



EUROPEAN COMMISSION

CORPORATE REFERENCE DATA MANAGEMENT IN THE EUROPEAN COMMISSION

POLICY

Endorsed by the Information Management Steering Board
on 3 February 2022

*Prepared by the Corporate Reference Data Management Coordination group, chaired by
DIGIT, OP and SG*

Table of Contents

1.	Introduction.....	3
2.	Rationale.....	3
2.1.	Challenges.....	4
2.2.	Benefits	4
3.	Policy objectives	5
3.1.	Making existing reference data assets accessible and reusable.....	5
3.2.	Establishing corporate governance for reference data	6
3.3.	Defining a common management approach for reference data.....	6
3.4.	Stimulating the reuse of reference data across the Commission	6
3.5.	Encouraging the use of corporate interoperable tools for reference data management.....	6
4.	Principles of governance.....	6
4.1.	Overall approach.....	6
4.2.	Specific approaches	7
4.3.	Responsibilities	8
4.3.1.	Strategic dimension.....	8
4.3.2.	Management dimension	8
4.3.3.	Operational dimension	8
4.4.	Interactions with existing governance structures	8
4.4.1.	Overall structure.....	8
4.4.2.	Role of the coordination group	9
4.4.3.	Role of the corporate asset liaisons	10
5.	Management principles and best practices.....	10
5.1.	Availability	10
5.2.	Communication and notification.....	11
5.3.	Extensibility	11
5.4.	Feedback and request handling	11
5.5.	Multilingual descriptions	11
5.6.	Multiple publication formats and access mechanisms	11
5.7.	Persistent identifiers.....	12
5.8.	Release, Versioning and Archiving	12
5.9.	Reuse of existing internal reference data	12
5.10.	Reuse of external reference data	12
5.11.	Metadata to describe reference data assets.....	12
5.12.	Metadata to describe the concepts inside the assets.....	13
5.13.	One concept type per asset.....	13
5.14.	Naming conventions	13
6.	Corporate reference data assets for shared use.....	13
6.1.	Focus on cross-domain assets	13
6.2.	Process to add reference data assets.....	14
Annex A	Glossary	16

1. INTRODUCTION

This document contains an updated version of the policy for Corporate reference data management in the European Commission, following the endorsement of the initial proposal¹ by the Information Management Steering Board (IMSB) of the European Commission in its meeting of 1 July 2020.

In the second half of 2020 and in the year 2021, the policy was implemented under the aegis of the Corporate Reference Data Management (CRDM) Coordination Group chaired by DIGIT, OP and SG. During this first implementation period, the governance and management principles described in the proposal have shown their value. Based on emerging experience, further details were discussed and agreed, and those further agreements led to the updates in this document.

In the area of governance, a description of the relationship between the coordination group on the central level and the local, DG/service level has been added in section 4.3.

In the area of management, the principles have led to a set of best practices that were developed and agreed by the coordination group. A summary of these best practices is included in section 5 with a link to the full text of the best practices.

In the policy proposal in 2020, the list of the selected corporate reference data assets was included. This list is now maintained separately from the policy document to allow for easier maintenance and evolution. The link to the list is included in section 6.

Finally, the agreed process for the selection of additional corporate assets is included in section 6.2.

2. RATIONALE

In October 2019, the IMSB decided to prioritise, among other actions, the development of data policies and guidelines for reference data. Therefore, a *reference data management coordination group* was set up. The objective of the coordination group is to build consensus on a corporate reference data management policy and to agree a set of principles for the management of reference data assets as well as to establish an initial set of reference data assets to be shared for common usage across the Commission.

As it is envisaged in the *Data governance and data policies at the European Commission*² document, Directorates General (DGs)/services will prioritise high-value data assets, such as reference data; a Commission-wide reference data domain will be created on the basis of the Commission's data inventory; DGs/services will give priority to the reuse of common reference data; and a common governance will ensure that reference data management will follow corporate good practices.

¹ Corporate Reference Data Management in the European Commission. Draft Policy.
https://op.europa.eu/documents/3938058/8065862/CorporateReferenceDataManagementPolicyInTheEC_v20210316_public.pdf.

² Data governance and data policies at the European Commission. July 2020.
https://ec.europa.eu/info/publications/data-governance-and-data-policies-european-commission_en.

2.1. Challenges

The *reference data coordination group* identified a number of challenges at its inception in 2020. While there was a wealth of reference data already available at various DGs/services, it was not widely shared and reused, which hindered interoperability and data exchange. In particular, the group identified the following challenges:

- *Finding relevant reference data*
Today it is often not clear how and where to find existing reference data. Some assets may be known through common knowledge, through personal contacts or through browsing various sources. It is also not easy to find out what the conditions are under which existing reference data can be reused and which data formats are supported for a particular asset. These ad-hoc approaches lead to inefficient use of the existing reference data.
- *Understanding which DG/service is responsible for the management*
Currently it is not always clear which DG/service assumes data management responsibilities, what the scope of those responsibilities is and how to get in contact with the maintaining organisation in case of problems or questions.
- *Knowing how a reference data asset is managed*
The way that reference data are managed currently in the DGs and services is based on local approaches to support applications and stakeholders of the IT systems used. This leads to an inconsistent landscape where users may not always understand how the management is organised, who is responsible for the maintenance, how decisions on changes are taken and how and when versions are released.
- *Avoiding duplicates and overlapping reference data*
Reference data of the same type may be created and maintained by different DGs and services as well as by external organisations, e.g. (inter)national public bodies and standards organisations, and the same type of reference data may exist in different places (e.g. lists of countries or languages), even inside one DG. This situation leads to duplication of effort as these different assets need to be maintained independently. At the same time, similar reference data maintained in different places is likely to lead to incompatibilities and possible data quality issues.
- *Reuse of existing reference data*
The reuse of existing reference data is currently left entirely to the discretion of a DG/service, for example in the design of new IT systems or in case an IT system is redesigned. This leads to limited reuse, incompatibilities between IT systems and inefficient use of resources.

2.2. Benefits

A common approach to the management of reference data and the sharing and reuse of a common set of reference data assets brings benefits to the Commission to gain efficiencies and effectiveness.

- Reference data management will be aligned with the general principles of data governance and policies in the Commission.

- Interoperability across Commission services and IT systems will increase because data can be easily cross-referenced and integrated.
- The adoption of corporate code lists and classifications will enable reusability and play a crucial role in linking datasets from different sectors and domains
- A common approach will remove obstacles in discovering, accessing, sharing, combining and re-using data assets to improve information discovery and support decision making processes across different policy areas and institutions for a tangible data-driven public sector
- The adoption of common classifications across different domain will reduce the costs and the effort, often manual, in managing and creating mappings
- The collection of corporate reference data assets will enable that those assets can easily be found and reused.
- A comply-or-explain policy will ensure that in all relevant cases a common set of reference data assets is used.
- Data duplication will decrease, as different services will be able to rely on the same set of reference data.
- Unnecessary (double) work in defining and maintaining reference data assets will be avoided.
- The quality of reference data will improve substantially.
- Reference data will be available in multilingual form across the EU in machine-readable format, in particular to further data exchange with and amongst EU institutions and Member States' public administrations.

The Commission can take a leading role by sharing the reference data assets, as much as possible, as open data, and demonstrating best practice in creating and maintaining reference data and fostering the reuse of public sector information for value creation to support interoperability and the multidisciplinary research on important social problems and challenges modern societies face.

3. POLICY OBJECTIVES

To overcome these challenges, this Commission-wide policy for reference data management has the following objectives:

3.1. Making existing reference data assets accessible and reusable

Reference data assets must be made available for reuse across the Commission and, if relevant, by other European institutions as well as by Member States' public administrations and enterprises.

A central registry, part of the Commission's Data Catalogue, will facilitate the access to a collection of reference data assets, especially those considered relevant for common use across the Commission, providing information about their reuse conditions and the formats in which they are available.

3.2. Establishing corporate governance for reference data

The responsibility to maintain every commonly used reference data asset must be made clear by assigning one DG/service to act as the authoritative source, i.e. the data owner. This implies a commitment to the basic management principles described in this policy.

3.3. Defining a common management approach for reference data

The management of reference data assets must be exercised following a set of common principles. All DGs/services should align with these principles.

3.4. Stimulating the reuse of reference data across the Commission

The reuse of the commonly used reference data assets by Commission DGs/services is strongly encouraged under a *comply-or-explain policy*, providing rules for the use of common resources, in order to promote reuse, but also to allow for exceptional cases where such reuse would be impossible, unreasonable or detrimental.

3.5. Encouraging the use of corporate interoperable tools for reference data management

Commission DG/services must reuse existing corporate solutions for the management of their reference data assets, once the list of corporate solutions will have been established. Using corporate solutions for reference data management encourages knowledge sharing, avoids reinventing the wheel and reduces unnecessary expenses for software development, consultancy and customisation.

4. PRINCIPLES OF GOVERNANCE

4.1. Overall approach

The governance of corporate reference data follows the framework defined in *Data governance and data policies at the European Commission* [Error! Bookmark not defined.](#)

The governance structure respects the three levels for data governance: **strategic, managerial, and operational**. As much as possible, these roles will be integrated into existing structures and competences at the DGs/services.

On the **strategic** level, the IMSB will be the ultimate steering body. This is natural, as reference data management falls within its remit. The IMSB oversees the work on the management level.

On the **management** level, a coordination group on corporate reference data in the Commission exists. This group reports to the IMSB and informs the Local Data Correspondents' (LDC) network, the IT Investment Team and other relevant groups. The members of the group represent the main stakeholder DGs and services, including DGs AGRI, BUDG, COMP, DGT, DIGIT, EAC, ENV, Eurostat, GROW, HR, INTPA, JRC, MARE, the OIB, the Publications Office, SCIC, and the SG. The group is jointly chaired by the Secretariat-General, DG DIGIT and the Publications Office.

On the **operational** level, for each corporate reference data asset, the coordination group designates a DG/service as the *data owner* and the same or a different DG/service as the *data steward*.

The *data steward* commits to:

- Keep the corporate reference data assets up to date.
- Provide support to the users of that asset and respond promptly to any requests concerning the asset it maintains.
- Make available a functional mailbox for requests and feedback and a notification channel to inform the users of upcoming changes and releases.

4.2. Specific approaches

Reference data defined in legal text: When commonly used reference data are defined in legal text, e.g. the INSPIRE themes or the Nomenclature for Statistical Territorial Units (NUTS), the data steward, in addition to the role described above, establishes contact with the DG that develops and maintains the legal instrument to be notified at an early stage for upcoming changes. It is recommended to remove the definition of the reference data from the body of the legal text and maintain the details of the reference data separately in a technical document referenced in the legal text to make the maintenance more agile.

Internally defined reference data: When corporate reference data have been defined by a DG/service without reference to an external authoritative source, the data steward acts as the contact point for all matters related to the reference data assets they maintain, e.g. the Publications Office does so for the EuroVoc thesaurus and the list of EU programmes.

Externally defined reference data, reused without modifications: When reference data are owned and managed by an external organisation, e.g. an international standardisation body, the data steward will act as a ‘proxy’ between the Commission and the external organisation, making sure that the users are notified of changes in the external reference data. An example is the list of the sustainable development goals³ from the UN.

Externally defined reference data, reused with modifications: When reference data reused from external sources are extended or modified, e.g. adding codes for internal use, the data steward, in addition to the role described above, also acts as a ‘proxy’, making sure that the users are notified of changes in the external reference data, and, if possible, maintains contact with the external organisation that maintains the source reference data. Examples are the list of languages based on ISO 639 and the list of currencies from ISO 4217.

³ <https://sdgs.un.org/goals>.

4.3. Responsibilities

4.3.1. Strategic dimension

The IMSB exercises oversight through regular reporting from the coordination group to meetings of IMSB, highlighting the activities around the corporate reference data, the tangible results achieved and any proposals for further evolution of the policy.

4.3.2. Management dimension

The coordination group is responsible for working out and agreeing further details of the policy, including the addition of additional assets to the corporate list, requirements for tools for the management of assets and common approaches on the operational level.

The coordination group consists of members that represent DGs and services, some of which, called **corporate asset liaisons**, have a responsibility for one or more corporate assets.

The coordination group may involve the Local Data Correspondents for review of proposals.

4.3.3. Operational dimension

Operationally, DGs/services that maintain one or more corporate reference assets collaborate to align their processes and procedures on the basis of agreements on best practices in the coordination group.

4.4. Interactions with existing governance structures

4.4.1. Overall structure

The corporate governance described in the previous section does not operate in a vacuum. Reference data assets that are included in the list of corporate assets are governed by existing governance structures in a DG or (inter)service environment as depicted in Figure 1. This governance can also have an international dimension (as it is for some assets managed by Eurostat, JRC and others).

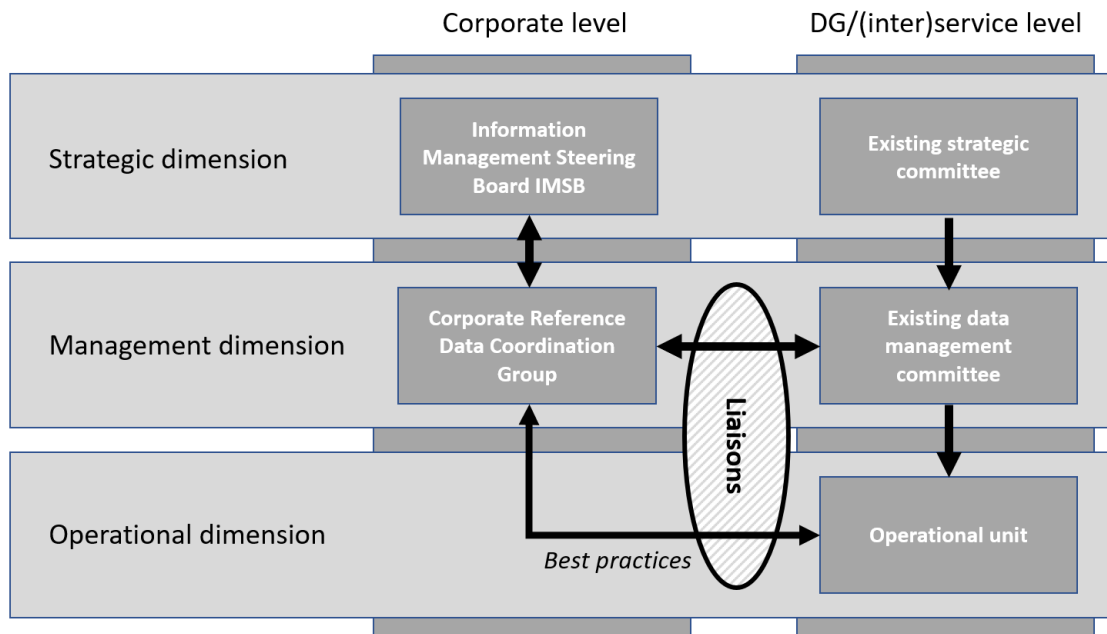


Figure 1 Interactions between corporate governance and DG/(inter)service governance

The relevant interactions in this diagram, from the point of view of the coordination group, are those with existing management committees, to align processes and decisions related to the corporate asset, and with the operational units, to coordinate the application of best practices defined on the corporate level.

4.4.2. Role of the coordination group

The main role of the coordination group, in addition to the role as the custodian of the corporate reference data management policy, is to provide a platform for the members of the group to discuss common challenges and to share experiences. In this sense, the group acts as a sounding board supporting the activities related to the management of the corporate assets.

The coordination group is also responsible for evaluating the policy, the interactions with the liaisons and the experience with the best practices. Discussions on issues raised may lead to updates of the policy, restructuring of the governance approach, refinement of existing best practices, or the development of new best practices.

The practical activities of the group include:

1. Discussing additional assets proposed by any member of the coordination group, and taking decisions on adding such assets to the corporate list following the agreed procedure for such inclusion;
2. Reviewing information about upcoming changes to corporate assets notified by the liaisons;
3. Discussing and providing advice on challenges raised by liaisons;

4. Discussing any other issue raised by any member of the coordination group related to the governance and management of reference data.

The coordination group reports to the IMSB on its activities and achievements on a regular basis.

4.4.3. Role of the corporate asset liaisons

The members of the coordination group that have responsibility for one or more corporate assets act as liaisons between the corporate level and the DG/(inter)service level. They are the persons that have full control over the corporate asset or represent groups that have that control.

Their activities include:

1. Suggesting additional assets for inclusion in the corporate asset list;
2. Informing the coordination group of upcoming changes in the corporate asset, ideally with sufficient time ahead of publication, in order to give the coordination group the opportunity to comment;
3. Raising any issue they encounter with the governance and management of reference data;
4. Ensuring implementation of agreed best practices for the management of the corporate asset and reporting on the ongoing experience and any challenges with such implementation.

In all the interactions between liaisons and coordination group, the final responsibility for changes and approaches remains with the local responsible person or group within the boundaries of the corporate reference data management policy

5. MANAGEMENT PRINCIPLES AND BEST PRACTICES

The management and operational levels need to adhere to clear and effective management principles that lead to a coherent and consistent management landscape, which in turn will deliver the benefits that the policy aims to achieve.

These management principles are further developed in the form of best practices that are developed, maintained and made available from <https://op.europa.eu/en/web/eu-vocabularies/corporate-reference-data-management>.

In addition to the best practices summarised below, additional best practices may be developed in response to emerging needs from implementation experience.

5.1. Availability

Corporate reference data assets must be available without interruption with sufficient measures to ensure performance and resilience through trusted services and platforms. The wider use of corporate reference data makes it necessary that the asset can be accessed at

anytime from anywhere at the Commission. In practice, the level of commitment for each reference data asset must be determined based on an analysis of how the asset is used.

5.2. Communication and notification

A communication and notification channel must be available for all users to be kept informed of any changes in the assets that they use, so that they can take necessary actions, for example to make change requests before a cut-off date, to participate in review of proposed changes or to plan changes to local systems to prepare for changes in the reference data. If the reference asset is associated with a clearly delineated user community, the existing channels for that community should be used. If usage is more differentiated and the extent of the user community is not known, any channels available should be used, for example a notice on the website associated with the asset.

5.3. Extensibility

Reference data assets have to allow for local extensions in which concepts are added that are only relevant for particular applications. Local applications may have a need for an extended or enriched set of concepts, either for additional concepts or for a deeper level of concepts underneath existing concepts. Extensions need to be visible also on the corporate level to other users of the asset.

5.4. Feedback and request handling

Clear and well-documented procedures and processes must be established to allow stakeholders of a reference data asset to provide feedback and make change requests. Information provided to the stakeholders must include a reference to the responsible person, unit, or committee, a functional mailbox for feedback and requests, and a link to an issue tracker if available. Submitted feedback and change requests must be handled in a reasonable timeframe, acknowledging receipt, and providing information on the process that will be followed to handle the submission including an indication when the submitter can expect to receive a response.

5.5. Multilingual descriptions

All assets and concepts must follow the language policy of the European Commission⁴. The European Union is a multilingual environment where the European Commission should set an example of good practice. The stated aim of the language policy is to strike a reasonable balance between respect for speakers of the EU's many languages and practical considerations such as the cost of translation.

5.6. Multiple publication formats and access mechanisms

A corporate reference data asset must be available for distributions in multiple, machine-readable formats to allow reuse by applications that use different technologies. These

⁴ https://ec.europa.eu/info/languages-our-websites_en.

formats may include schema expressions using RDF/XML⁵, JSON-LD⁶ and XML Schema Definition⁷. The choice of expression should be made based on the expectations of the primary users of the asset.

5.7. Persistent identifiers

Concepts in a reference data asset must be identified by persistent identifiers in order for them to be accessed and reused consistently over time. As the concepts of a reference data asset will be used as values in the description of data resources, it needs to be clear in all cases what the exact meaning of such a concept is, and therefore there needs to be an unambiguous way for the user to look up the meaning of the concept.

5.8. Release, Versioning and Archiving

The current and all previous versions of a reference data asset must be kept for access. If possible, also all versions of the individual concepts in the asset should be kept. In most cases, reference data need to reflect changes in the environment. New countries or organisations are established, old ones are discontinued, current affairs and evolving knowledge require new concepts to be described. The release schedule must be documented and published, even if releases are done ad-hoc. The current and all previous versions of a reference data asset must be kept for access.

5.9. Reuse of existing internal reference data

As a default, new IT systems developed by and for the Commission must reuse reference data that already exist at the Commission, and in particular, the assets that are managed under this reference data management policy. This reuse is governed by the comply-or-explain policy that is designed to encourage the reuse of such existing reference data unless such reuse would not be possible or feasible.

5.10. Reuse of external reference data

As far as possible, externally defined reference data have to be reused, provided that an external asset is managed by a trustworthy source which satisfies criteria of good design and good maintenance. Such trustworthy sources include decentralised EU agencies like EEA, EIGE and ECDC, international organisations like the UN, FAO, WHO or OECD, or standardisation bodies like ISO or CEN.

5.11. Metadata to describe reference data assets

All corporate reference data assets must be described with metadata according to a common specification, aligned with the metadata specification for the EC Data Catalogue.

⁵ <https://www.w3.org/TR/rdf-syntax-grammar/>

⁶ <https://json-ld.org/>

⁷ <https://www.w3.org/TR/xmlschema-0/>

5.12. Metadata to describe the concepts inside the assets

A common, minimum set of metadata elements to describe the concepts in the assets must be agreed. This is to ensure that sufficient information is provided for the user to decide whether a particular concept is useful. A common machine-readable schema will also enable a search function to be provided in a central registry on the level of concepts. In particular, when external reference data are used as the basis for an internal asset, each concept in the internal asset should reference the concept in the external asset on which it is based.

5.13. One concept type per asset

A reference data asset should contain only concepts of a single type. This principle makes it easier to manage the reference data. If, on the contrary, an asset would combine different dimensions, such an asset is harder to manage than two assets separately.

5.14. Naming conventions

For new assets and concepts, a common naming convention must be agreed. A name has to appropriately reflect the scope and meaning and be neither too broad nor too narrow. The naming convention needs to take into account existing approaches as well as allowing for possible evolution of the assets and concepts.

6. CORPORATE REFERENCE DATA ASSETS FOR SHARED USE

6.1. Focus on cross-domain assets

In the list of corporate reference data assets, the priority is on cross-domain assets, i.e. those ones that can be used across several or all DGs and services. Examples are languages, economic activities, and geospatial reference data. In cases where there are multiple candidates for a particular type of concept, a dedicated working group is established in coordination with existing expert groups to identify the possible reference sources and to define the context in which the candidates will be recommended.

The current list is made available at <https://op.europa.eu/en/web/eu-vocabularies/corporate-reference-data-list>. It is managed by the coordination group for corporate reference data and will evolve over time.

The list of corporate reference data assets is made available to the IT Investments Team (ITIT) at the Commission, to make sure they are considered in the development of new IT systems. In addition, it will be included in the EC Digital Strategy handbook⁸, which is the checklist for the ITIT and for IT projects to adhere to the principles of the digital strategy and to other standards and obligations related to the execution of IT projects.

⁸ EC Digital Strategy handbook - Corporate IT Governance wiki - European Commission - Intranet Wiki.

6.2. Process to add reference data assets

A reference data asset can be considered for inclusion in the corporate list if it meets the following criteria:

- Documented use by two or more DGs/services, or a credible business case of how the asset would increase efficiency if it were used across multiple DGs/services.
- Governed according to the reference data governance policy of the European Commission with clearly assigned management and operational roles, or a commitment to adapt the existing governance (if any) to the corporate one.
- Managed according to the management principles specified in the corporate reference data management policy, or a commitment from the proposing DG/service to align with the management principles.

Figure 2 shows the steps taken to determine if an asset qualifies as a corporate reference data asset.



Figure 2: Process for corporate reference data assets

Any DG/service may propose a reference data asset managed by them by submitting a request to the coordination group on corporate reference data with documentation of how the asset meets the criteria mentioned above.

If a proposed reference data asset is defined and managed outside of the Commission, the DG/service will be asked to provide evidence that the external reference data asset is maintained by a trustworthy organisation following governance and management policies equivalent to the ones defined in this policy.

If, in the opinion of the coordination group, the proposed asset covers a type of concept for which also other candidates exist, it will bring together the DGs/services that manage assets for that same topic. Such assets may be related in a hierarchy, for example, global/European/regional/local levels in geographic information. The coordination group will determine whether assets might be combined or provide clear guidelines about the context in which each of those candidates should be used.

The coordination group will take a decision on inclusion of the proposed asset in the corporate reference data assets list and will report on changes to the list and on planned

work on new reference data assets to the IMSB on a regular basis. For politically sensitive sets of reference data, the group will be guided by the IMSB, its chair or the IMSB secretariat in the Secretariat-General, to ensure a broader/higher level consultation/endorsement, before including them in the corporate reference data assets list.

Annex A Glossary

Data asset means any collection of data, data sets or information somehow linked, e.g. by common codes or metadata, which has been created by the Commission, collected from Member States or other stakeholders, or acquired from third parties in the context of projects, policy or administrative processes. Data assets may be structured or unstructured. Structured data assets are organised according to a predefined data model or schema and the content of each field/variable can assume only predefined values. Unstructured or semi-structured data assets are not structured via predefined data models, schemata or code lists. Data assets may be static or dynamic. Data assets may be raw or curated. Data assets are in digital formats [Based on C(2016) 6626 final and Directive (EU) 2019/1024].

Data domain means a logical representation of a category of data assets that has been designated and named. Data domains are not physical repositories or databases. Instead, they are 'logical' categories or groupings of data that are deemed important and necessary to the Commission's operation. Data domains include both internally generated data as well as externally acquired data. [Source: EDM Data Management Glossary (85)].

Data policy means a set of broad, high-level principles which form the guiding framework in which data assets in the Commission can be managed [Source: European Interoperability Reference Architecture].

Reference data, such as code lists and authority tables, means data that are used solely to characterise or relate to other data. They are fundamental building blocks of most information systems. Using common interoperable reference data is essential for achieving interoperability [Source: Data governance and data policies at the European Commission. July 2020].