considered for the maturity assessment are:

- **Open Data Policy** refers to the adoption of ad-hoc policies and strategies to increase Open Data dissemination at national level. Moreover, this dimension points out the existence of a governance framework allowing the involvement of private and third sector stakeholders, and the implementation actions fostering Open Data initiatives at national, regional, and local level.
- **Open Data Portal**. This dimension refers to advanced portal features that allow expert and less expert users to access Open Data, as well as portal functions that enable publishers and reusers interaction.

Furthermore, the dimension measures the use of web analytics tools to gain insights on users' needs and behavior, and then update the portals functionalities. In addition, the dimension assesses the Open Data coverage, according to the domain and the strategy adopted to ensure the portal sustainability.

[6] https://www.europeandataportal.eu/en/impact-studies/open-data-maturity#2018

Co-financed by the Connecting Europe Facility of the European Union

- **Open Data Impact** focuses on the activities in place for monitoring and measuring Open Data reuse, as well as the assessment of the impact resulting from the reuse. In addition to the strategic awareness, this dimension relates to four main impact areas: political, social, environmental, and economic. For each area, the questionnaire aims at detecting the extent of monitoring activities, the development of applications, products, and services to face challenges in these sectors, the existence of civil society initiatives related to Open Data and supported by public authorities. Concerning the economic area, an additional analysis relates to assessments, reports and/or studies for measuring the value of Open Data, at both micro and macro-economic level, and the efficiency increase gained by the public sector.
- **Open Data Quality** highlights the measures developed to harvest metadata from different sources at national level. Additionally, the fourth dimension assesses data and metadata currency, the compliance with the DCAT-AP metadata standard, and the quality of disseminated Open Data. This dimension also offers suggestions to portal managers and policymakers for improving the quality of published Open Data.
- For each dimension, a set of indicators allows to assign a score, thus obtaining comparable results over time and among different countries [31]. According to the overall maturity score, the respondent countries are classified by total maturity level, and divided in the following clusters, starting from the most mature to the least:
- **Trend-setters**. This group includes countries with structured Open Data policies and coordinated Open Data initiatives throughout the public sector. The national portal offers several functionalities, fulfilling the needs of skilled users and publishers. The quality of published Open Data is high and assessed by several initiatives, monitoring also the compliance with DCAT-AP and the impact of Open Data reuse among different domains. In general, there are no particular limitations to Open Data publication and reuse.
- **Fast-trackers**. This subset groups all countries with a good maturity level registered in all dimensions. In general, these countries have adopted a strategic approach to achieve standard compliance and improve Open Data quality. The national portal meets the needs of expert and basic users. Although the efforts to monitor Open Data impact are limited, several activities are in place to track and enhance reuse, as well as to deal with some issues remaining.
- **Followers**. This cluster is composed by those countries having an Open Data policy and a sufficient degree of coordination among Open Data activities. The national portal provides standard features and a limited number of functionalities for experienced users**.** A limited number of activities are carried out to increase the dissemination of high-quality data provided by different sources. Other limitations concern the reuse monitoring, and the assessment of Open Data impact.
- **Beginners**. The countries belonging to this group have an initial level of maturity, considering the four dimensions. Despite some progresses related to the Open Data policy dimension, an Open Data portal has not been implemented, or has limited functionalities and datasets. Considering Open Data reuse and the assessment of its impact, the performed activities are limited, or nonexistent, as well as the actions to publish Open data in higher quality or adopt the DCAT-AP standard.

The following figure shows the distribution of the four clusters in Europe in 2020[7]. More in detail: 7 countries are in the Trend-Setters group: Denmark, Spain, France, Ireland, Estonia, Poland, and Austria, while 13 countries are in the Fast-Trackers group. The Followers group includes 7 countries, and the residual 8 countries belong to the Beginners group.

[7] Source: Open Data Maturity Report 2020, available from: https://www.europeandataportal.eu/it/dashboard/2020#country-overview

Co-financed by the Connecting Europe Facility of the European Union

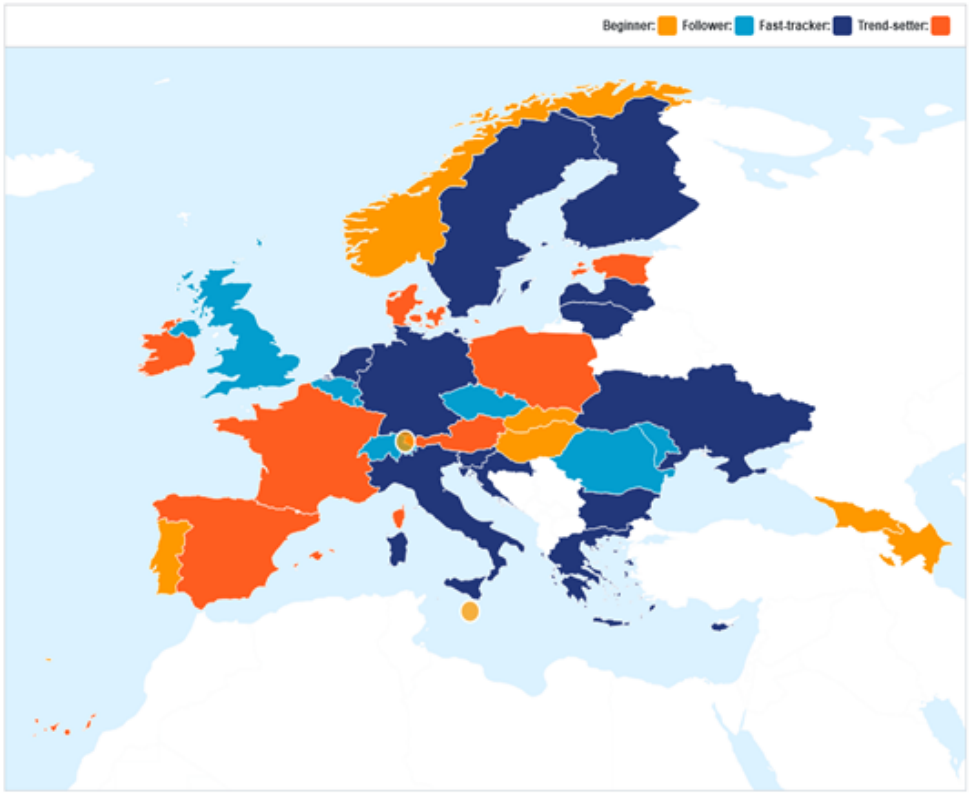


Figure 3 - European Countries maturity map in 2020[8]

2.2 Digicom Project

General context

The ESS Vision 2020 was adopted in May 2014, to develop and implement a joint strategy to help all member countries to modernize statistical processes and their output, thus facing the new challenges for official statistics. Among the several objectives, the following is an excerpt from the ESS Vision 2020 [32] that aims at:

- Engaging users proactively and meets their demands in a cost-efficient and responsive manner
- Promoting efficiency and realizing productivity gains through collaboration in sharing methods, tools, technological infrastructure
- Embracing opportunities provided by the digital transformation
- Delivering information in an interactive and easily understandable way
- Improving statistical literacy of European citizens and institutions by guiding them through the deluge of data and information from various origins.

Main objectives

[8] Source: European Data Portal Dashboard

Co-financed by the Connecting Europe Facility of the European Union

As part of the ESS Vision 2020 portfolio, the Digital communication, User analytics and Innovative products (DIGICOM) project was launched to accomplish most of the above areas. This goal has been achieved by testing and developing innovative products and services for data dissemination, starting from the ESS experiences and the requirements of European statistics users. The establishment of collaborative networks is another key feature of the project, thus building key capabilities within the ESS. The project has been structured following a dual approach, to preserve agile management and quickly adjust according to users' needs. Therefore, the main drivers of the project have been the improvement of methods, to identify unfulfilled needs of current and potential users, and the development of innovative solutions to meet those needs.

Key outcomes

DIGICOM project started in 2016 and lasted for three years, producing 55 deliverables [33]. The scheduled activities were divided into the following subsets, described in terms of achieved goals and results:

1) User Analysis

- Achieve a broader understanding of users of European statistics by applying different methods for user profiling, such as an in-depth user analysis starting from available sources, ad-hoc field studies, several usability tests related to statistical products disseminated online by Eurostat and the NSIs involved in the project.
- Identify the ever-changing needs of users by creating specific communication channels, more precisely: an ESS social network for expert users, and an ESS Facebook page for the potential ones.

In addition to the above objectives, outcomes also include several guidelines about user analytics, user research and user personas, a set of user personas tailored on actual users for guiding the design of new products to disseminate.

2) Innovative and shareable products and tools

The focus of this task was the development of shareable and customizable products with a standardized content, such as short texts, images, animations, and interactive tools. Furthermore, this subset of activities was conceived to enhance the visual dissemination of European statistics, by exchanging best practices, scheduling regular trainings of ESS staff, providing common guidelines, and sharing visualization tools. Another relevant outcome of this task is the European digital publication "The life of women and men in Europe" which was translated into all EU official languages and published simultaneously throughout the ESS and in the main EFTA countries. This initiative was very successful, having a great impact on social media and mainstream media.

3) Open Data Dissemination

This strand analyzed new potential dissemination channels, building enabling capabilities to support Open Data dissemination throughout the ESS. This task was accomplished by several actions, starting with an assessment of several data dissemination options, to identify the most effective ones and was structured in five main threads [34], namely:

- Linked Open Data, focused on stocktaking and requirements gathering
- Facilitation and harmonisation of APIs to European data, in order to enhance extraction tools, to harmonize and standardize APIs and improve data access at European level

Co-financed by the Connecting Europe Facility of the European Union

- Development of a European reuse policy [35], starting from the Common ESS conditions for access to and reuse of data
- Improved access to microdata, for stocktaking and developing improved solutions for microdata access
- Search and extraction tools, including the revision of best practices for the semantic search of statistical concepts, the requirements for search and extraction tools, and the development of reusable solutions for end-user search.

The main outcome of the first thread was the development of Public Use Files (PUFs), containing anonymized micro data, for the Labour Force Survey (LFS) and the European Union Statistics on Income and Living Conditions (EU-SILC), available for more than 20 NSIs. The focus of most activities was the addition of a semantic layer to data, in order to make them LOD. To this aim, the following actions have been particularly significant:

- Extension of the stocktaking exercise to include also an initial revision of relevant ontologies for statistical data and metadata
- Provision of a set of recommendations towards Linked Open Data
- Design of a reference architecture for LOD, shared with the ESS Enterprise Architecture Board
- Synchronization between the Open Data Portal and Eurostat data catalogue
- Analysis of the existing ESS Linked Open Data portals and inventory of LOD use cases
- Development of a semantic format converter service.

The development activities were performed in collaboration with the ESSnet on Linked Open Statistics and resulting in the Linked Open Statistical Data hub [36], to allow all NSIs to convert, upload and link their data. This outcome was complemented by a proof of concept developed by Eurostat to test the use of pseudo knowledge graphs for data search and discovery. This prototype [37], developed in a test environment, on a set of predetermined use cases and data sources, is freely available online. In general, the achieved outcomes have underlined the relevance of semantic technologies and standards for data dissemination, highlighting the need of building LOD capabilities for the whole ESS.

4) Communication and promotion

These activities dealt with the main issues related to the communication and promotion strategy of official statistics. In addition, at the beginning, the planned activities included a platform for experimental statistics and statistical literacy, as well as gamification. Regarding statistical literacy, the outcomes achieved exceeded the initial expectations, arranging several tools and event such as:

- The European Statistics Competition
- Statistical games[9] developed by ESS NSIs and Eurostat
- An online e-learning tool and an e-learning portal
- Educational videos for teachers, students, and for open online course
- Translation of statistics4beginners.

Furthermore, an ambitious program to engage multipliers (teachers, researchers, data journalists, NGOs) was planned, based on the “knowledge disseminator” scenario, resulting from the DIGICOM “branding study”, a study for the analysis of users’ perception and the improvement of official statistics communication.

[9] https://ec.europa.eu/eurostat/help/education-corner/games

Co-financed by the Connecting Europe Facility of the European Union

2.3 ESSnet LOS

General context

After an initial stocktaking and requirement gathering, most of the development work within the Digicom Project was performed in partnership with the ESSnet on Linked Open Statistics [38], for the implementation of the Linked Open Statistical Data hub.

In order to harmonize data dissemination, one of the main goals of the Eurostat "Linked Open Statistics" ESSnet was the development of a common framework for Data Sharing through LOD.

The ESSnet started with the following four national Institutes participating in a pilot project, aimed to design a common reference framework:

- National Statistical Institute of Bulgaria
- Institut National de la Statistique et des Etudes Economiques (INSEE)
- Istituto Nazionale Di Statistica (ISTAT)
- Central Statistics Office (Ireland).

The interest in the project was so high, that many other institutions and NSO_s were interested in the results. By the end of the ESSnet, in July 2019 four more NSO_s expressed the will to participate actively in any following activity about LOSD. Every single NSO had its own experience with Open Data dissemination and a different level of maturity, starting from none to having built a custom framework. This scenario highlighted the need to integrate several different preexistent frameworks into a common standardized LOSD pipeline.

Main objectives

The ESSnet goals, as well as the DIGICOM project, were related to all key areas of the ESS Vision 2020, and particularly to Key Area 1: "Focus on Users" and Key Area 5: "Improve Dissemination and Communication". Aiming at enhancing LOSD value for users and stakeholders, developing open and reusable solutions, and promoting the adoption of existing standards, the overall objectives of the ESSnet were:

- To explore NSIs publishing statistical data as LOSD on the semantic web
- To allow users to engage with statistical data easily as Linked Open Data
- To prepare NSIs to publish LOSD in the wider ESS
- To recommend a way forward for the ESS to adopt and exploit LOSD.

Each objective was associated to a business case [39], converted in practical use-cases, to validate the initial assumptions in the application domains. In order to achieve these objectives, part of the work was subcontracted to external stakeholders: Derilinx, the ADAPT Centre and Insight Centre for Data Analytics, responsible for the provision of:

- A service for publishing and linking statistical data on the Semantic Web (Linked Open Statistical Data publication project)
- A set of services, tools and training materials for supporting Linked Open Data dissemination, use and visualization, thus creating capacity building within ESS network (LOSD Capacity Building and Training Project).

Co-financed by the Connecting Europe Facility of the European Union

Following a workshop held in Malta in 2017 and adopting a practical approach to accomplish the above objectives, the first ESSnet deliverable was the design of a generalized pipeline for LOSD publication and dissemination. This pipeline, shown in Figure 5, was the starting point for the development of a reusable service for LOSD.

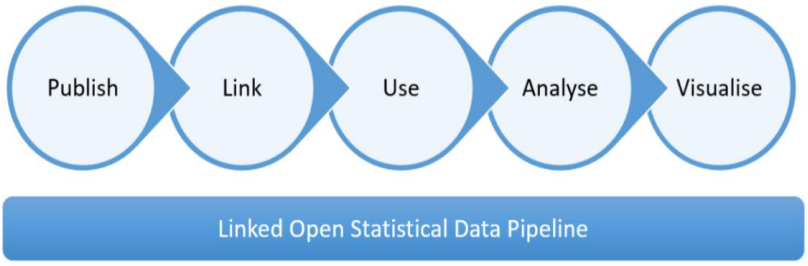


Figure 4 - LOSD pipeline[10]

For each element of the pipeline, a preliminary analysis was carried out to design a LOSD reference architecture, and a LOSD publication architecture [40], starting from an inventory of available resources for the statistical data management. In addition, for the description of the service functionalities, the following three practical use-cases [41], involving several types of potential stakeholders, were examined:

- Use Case 1: NSO Statisticians Publishing LOSD (related to the objective: Explore NSIs publishing statistical data as LOSD on the semantic web)
- Use Case 2: General Public Users Accessing LOSD Data (related to the objective: Allow users to engage with statistical data easily as Linked Open Data)
- Use Case 3: NSO Collaboration with third Party Partner for Developing Added-Value Services.

The ESSnet was structured in the following technical activities:

- LOSD Pilots, User Assessment and Recommendations on Horizontal Topics
- ESS Networking, Cooperation and Capacity Building.

The main goal of these activities was to set up a platform for linked data, to develop pilot dissemination of LOSD based on different use cases, and to produce a set of recommendations related to common aspects. Another achieved result was the setup of a platform [42] enabling the collaborative work among LOD experts involved in the ESSnet, and the definition of a strategy for capacity building, promoted by several events and training material.

Key outcomes

Among the main results listed in the ESSnet webpage, the most relevant outcomes for this context are:

[10] Source: D1.1 Vision, Stakeholders and Business Case Definition, available from: https://ec.europa.eu/eurostat/cros/system/files/d1.1_vision_stakeholders_and_business_case_v3.pdf

Co-financed by the Connecting Europe Facility of the European Union

- Linked Open Statistical Data hub[11] and related software components. More precisely, the subcontractors, allowing the access to datasets, Linked Data tools and training materials, implemented a test portal. The portal is composed of the following modules [43]:
  - Web portal for data loading and management (CKAN)
  - WYSIWYG interface for data mapping capabilities (JUMA)
  - Data navigation tools
  - OLAP browser for Hypercubes visualization
  - Table visualizers
  - Triplestore storage (Virtuoso)
  - Endpoint for data querying.

  These modules allow data semantics standardization, and external source data linking, and provide a set of capabilities for:
  1. Open Data publication by several open standards (CSV, RDF, Turtle).
  2. Data integration by LOD format
  3. Data access point for data query using appropriate query language (SPARQL)
  4. Dataset management and particularly: Data analysis, Data visualization, Data Report

- The Catalogue of Linked statistical datasets and triplestore population implemented for the use cases
- The following datasets published, by the ESSnet participants, in Derilinx's portal:
  1. Labour forces
  2. Income and living conditions
  3. Census 2011
  4. Unemployment rate
  5. Tourism Data

- A Catalogue for the description of semantic resources
- Several documents describing LOSD pilots and their assessment from a user perspective, as well as the recommendations on horizontal topics and on software and infrastructure solutions
- Platform for collaborative work
- Stakeholders' workshop and two Webinars: "Introduction to Statistical LOD", "Statistical LOD from a user perspective".

[11] https://losd-data.staging.derilinx.com/

Co-financed by the Connecting Europe Facility of the European Union

2.4 High Value Datasets

In order to foster Open Data dissemination and reuse, Section V of the Open Data Directive (ODD) has entrusted the European Commission to adopt, as an implementing Act, a specific list of High-Value Datasets (HVD_S), as well as requirements and arrangements for their access and reuse. Due to the relevance of the information provided in terms of economic and social benefits, these datasets have a high commercial potential. In addition, they foster the development of value-added information products at European level and provide key data sources for Artificial Intelligence development.

As mentioned in Section 2.2.3, the ODD provides the following thematic categories of HVD_S:

- Geospatial
- Earth observation and environment
- Meteorological
- Statistics
- Companies and company ownership
- Mobility

Depending on technological and market enhancements, the Commission may add further categories to the above list. In compliance with the Directive, each Member state of the European Economic Area (EAA) will provide HVDs, according to the following requirements:

- Free of charge access
- Machine readable
- Provision via Application Programme Interfaces (APIs)
- Bulk downloadable

The list of specific HVD_S to be adopted by the Commission in 2021 is the result of an impact assessment, carried out in cooperation with a Committee on Open Data and the reuse of public sector information. Due to the inclusion of Statistics in the list of thematic categories, Eurostat has been involved in the preliminary studies for the impact assessment, namely for the definition of an ESS list of HVDs in the statistics category. Considering ESS capabilities and the valued added for the users, in April 2020 Eurostat proposed an initial list [44] of HVDs related to the following statistical domains:

- Business statistics
- Macroeconomic statistics
- Social statistics

Each dataset has been described in terms of:

- Main variables
- Sustainable Development Goal (SDG) indicators
- Frequency
- Granularity/breakdowns (depending on the sample size)

Co-financed by the Connecting Europe Facility of the European Union

The release of the definitive list of HVDs by category, through the adoption of the implementing Act, is expected in the first half of 2021. The specification of statistical HVDs has underlined several crosscutting issues, such as:

- Metadata description for each HVD, to be addressed by defining a conceptual model [45] for statistical HVDs
- Fulfilment of the availability requirements (free availability, machine-readable format through APIs, bulk download), related to the interoperability and accessibility dimensions. Interoperability level may affect the combination between:
  - HVDs concerning different domains, or different categories, or produced by other countries
  - HVDs and other official statistics
  - HVDs and other types of datasets, not included in the list of HVDs and external to the statistical domain.

In relation to accessibility, although the compliance with the first two requirements (free of charge access and machine-readable format) is achieved, further joint actions and developments are needed to improve availability via APIs and bulk download. The definition of a common conceptual view of the datasets to be exchanged and reuse is a prerequisite for interoperability. The model is based on the assumption that statistical HVDs can be considered as 'slices' of greater general statistical data cubes and can be described in terms of measures and dimensions. Modelling the general statistical cubes before the definition of statistics HVDs will help to generalize concepts and descriptions, thus facilitating their governance and the revision of HVDs over time. The conceptual model, to be detailed together with Member States, allows overcoming technological gaps among different countries.

A proposed strategy for metadata representation and enhancing semantic and technical interoperability is based on the combination of the conceptual model for data and metadata descriptions, and SDMX standard or Linked Open Data standards. More in detail, after the definition of the conceptual template, Eurostat would use SDMX to model statistical HVDs, and store them in a central Registry. These artefacts can be modeled and published as Linked Open Data (mainly, using SKOS and XKOS ontologies). On the other side, Member States may use APIs to provide the requested information, modelled according to the common conceptual model and the established format.

2.5 ISA² - Interoperability solutions for public administrations, businesses and citizens

General context

The ISA² (ISA squared) [46] programme was conceived to foster the reduction of administrative burdens and the digital communication between European public administrations, businesses, and citizens. Starting from the results of the previous Interoperability Solutions for European Public Administrations (ISA) programme and extending its scope, ISA² supports the implementation of digital solutions to increase cross-border and cross-sector interoperability of public services. More precisely, ISA² is the new programme, running from January 2016 until December 2020, in order to ensure the coordination of the activities performed at EU level, and develop solutions and tools based on:

- A reviewed European Interoperability Framework (EIF)

Co-financed by the Connecting Europe Facility of the European Union

- A reviewed European Interoperability Strategy (EIS)
- The European Interoperability Reference Architecture (EIRA)
- A repository of interoperability solutions, the European Interoperability Cartography[12] (EIC).

ISA[2] is part of a broader framework of initiatives for enhancing interoperability at European level, such as:

- The roadmap related to "A digital single market strategy for Europe", the Commission's Communication emphasizing the role of interoperability and standards, and particularly the Priority ICT Standards Plan, and the EU Catalogue of ICT standards
- The European eGovernment Action Plan 2016-2020, aiming at removing digital barriers, reducing administrative burdens and facilitate interactions with public institutions
- The IT Governance of the European Commission
- A set of initiatives fostering the innovation of the public sector and setting up a cross-border interoperability infrastructure, such as Horizon 2020 and the Connecting Europe Facility (CEF).

Main objectives

ISA[2] was adopted [47] in November 2015 by the European Parliament and the Council of European Union, to run from January 2016 until December 2020, to achieve the following objectives:

- The development, enhancement and reuse of available cross-border and cross-sector interoperability solutions
- Assessment of ICT effects of EU regulations and identification of fields that could benefit from new legislation fostering interoperability
- Design of a European Interoperability Reference Architecture (EIRA) for the development and assessment of interoperability solutions
- Set up of a tool to promote the reuse of implemented interoperability solutions and identify areas in which such solutions are not available
- Assessment and promotion of existing common specifications and standards, and development of new ones
- Development of a system for the assessment of the benefits resulting from interoperability solutions.

The programme is put in place through rolling working programmes, revised every year and selecting specific actions according to a prioritization criterion. These actions are grouped in the following 9 tasks, related to similar initiatives:

- **Key and generic interoperability enablers**, for the development of generic key interoperability solutions to support public services provision
- The **semantic interoperability** task, grouping the initiatives launched to improve semantic interoperability among public organisations
- The access the **data/data-sharing/Open-Data** task supporting the conversion of national data repositories in the Open Data format, fostering cross-borders and cross-sectors reuse, and facilitating the access to public data

[12] https://joinup.ec.europa.eu/collection/cartography/solution/european-interoperability-cartography-eic

Co-financed by the Connecting Europe Facility of the European Union

- The **geospatial solutions** task groups the initiatives related to geospatial data produced by public and private providers
- The **e-procurement/e-invoicing** task includes all Commission e-procurement initiatives
- The **decision-making and legislation** task refers to all initiatives that support the decision-making process for improving interoperability between EU institutions and the Member States
- The EU **policies - supporting instruments** for the implementation of EU policies
- The **supporting instruments for public administrations** include all actions for the development of solutions that support public administrations in implementing interoperable services
- The **accompanying measures** task concerns the initiatives for monitoring of actions and for promoting the interoperable solutions funded under ISA².

As a whole, the ISA² programme supports 54 actions, organized in the above subsets and addressing several aspects of interoperable solutions development. A dashboard[13] has been implemented to guarantee the coherence between long-term goals and the performance of the work programme, and monitor the state of play of underway actions.

Key outcomes

In synergy with LOD initiatives in the statistical domain, the third task (access the data/data-sharing/Open-Data) has contributed to the implementation of ESS Vision 2020, by funding the following actions:

- 2016.06 [48] - Sharing statistical production and dissemination services and solutions in the European Statistical System. The project, started in 2016 until 2019, achieved the following results:
  - Revision of the reference architecture for the statistical production
  - Implementation of the ESS Service Catalogue fostering the reuse of shared statistical services
  - Development of new reusable services and solutions, also for the statistical dissemination
  - Benchmarking of Member States architecture for the following domains: Data validation, European system of Business Registers, Data, and metadata management service, Linked open data.

- 2018.03 [49] - Improving statistical data and metadata discoverability and analysis. The main outcomes of this action, scheduled from 2018 to the end of 2020, are:
  - Design of reference ontologies for harmonising statistical data and metadata
  - Development and test of a reference architecture for metadata handling, as well as end users data services
  - Relevant metadata assets (classifications, code lists) released as semantic objects.

  Other results expected by the end of 2020 are:
  - Generalized models to represent the main statistical metadata as Linked Open Data
  - A reference architecture for managing statistical metadata assets in selected use cases, such as accessing data, integrating and analysing data
  - Relevant European Statistical metadata transformed in open format
  - A prototype of a data analytics platform for accessing and analysing Eurostat Open Data.

These actions, partially funded by Eurostat, have contributed to the following ESS initiatives for implementing Vision 2020. Particularly, the first action is related to:

[13] https://ec.europa.eu/isa2/dashboard/

Co-financed by the Connecting Europe Facility of the European Union

- ESS Enterprise Architecture (ESS EA) – aiming at modelling a common reference architecture for NSIs and Eurostat, to harmonise statistical processes and reducing the gap between business needs and ITdevelopments
- "Shared SERVices" (SERV) – a project for sharing technical statistical services (including dissemination) and supporting their integration in the statistical processes across the ESS
- DIGICOM Project, involved in the second action also, along with the ESSnet LOD.

2.6 International initiatives at a glance

Every year an increasing number of governments and international organisations launches Open Data initiatives to foster free use, reuse and distribution of data for any purpose and without restrictions. While a large amount of data is published on government websites, the majority of published data is intended only to be read as stand-alone documents, not reused for other purposes. To be considered "open," the data must be re-usable, meaning they can be downloaded in open formats and read by software, and users have a legal right to reuse it.

Within the international Initiatives, it is interesting to have a look at the Open Data Inventory (ODIN), developed by the Open Data Watch (ODW) [50]. It is an international, non-profit organization of data experts working to bring change to organizations producing official statistical data. The main mission is to support the efforts of NSOs, particularly those in low- and middle-income, to improve their data systems and harness the advancements of the data revolution and enhancing capacity in Open Data. The ODW Openness Assessment has five elements, namely:

1) Machine readability
2) Use of non-proprietary formats
3) Availability of multiple download options
4) Availability of metadata providing sufficient context to understand the data
5) Open licensing.

ODW's assessment methodology is available to NSOs or other statistical agencies for self-assessment. The Open Data Inventory (ODIN) is a platform that measures the completeness of a country's statistical offerings, and whether data meet international standards of openness. Data assessed in ODIN must be official country data published on the NSO website, or any other official country website linked from the NSO website. The ODIN assesses the coverage and openness of official statistics to identify gaps, promote Open Data policies, improve access, and encourage dialogue between NSOs and data users.

The availability of the Report of ODIN[14] provides an overview of the situation of countries in Open Data availability and the evidence of leading countries in the field. The 2020/21 Open Data Inventory (ODIN) is the fifth edition of the index compiled by Open Data Watch. It covers 187 countries. The platform allows exploring and downloading the results from 2020 or previous years, and visualizing on the map more information on a specific country. Coverage scores are based on the availability of key indicators, disaggregated over time and for geographic subdivisions. Openness scores are based on the following indicators:

- Data can be downloaded in machine-readable and non-proprietary formats

[14] https://odin.opendatawatch.com/report/pressReport

Co-financed by the Connecting Europe Facility of the European Union

- Data are accompanied by metadata
- Download options exist such as bulk download, user-selection, or APIs
- Data released under Open Data license.

Concerning the international scenario, the Collaborative on SDG Data Interoperability[15] convened by the UN Statistics Division and Global Partnership for Sustainable Development Data (GPSDD) launched Data Interoperability: A practitioner's guide to joining up data in the development sector at the 2018 World Data Forum, held in Dubai, UAE in October 2018. The Guide identifies five dimensions of interoperability that are required for the development of data systems and processes capable of integrating data from numerous sources, including data published in Open Data-friendly formats.

Finally, the World Bank established the Open Data Initiative, developing the Open Government Data Toolkit[16] getting started with Open Data. It is a toolkit designed to help governments, Bank staff and users understand the basic principles and resources of Open Data, then get "up to speed" in planning and implementing an open government data program, while avoiding common pitfalls.

[15] https://www.data4sdgs.org/initiatives/data-interoperability-collaborative
[16] http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html

Co-financed by the Connecting Europe Facility of the European Union

3 Towards a joint LOSD strategy

3.1 Key dimensions of LOSD maturity assessment

One of the main outcomes of the Digicom project is the study of current ESS experiences in LOSD dissemination, to assess the maturity of statistical linked data, and provide insights for a common strategy. This study, performed by Price Waterhouse Coopers [51] to analyse LOD strategies in ESS countries, was carried out by desk research and interviews with NSI's representatives. More in detail, the analysis of the state of play in some ESS countries (including France, Ireland, United Kingdom and Switzerland) has revealed three main scenarios, reflecting different maturity levels of data linking. The following scenarios summarize the LOD strategy at national level, providing a glimpse of the achieved maturity:

- Individual local activities to use linked format for interconnecting official statistics datasets (considering both data and metadata) stored in different repositories within an NSI, thus increasing data integration efficiency
- Joint activities between NSI_s and Eurostat to adopt the linked format for interconnecting official statistics datasets (including both data and metadata) stored in the repositories of different NSIs and/or Eurostat. These initiatives, allow to overcome interoperability barriers, and to reduce the exchange of physical data within the ESS
- LOD for publishing official statistics. The main goal of these activities is to disseminate official statistics according to linkable, machine-readable formats, at a granularity level that fulfils users' needs. In this scenario, the adoption of open standards (e.g. RDF Data Cube, SDMX) supports the reuse of official statistics, enhancing cross-sector interoperability.

During the study, the investigation of ESS strategies was structured following the nine areas in the Business Model Canvas [52]. Among these areas, an overview of key resources and key activities implemented by NSI_s to create value for statistical users allows to identify relevant aspects of LOD strategies to consider for further improvements. Concerning key resources, the interviewed NSI_s have pointed out the following needs:

- Clear data licensing and reuse conditions
- URI policy to guarantee HTTP URIs consistency, resolvability, persistence
- Linked data infrastructure to provide data as LOD. The main components of this infrastructure should be:
  - A data preparation building block
  - A SPARQL endpoint for LOD publishing
  - Some REST API for data access
  - Data browsers for data exploration
- Skills and capacities for LOD publishing and management
- Data standards to support data structures modelling and semantic description.

Co-financed by the Connecting Europe Facility of the European Union

In relation to key actions for LOD implementation and management, during the interviews, NSIs's experts have underlined the following issues:

- Scoping and requirements analysis to specify business needs, technical requirements and to assess and monitor development activities
- Development and grouping the several activities to carry out for LOD dissemination, such as data modelling, data transformation and harmonization
- Management and Maintenance to support LOD publishing over time
- End-user perspective and promotion to facilitate LOD use. Among these actions, some experts have enhanced the relevance of providing visualization tools for data extraction and investigation.

In addition to the inventory of LOD strategies of several ESS countries, the study has provided a representation of the maturity level based on the 5 stars model developed by Tim Berners-Lee [53]. According to this model, the following dimensions have been considered:

1) Maturity level in terms of:

- Open data dissemination on the web
- Data dissemination in a machine-readable format
- Data dissemination in an open machine-readable format
- Publishing of some datasets in a linkable format
- Publishing of a relevant number of datasets as linked data

2) LOD availability, in terms of published resources (e.g., code lists, datasets, metadata).
Although the depicted state of play may be updated according to NSIs feedback and enhancements, the following figure shows that most part of countries are in the central levels of maturity, so there is large room for improvements at European level.

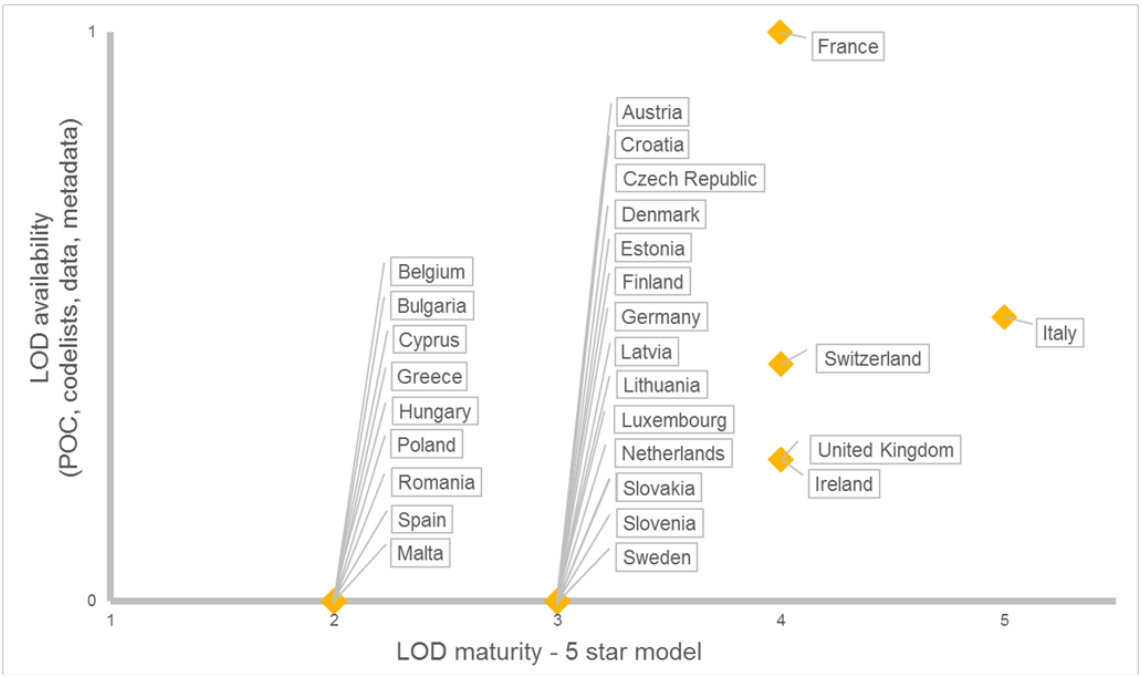


Figure 5 - ESS countries LOD maturity level (Source: Price Waterhouse Cooper, Study on LOD strategies and use cases in the ESS)

Co-financed by the Connecting Europe Facility of the European Union

3.2 A common framework for LOSD actions: drivers and stakeholders

The strategy proposed for improving interoperability involves several stakeholders that can be divided in three main subsets, according to their prevailing role in LOSD development and reuse: LOSD providers, Reference bodies, Research, and LOSD final users. The first group includes the following entities:

- **European Statistical System**

The ESS, built to provide consistent and comparable statistics for Community policies, is a network of several organisations responsible for the development, production and dissemination of official European statistics. The partnership involves the statistical office of the European Union (Eurostat), NSI_s and other national or European agencies producing statistics for relevant business sectors (e.g., energy, environment, and telecommunications). The ESS has an active role in the set-up of legal and technical frameworks and infrastructure enabling all Member States to publish LOSD. In fostering harmonisation and capacity building, the ESS cooperates with several Commission services and international organisations, such as the UN Statistics Division, the UNECE and OECD.

- **National governments**

National governments play a relevant role in fostering the reuse of public sector information by implementing at local level the current European regulation. Although they are also statistical data users, the coordination with statistical offices is essential to extend the benefits of the Smart Government paradigm introduced by the European Commission, to the development of LOSD capabilities.

- **Non-governmental organisations (NGO_s)**

NGO_s support the definition, the achievement and the assessment of sustainable development goals and are both, data producers and statistics consumers. Based on this assumption, they should be engaged in LOSD initiatives for their expertise related to particular aspects, such as privacy protection or the assessment of LOSD relevance in specific domains.

- **Private companies data holders**

The private sector includes several businesses that play a key role in the statistical scenario, both as data holders and data users for their operational goals. Fostering the cooperation with companies, providing innovative services and implemented software solutions, allows to increase interoperability, thus reducing the cost of LOSD technical development. The establishment of public-private partnerships fosters the creation of a win-win strategy where public and private sectors complement each other to achieve shared goals. This premise should reduce competition and avoid market distortions.

European Commission's decisions, and Data Protection Authorities monitoring affect the activities of the institutions involved in LOSD development and dissemination.

- **European Commission**

The policy initiatives of the European Commission to promote and regulate the adoption of open format have resulted in the latest Open Data Directive and the creation of Common European data spaces to foster data access and the implementation of new products and data services. The European Commission activity relates also to the private data market, as well as the reuse of public information for statistical purposes.

Co-financed by the Connecting Europe Facility of the European Union

- **Data Protection Authorities**

In performing their activities, LOSD providers must follow GDPR principles throughout the whole implementation process. Such authorities supervise the application of the data protection regulation, assessing also the data protection issues arising from the development of new methods for data dissemination and reuse.

The second subset of stakeholders groups several public authorities, supporting technical and semantic standardisation, as well as the scientific community, and includes:

- **European and international organisations**

In addition to the ESS, several external organisations, such as the World Bank, the UNECE, the UN Statistics Division or OECD are actively involved in LOSD development. As an example, within the United Nations Statistics Division's (UNSD) current work, the LinkedSDGs[17] platform supports the 2030 Agenda implementation. This initiative fosters the use of common data standards for monitoring the achievement of sustainable development goals. In order to avoid duplication and establish a structured collaboration, the integration between the ESS work and similar initiatives is achieved through partnership in project groups and task forces, as well as participation in high-level management groups.

- **Standardisation bodies**

The use of technical standards and communication protocols is essential to implement and reuse LOSD solutions, and gain efficiency in fulfilling interoperability requirements. Standardisation bodies provide guidance and frameworks to improve semantic and technical harmonization, facilitating the dissemination of official statistics as LOSD.

- **Scientific community and Academia**

The research and academic communities are relevant users of official statistics, and at the same time key partners in LOSD technical implementation. The exchange of Know-how and expertise with the ESS, achieved through collaborative projects and staff trainings, promotes the development of problem-driven solutions. Another advantage of joint research activities is the contribution from disciplines that are usually outside common statistical partnerships, such as Computer science. This approach allows to achieve greater results in terms of skills and capacity building.

The third group is composed by several types of data consumers that could benefit from the release of official statistics in LOSD format, such as:

- **Citizens, Private enterprises, Policy makers**

Official statistics support the decision making in relevant economic and social domains. The combination of several statistical data, coming from recognized trusted sources, provides a greater understanding of a subject, enabling policy design and management. From a company perspective, the integration between official statistics and private market data assists the definition of a business strategy, taking into account internal and external factors. The development of Open Data portals enables to widen the range of citizens that access to statistical data to analyse a specific topic. In addition, the increase of reusable statistical information could result also in a reduction of the respondent burden in the most investigated domains. The adoption of LOSD format underlines the need to meet users' requirements to facilitate data sharing and integration. These requirements concern not only the harmonisation of data structures models and formats, but also the description of data and metadata meaning, according to the different sources combined. Therefore, an interactive

[17] For more information, see: https://sustainabledevelopment.un.org/LinkedSDGs/about

Co-financed by the Connecting Europe Facility of the European Union

dialogue with LOSD end users, as well as a regular investigation of their needs are essential to promote the value of official statistics.

- **Media**

Official statistics, due to their trustfulness and easy access without charge, are widely used by the media to describe a particular phenomenon and encourage public debate. At the same time, having a great impact on public opinion, the media could be involved by NSI_s in the assessment and promotion of interoperability solutions for improving data access, navigation and analysis. Moreover, the engagement of media is a key factor for the success of a communication strategy underlining LOSD benefits. Particularly, outside of specialized media, the mains messages to citizens are to convey the importance of secure data sharing and data integration to reduce burden, as well as the power of data linkages to increase interoperability.

For each group of stakeholders, the following figure shows the common drivers, and the related assessments, goals and outcomes, represented according to ArchiMate language[18]. The ArchiMate objects used to model these concepts are described in Annex A.

[18] ArchiMate is an open and independent language for modeling domain architectures according to the Enterprise Architecture standard, available from: https://www.archimatetool.com/

Co-financed by the Connecting Europe Facility of the European Union

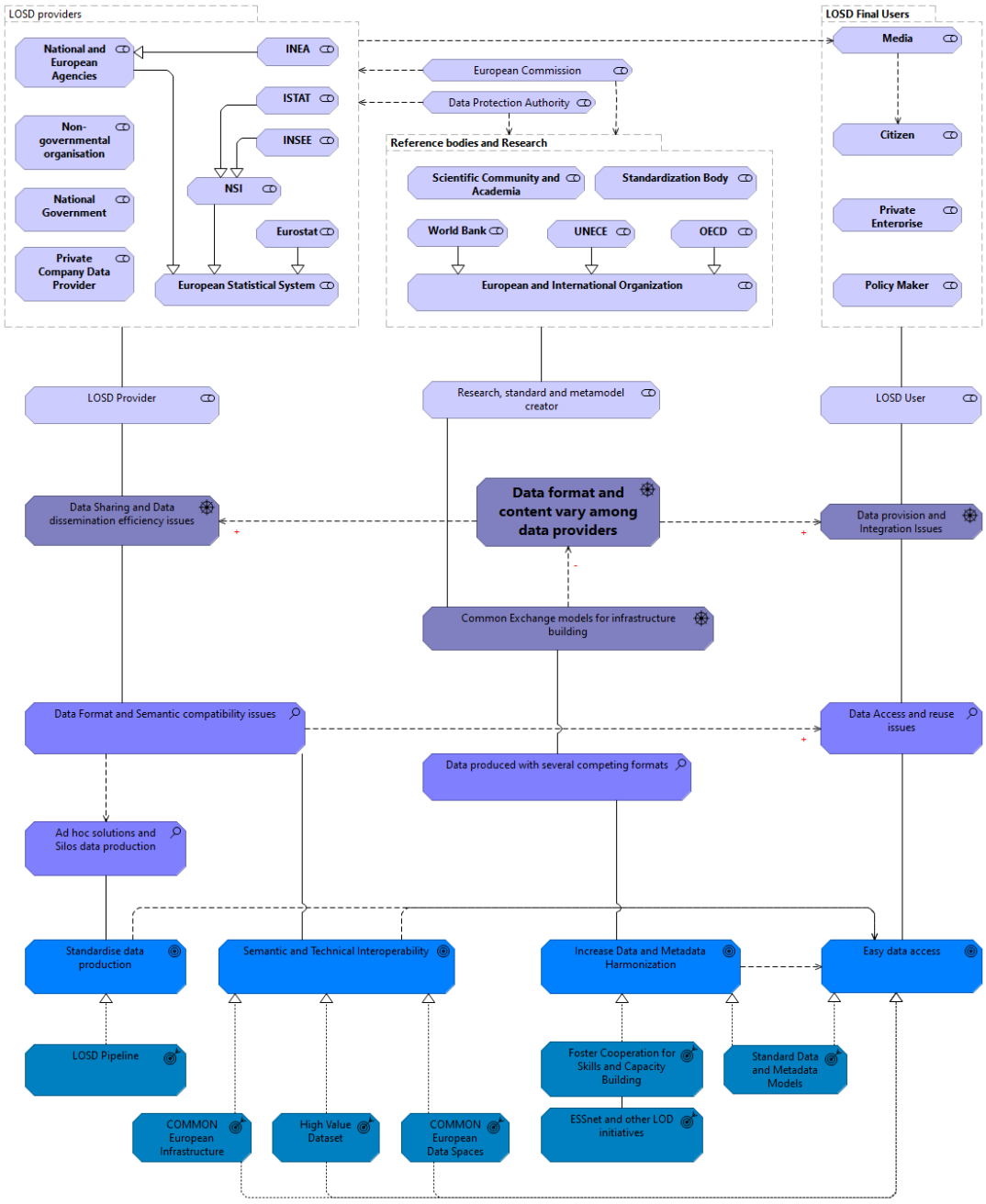


Figure 6 - LOSD stakeholders, drivers, assessments, goals and outcomes

The main driver, affecting the activity of all stakeholders involved in LOSD production and use, is the heterogeneity of data format and content among data providers. Particularly, for LOSD publishers, this interoperability barrier results in data sharing and data dissemination inefficiencies. On the other hand, LOSD users have to face several challenges related to data provision and integration. Regardless of technical aspects, another driver of the second subset of stakeholders (Research, standard and metamodel creators) is the development of common exchange models for statistical data, to build a shared infrastructure for LOSD dissemination.

Co-financed by the Connecting Europe Facility of the European Union

In order to identify suitable corrective actions, all stakeholders are involved in the assessment of the AS-IS context. Particularly, lack of semantic interoperability pushes LOSD providers to develop ad hoc solutions and silos data production, which, in turn, produces further interoperability issues. The analysis performed by LOSD providers concerns also data access and reuse challenges that involve LOSD final users as well. The assessment from the perspective of Research, standard and metamodel creators underlines the existence of several competing data formats.

Starting from the analysis of the current situation, each group of stakeholders sets specific goals, to address the identified issues and achieve the desired outcomes. More in detail, at national and European level, the main goals of LOSD providers are data production standardisation, and the improvement of semantic and technical interoperability to facilitate data access. In relation to these goals, the expected outcomes are the design of a LOSD pipeline, the implementation of a common European infrastructure, as well as the provision of High Value Datasets and the creation of Common European data spaces for relevant domains. The last three outcomes, related to semantic and technical interoperability, have a positive impact also on data access and reuse. The main goal of Research, standard and metamodel creators is the enhancement of data and metadata harmonisation. The outcomes related to this goal are an increase of cooperation for skills and capacity building through the organisation of ESSnets and other LOD initiatives, and a greater standardisation of data and metadata models. The last outcome has a positive impact on the easy data access goal. The main result of this analysis is an overview of the connections between the described stakeholders and the effects of their achievements that have a positive impact on the system.

3.3 A joint strategy for enhancing interoperability

The analysis of LOSD strategies in the ESS, carried out by desk research and interviews with NSI's representatives, as well as the Malta workshop held in January 2017, have led to the definition of a joint ESS LOD strategy [54], introduced in Section 2.1. The proposed approach is based on the combination of the following five dimensions:

- Strategy and Policy
- People and Capacities
- Data and Metadata
- Governance
- Technology and Infrastructure.

The objective of the first dimension is the creation of a common vision and a shared implementation strategy, based on iterative cycles with specific outcomes, to foster the assessment of LOSD benefits before investing a larger amount of resources. The second dimension aims at laying the foundation for increasing internal knowledge, to achieve the required skills for providing LOSD. The third dimension aims at prioritizing the types of statistical data and metadata that should be disseminated as LOSD, as well as data standards and formats to use, according to best practices. The governance dimension relates to the coordination of LOSD ownership and to the management of LOSD life cycle, including some processes, such as harmonising, publishing, updating, disseminating, and archiving. The establishment of governance structures, both at national, European level and international level is essential for building LOSD capabilities. The dimension related to Technology and Infrastructure will be analysed more in detail in deliverable D.1.2, dedicated to the technical aspects.

Co-financed by the Connecting Europe Facility of the European Union

From an organizational perspective, the implementation of the proposed strategy is based on both joint activities carried out by NSI_s in partnership with Eurostat, and national activities individually performed by NSI_s. This approach has several benefits and particularly:

- Enhanced semantic harmonization of official statistics across the ESS
- Improvement of cross-border and cross-domain interoperability by building common infrastructures, skills and capability
- Development of common innovative solutions to meet users' requirements
- Compliance with the ESS Enterprise Architecture reference framework[19] that fosters the development of common services available to all Member States, acknowledging the independence of NSI_s to set up their own LOSD infrastructures and engage with national stakeholders.

Among the future actions related to each dimension for implementing the proposed strategy, the followings are particularly relevant, in terms of NSI_s involvement and commitment:

- Cooperation with other organisations beyond the ESS to increase awareness of LOSD benefits
- Acquisition of external funding by participating to the programmes launched by the European Commission
- Analysis of new business models to introduce innovations in the dissemination process
- In-house development of Key LOSD skills and capabilities for an efficient management of available resources
- Active participation in standardisation initiatives
- Improvement of NSI_s data portals to facilitate data search and exchange
- Exploration of open format for improving internal steps of the statistical process, such as data integration
- Additional assessment of the impact of LOSD format on confidentiality requirements
- Enhancement of provenance metadata with a quality assessment of published LOSD.

3.4 Lessons learnt from previous experiences

The initiatives described in the previous chapter have highlighted the benefits of creating networks and sharing knowledge and tools, to increase interoperability across sectors and ESS countries. An assessment of the several initiatives launched has underlined the need of adopting a wider approach covering the described dimensions. In addition, iterative implementation activities and the specification of expected outcomes are recommended before increasing the amount of resources to invest.

An overview of the main insights gained from the latest projects is provided by the analysis of joint activities to be performed with high priority. Particularly, from a wider perspective the lessons learnt may correspond to the rationale behind the primary activities proposed for each dimension within the common ESS LOD strategy. Concerning the first dimension, Strategy and Policy, the knowledge gained in the previous experiences has suggested the following actions:

[19] https://ec.europa.eu/eurostat/cros/content/ess-enterprise-architecture-reference-framework_en

Co-financed by the Connecting Europe Facility of the European Union

- Build proofs-of-concepts, in order to foster collaboration, implement capabilities, and share knowledge
- Align LOSD national strategy with the common strategy coordinated by Eurostat
- Commitment of highest management levels to support LOSD national strategy
- Identify target stakeholders to engage them and adopt ad-hoc communication strategies
- Promote the development of LOSD standards interacting with other stakeholders outside the ESS.

In relation to People and Capacities, the cultural change needed to increase LOSD dissemination is based on the following insights:

- Foster cooperation and knowledge sharing by participating to the initiatives launched to address common issues, such as joint pilots, workshops
- Plan development activities to reduce knowledge gaps arisen during LOSD implementation
- Build internal LOSD expertise by integrating different skills and experiences to harmonize concepts, methods, technical procedures, and business goals
- Create effective partnership with academic community and private companies providing skills and capabilities for LOSD dissemination and reuse.

Regarding Data and Metadata dimension, the assessment of the state of play has highlighted the need to:

- Provide linked statistical metadata starting from statistical classifications, code lists and concepts, to facilitate data understanding
- Adopt common and flexible standards, to be specialized for statistical domain
- Identify LOSD to publish according to priorities criteria, taking into account requested data and the benefits of data reuse
- Analyze LOSD dissemination from an architectural perspective to assess the impact of LOSD format on statistical process phases.

Referring to the Governance dimension for managing the several steps of LOSD pipeline, the previous experiences have suggested the following primary joint actions:

- Develop a shared ESS governance system, integrated with the implemented framework
- Increase the cooperation with standardization authorities and the participation in relevant open projects
- Harmonize data and metadata models to increase interoperability across domains and countries
- Alignment of European and national policies to improve URI_s persistence
- Define a set of recommendations to provide guidance for LOSD dissemination and describe the steps to execute and the related standards and tools to implement.

Co-financed by the Connecting Europe Facility of the European Union

4 Guidelines to overcome legal and procedural barriers

4.1 Overview of legal and procedural barriers

The analysis of the current ESS scenarios has underlined several legal and procedural barriers that prevent LOSD development and dissemination. Although the current legal framework provides a basis for Open Data publishing and reuse, there is the need of further efforts to align national and European legislation. In addition, as resulted in the analysis of Open Data barriers, performed by the European data portal [55], the harmonisation between the Open Data regulation and the national and regional Open Data policies is essential to prevent and remove potential legal issues. Other challenges described in Section 2.2.4 are:

- Consistency between Open Data licenses and reuse policies
- Analysis of privacy issues that may arise from the combination of personal and Open Data
- Assessment of potential confidentiality constraints related to Open Data format.

In order to prevent these legal issues, the Open Data Gold-book for Data Managers and Data Holders [56] suggests adopting an Open Data policy to support the implementation of the Open Data strategy within an organization. In addition to Open Data definition and benefits, scope and goals of Open Data initiative, legal aspects and specification of a list of data to publish in open format, the policy should consider the following tasks:

- Evaluate the current and future situation for the definition of measurable goals
- Identify, schedule and monitor the actions needed to achieve the planned objectives
- Assess technical and budgetary requirements
- Set up a reference point of contact for any internal question about Open Data
- Measure the results achieved by a set of Key Performance Indicators.

The procedural barriers depend on both external factors, related to the reference scenario, and internal organization of NSI_s' activities. More precisely, the analysis of users' needs, the compliance with technical standards and reference frameworks, as well as data publishing on existing data portals may foster the release of new products, or the development of software solutions, thus affecting the statistical process. From the internal perspective, in most cases, the low integration between LOSD pipeline and the production chain prevents the set-up of capabilities and reusable tools. Particularly, the initiatives described in the previous chapter have highlighted the following procedural challenges, related to interoperability enhancement:

- Definition of an internal LOSD strategy from an architectural perspective that considers the whole statistical process, and not only the dissemination phase
- Ontology development for harmonizing concepts and definitions
- Implementation of ad-hoc solutions for specific tasks instead of generalized tools
- Data and metadata standardization
- Lack of skills and resources.

An architectural analysis of the production chain is essential to identify the main actions to cope with internal procedural barriers.

Co-financed by the Connecting Europe Facility of the European Union

4.2 Turning challenges into opportunities

In order to face the challenges listed above and turn them into opportunities, the following recommendations provide guidance from an architectural perspective, focusing on in-house improvements. The overview of legal barriers has highlighted the need of harmonising national regulations, Open Data licenses and reuse conditions. In order to achieve this result, the following recommendations may foster the alignment with the European LOSD regulation:

- Analyze legal aspects related to LOSD pipeline for each type of data, to identify the tasks to perform to ensure the compliance with the European legal framework and the European reuse policy
- Set up a protocol to check the consistency between Open Data license and dataset terms of use
- Develop Proof-of-Concepts to assess the impact of privacy constraints on LOSD access and reuse
- Provide a set of guidelines to specify data transformations that are compliant with the legal requirements of data reuse.

The focus of the following analysis is to highlight the relationship between the LOSD pipeline and the statistical process, to identify and remove potential procedural barriers to LOSD dissemination. This approach is consistent with the "Open By Default" principle stated in the Open Data Charter[20]. In other words, interoperability enhancement may concern several stages of the statistical process, rather than the dissemination phase only. Therefore, the following analysis focuses on the statistical process, described in detail using ArchiMate language. Before even attempting any implementation, process design allows abstracting from the semantic context, and provides a conceptual description of the tasks to perform and the data structures involved.

The following figure shows GSBPM phases that include one or more steps of the LOSD pipeline, and the statistical roles involved in each stage. Starting from the first GSBPM phase, in order to specify statistical needs, domain experts and ontology designers cooperate to:

- Identify the subset of official statistics to publish as LOSD, taking into account stakeholders' needs
- Specify the selection of data to extract for LOSD dissemination
- Define core data properties, such as: concepts, attributes, measures, dimensions, and roles
- Analyse LOSD data sources, such as Registers and administrative sources.

In GSBPM Design phase, ontology designers model the semantic layer, based on the core concepts previously defined, and in particular:

- Design ontology for data and metadata modelling, starting from available meta-ontologies (e.g., Data Cube Vocabulary, SKOS, XKOS), or reusing concepts from existing ontologies with equal definitions
- Design or reuse statistical classifications.

[20] https://opendatacharter.net/principles/

Co-financed by the Connecting Europe Facility of the European Union

The outcome of this task is a specific domain ontology used to generate RDF triples, together with LOSD data source. More in detail, during GSBPM Build phase, the mapping designers specify the conversion rules to transform data from a general application format (e.g., csv, txt, database tables) into RDF. This stage refers to the development from scratch, or the reuse of software solutions for:

- Converting data (micro/macro) in RDF format
- Providing web services for LOSD dissemination
- Producing technical documentation and user manuals.

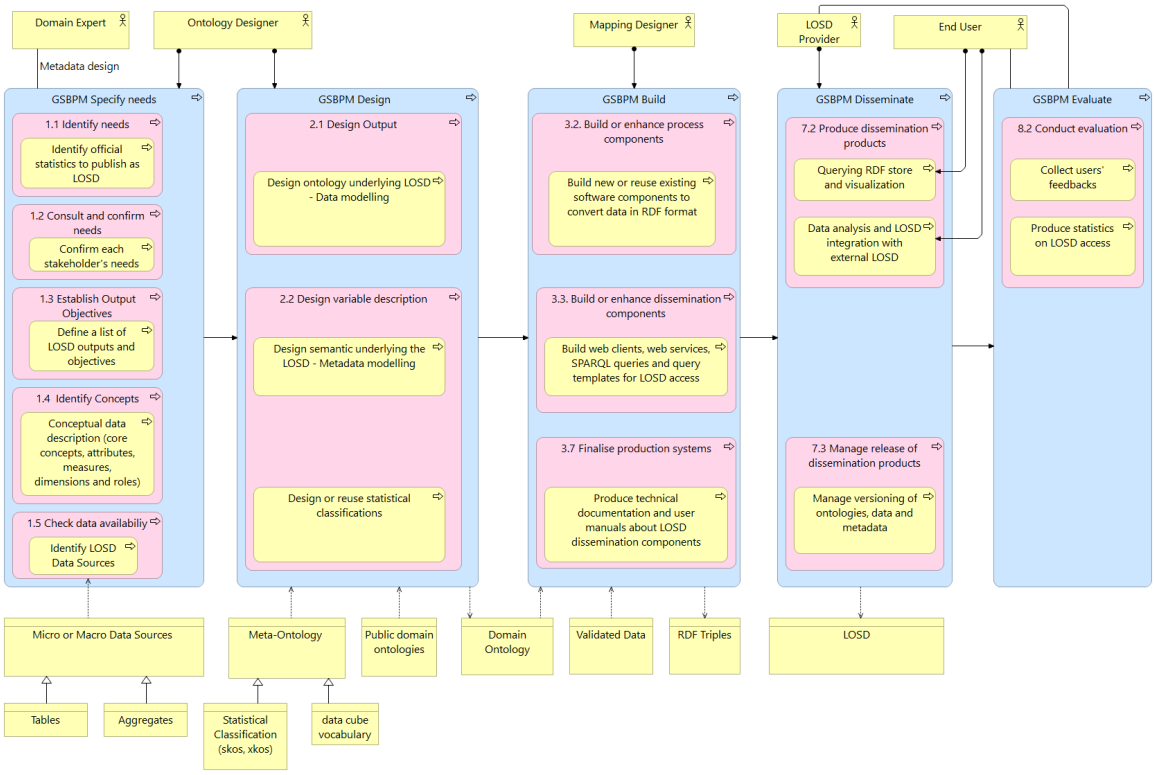


Figure 7 - GSBPM phases and LOSD pipeline

GSBPM Disseminate phase involves several stakeholders, divided in two main groups: data providers and final users. While the former publishes LOSD on the web, managing and monitoring several issues related to disseminated products, such as data and metadata versioning, security and privacy, the latter access LOSD for data analysis and visualization. The final stage of the statistical process concerns the assessment of results, to identify and prioritise potential enhancements. In GSBPM Evaluate phase, data providers may analyse LOSD access data or collect users' feedbacks to improve LOSD pipeline and increase official statistics disseminated as LOSD.

The following recommendations may foster the removal of procedural barriers:

- Set up of a LOSD pipeline starting from HVDs specific domains and Common European data spaces sectors, to join the European LOSD framework
- Deal with semantic harmonization in the GSBPM Design phase, and manage potential enhancement during the Dissemination phase
- Describe LOSD pipeline in terms of process metadata, to track data sources, data transformations, methods applied to calculate derived variables and/or dimensions

Co-financed by the Connecting Europe Facility of the European Union

- Model LOSD provenance to complement metadata provision
- Standardize metadata content to facilitate LOSD querying through natural language.

The suggested approach grounds on iterative improvements, to fill the gap between the current situation and the intended outcomes. The definition of a set of Key Performance Indicators is essential to monitor the achieved results and define a long-term strategy to enhance LOSD dissemination.

Co-financed by the Connecting Europe Facility of the European Union

References

[1] Linked Data Principles, available from: https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html#linked-data-principles

[2] Fundamental Principles of Official Statistics, available from: https://www.unece.org/stats/fps.html

[3] European Statistics Code of Practice, available from: https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7

[4] Common Quality Framework, available from: https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf

[5] Regulation (EC) no 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and amended by Regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 2015

[6] Generic Law on Official Statistics, available from: http://www.unece.org/DAM/stats/documents/ece/ces/2016/mtg/ECE_CES_2016_8-1601555E.pdf

[7] Guidance on common elements of statistical legislation, available from: legislation http://www.unece.org/DAM/stats/documents/ece/ces/2018/CES_6_Common_elements_of_statistical_legislation__Guidance__for_consultation_for_upload.pdf

[8] Charter of Fundamental Rights of the European Union, available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN

[9] European data strategy, available from: https://ec.europa.eu/digital-single-market/en/european-strategy-data

[10] Communication "Towards a common European data space", available from: https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space

[11] Directive (EU) 2019/1024 on open data and the reuse of public sector information, available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN

[12] Legal aspects of public sector information (LAPSI) thematic network outputs, available from: https://ec.europa.eu/digital-single-market/en/news/legal-aspects-public-sector-information-lapsi-thematic-network-outputs

[13] Dulong de Rosnay M., Janssen K., Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. Journal of Theoretical and Applied Electronic Commerce Research, VOL 9 / ISSUE 3 / SEPTEMBER 2014 / 1 -14

[14] European Data Portal: Protecting data and opening data, available from: https://www.europeandataportal.eu/it/highlights/protecting-data-and-opening-data

[15] 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)

[16] Analytical Report 3: Open Data and Privacy, available from: https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf

[17] 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, available from:

Co-financed by the Connecting Europe Facility of the European Union

https://eurlex.europa.eu/legalcontent/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066
[18]Underlying principles of European public services, available from: https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/2-underlying-principles-european-public-services
[19]Generic Statistical Business Process Model (GSBPM), version 5.1, available from: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1
[20]Generic Statistical Information Model (GSIM), version 1.2, available from: https://statswiki.unece.org/display/gsim/Generic+Statistical+Information+Model
[21]SDMX - Statistical Data and Metadata eXchange, available from: https://sdmx.org/
[22]Data Documentation Initiative (DDI), available from: https://ddialliance.org/
[23]Cunningham J. A., Van Speybroeck M., Kalra D., Verbeeck R., Nine Principles of Semantic Harmonization. AMIA Annual Synposium Proceedings Archives, February 2017, available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333211/
[24]Understanding the European Data Portal High level presentation of the architecture, available from: https://www.europeandataportal.eu/sites/default/files/edp_factsheet_portal_architecture_online.pdf
[25]DCAT Application Profile for data portals in Europe, available from: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/about
[26]European Data Portal Metadata Quality Assurance, available from: https://www.europeandataportal.eu/mqa/?locale=en
[27]W3C Data Quality Vocabulary, available from: https://www.w3.org/TR/vocab-dqv/
[28]FAIR Guiding Principles for scientific data management and stewardship, available from: https://www.go-fair.org/fair-principles/
[29]European Data Portal website: https://www.europeandataportal.eu/en
[30]EU Open Data Portal website: https://data.europa.eu/euodp/en/home
[31]Open Data Maturity Report 2020, available from: https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf
[32]ESS VISION 2020 Building the future of European statistics, available from: https://ec.europa.eu/eurostat/documents/7330775/7339647/ESS+vision+2020+brochure/4baffcaa-9469-4372-b1ea-40784ca1db62
[33]DIGICOM project results, available from: https://ec.europa.eu/eurostat/cros/content/digicom-results_en
[34]ESS.VIP DIGICOM Business Case. Version 1.0.0, October 2015, available from: https://ec.europa.eu/eurostat/documents/7330775/7339647/DIGICOM+Business+Case+1.0/8a2f1b42-4101-4208-9c49-1d1819604047
[35]ESS policy on free access to and re-use of data, available from: https://ec.europa.eu/eurostat/web/european-statistical-system/reuse-ess-statistics
[36]Linked Open Statistical Data hub, available from: https://losd-data.staging.derilinx.com/
[37]Eurostat Linked Open Data prototype (LOD2019), available from: https://ec.europa.eu/eurostat/cros/DIGICOM/LOD2019_en
[38]ESSnet on Linked Open Statistics, available from: https://ec.europa.eu/eurostat/cros/content/essnet-linked-open-statistics_en
[39]LOSD Publication and Capacity Building. D1.1 Vision, Stakeholders and Business Case Definition, available from:

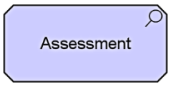
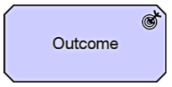
Co-financed by the Connecting Europe Facility of the European Union

https://ec.europa.eu/eurostat/cros/system/files/d1.1_vision_stakeholders_and_business_case_v3.pdf
[40]LOSD Publication and Capacity Building. D2.1 LOSD Publication Pipeline and Overall LOSD Service Architecture, available from: https://ec.europa.eu/eurostat/cros/system/files/d2.1_p_losd_publication_pipeline_and_overall_losd_service_architecture.pdf
[41]LOSD Publication and Capacity Building. D1.2 CSO LOSD Publishing Use cases, available from: https://ec.europa.eu/eurostat/cros/system/files/d1.2_cso_use_cases_v2.pdf
[42]Work package 2ESS Networking, Cooperation and Capacity Building, available from: https://ec.europa.eu/eurostat/cros/system/files/kickoff_-_wp2_presentation.pdf
[43]LOSD Publication and Capacity Building. D4.1 Complete Documentation of the LOSD Service, available from: https://ec.europa.eu/eurostat/cros/system/files/d4.1_complete_documentation_of_the_losd_service_v2_1.pdf
[44]DIME/ITDG document: High-value datasets (HVDs) - Detailed dataset specifications, available from: https://ec.europa.eu/eurostat/cros/content/item-2-high-value-datasets_en
[45]DIME/ITDG document: High-value datasets (HVDs) Opportunities for improved interoperability and accessibility, available from: https://ec.europa.eu/eurostat/cros/system/files/2a._dime-itdg-2020-2a-hvdsintacc.pdf
[46]ISA² - Interoperability solutions for public administrations, businesses and citizens, available from: https://ec.europa.eu/isa2/isa2_en
[47]Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector, available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1449478436536&uri=CELEX:32015D2240
[48]2016.06 Sharing statistical production and dissemination services and solutions in the European Statistical System, available from: https://ec.europa.eu/isa2/actions/sharing-statistical-services-and-solutions_en
[49]2018.03 Improving statistical data and metadata discoverability and analysis, available from: https://ec.europa.eu/isa2/actions/enhancing-discoverability-and-analysis-statistical-data-and-metadata_en
[50]Open Data Inventory (ODIN), available from: https://odin.opendatawatch.com/
[51]Price Waterhouse Cooper, Study on LOD strategies and use cases in the ESS, available from: https://ec.europa.eu/eurostat/cros/content/task-2-use-cases-and-analysis_en
[52]Osterwalder, A., & Pigneur, Y., Business Model Generation (2009)
[53]5 ★ Open Data, available from: https://5stardata.info/en/
[54]Item 2- Annex 1: A joint framework for action on linked open data at ESS level, available from: https://ec.europa.eu/eurostat/cros/content/item-2annex1proposallodessframework_en
[55]Analytical Report 5: Barriers in working with Open Data, available from: https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf
[56]Open Data Goldbook for Data Managers and Data Holders, available from: https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

Co-financed by the Connecting Europe Facility of the European Union

ANNEX A

Table A.1: ArchiMate motivation objects used to model LOSD drivers and stakeholders

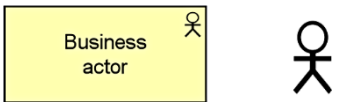
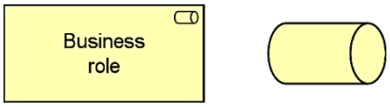
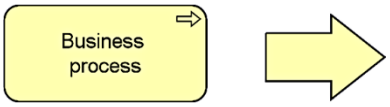
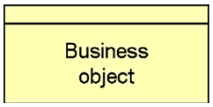
Element	Description	Notation
Stakeholder	A stakeholder represents the role of an individual, team, or organization (or classes thereof) that represents their interests in the effects of the architecture.	
Driver	A driver represents an external or internal condition that motivates an organization to define its goals and implement the changes necessary to achieve them	
Assessment	An assessment represents the result of an analysis of the state of affairs of the enterprise with respect to some driver	
Goal	A goal represents a high-level statement of intent, direction, or desired end state for an organization and its stakeholders	
Outcome	An outcome represents an end result	

Source: ArchiMate® 3.1 Specification[21]

[21] https://pubs.opengroup.org/architecture/archimate3-doc/

Co-financed by the Connecting Europe Facility of the European Union

Table A.2: ArchiMate business layer objects used to model LOSD pipeline and the statistical process

Element	Description	Notation
Business actor	Represents a business entity that is capable of performing behavior.	
Business role	Represents the responsibility for performing specific behavior, to which an actor can be assigned, or the part an actor plays in a particular action or event.	
Business process	Represents a sequence of business behaviors that achieves a specific result such as a defined set of products or business services.	
Business object	Represents a concept used within a particular business domain.	

Source: ArchiMate® 3.1 Specification

Co-financed by the Connecting Europe Facility of the European Union