European Commission

Data governance and data policies

at the European Commission

Secretariat-General

EUROPEAN COMMISSION
Secretariat-General

July 2020

CONTENTS

EXECUTIVE SUMMARY

The Juncker Commission put great emphasis on improving policymaking and internal processes by making it easier to access internal and external data, and extract insights from these data with advanced digital technologies. The Commission's data strategy constitutes a corporate commitment to transform the Commission into a data-driven organisation, enabled by a data ecosystem governed by corporate data governance and data policies. President von der Leyen's 'political guidelines' ([1]) set out a vision of a Commission that leads by example and is fully digital, agile, flexible and transparent, and emphasise the 'need to share' data, considering data protection, information security and intellectual property.

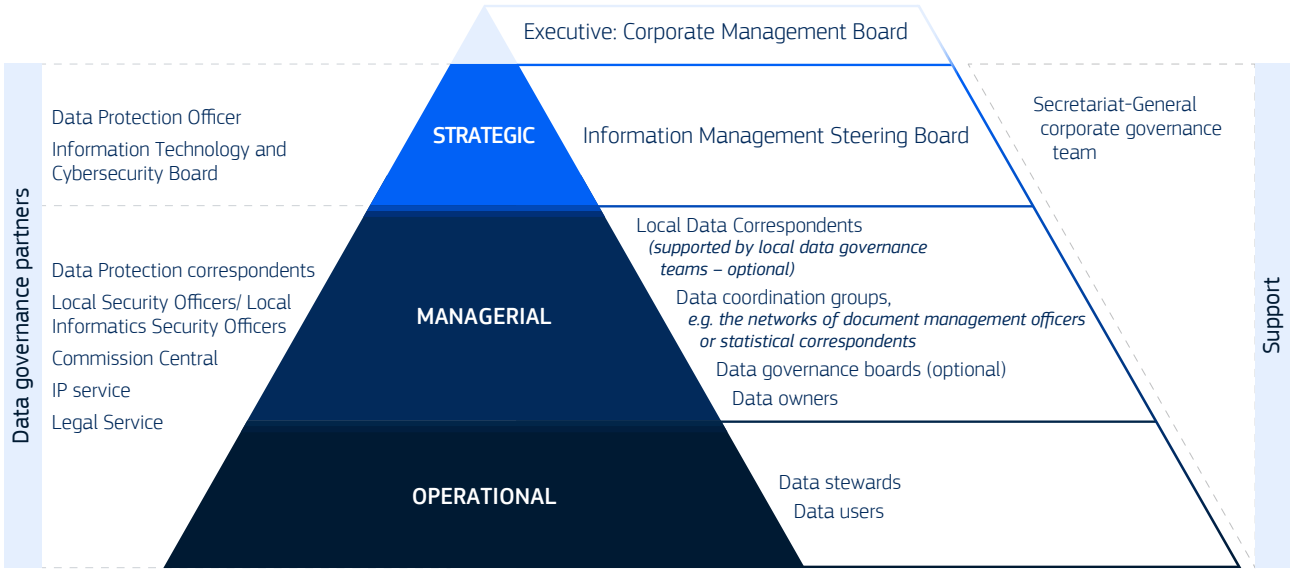
To ensure that the Commission can build on the available data and information as effectively and efficiently as possible when developing its policies, it must invest in data governance and data policies today. To implement the von der Leyen Commission's 'whole of government approach', obstacles to internally sharing, combining and reusing data assets will need to be removed, where and when possible. The 'need to share' principle should become the norm for sharing data, information and knowledge in the organisation.

This is a 'living' document. Its purpose is: (i) to show how data governance and data policies can allow the Commission to transform into a data-driven organisation; (ii) to provide direction; and (iii) to identify areas for further work. It is the result of joint work between the Secretariat-General and the local data correspondents network. It is informed by and consolidates existing local initiatives under a common corporate framework, and is aligned with international standards and good practices in the field.

Trusted and reliable analytics and artificial intelligence — which are key ingredients for transparent, evidence based policy-making— require findable, accessible, interoperable, secure and high-quality data.

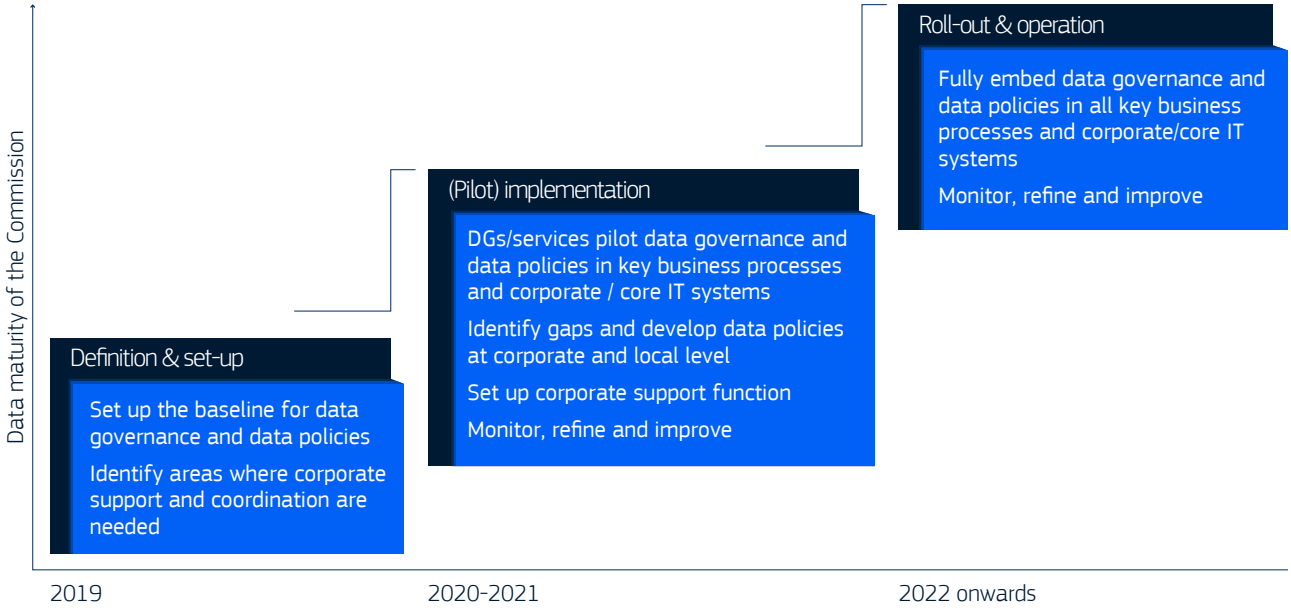
Data governance and data policies help the Commission comply with regulatory and legal requirements, notably those linked to data and document management, access to data and documents (including open data), data protection, intellectual property and information security, thereby reducing associated risks. They deliver medium to long-term efficiencies in resources spent on data management, due to the optimisation of data creation, collection, acquisition, access, use, processing, sharing, preservation and deletion, and to better data quality.

Data governance sets out a framework with clear roles, and the responsibilities and interdependencies of those roles.
Data policies introduce common principles, guidance and working practices in the areas of data management, data interoperability and standards, and data quality. Equally important are the areas of data protection, information security and intellectual property. However, these are not the focus of this document. Corporate data policies do not specify detailed processes. This allows Directorates-General (DGs)/services to organise themselves in the way that best suits their internal organisation, while ensuring coordination and alignment across the Commission, including its executive agencies.



([1]) *A Union that strives for more: my agenda for Europe – political guidelines for the next Commission 2019-2024.*

Implementing data governance and data policies will require action and investment at both corporate and local level. The Commission is steadily improving the processes underpinning its data strategy. In 2019 it started with the definition and set-up of the framework. From 2020 onwards it will proceed with iterative development, progress monitoring and other improvements. At first, DGs/services will prioritise high-value data assets, such as master and reference data. Implementation will be iterative, in that sets of principles will be introduced and then reviewed at the end of each phase to analyse the impact on business processes and IT systems. Unless required by binding legislation, such as a regulation or a Commission decision, or specifically mentioned as 'optional', these policies are implemented on a 'comply-or-explain' basis. In line with the proportionality principle, the Commission's DGs/services ([2]) are expected to implement these policies, and any diverging local approaches should be justified.



The following initiatives/bodies will help DGs/services implement data governance and data policies locally:

- The local data correspondents network, which is operational since May 2019. The members of the network share knowledge, experience and practices for successful implementation. They will also receive training to prepare them better for their roles.
- A data advisory service by the Joint Research Centre and the Directorate-General for Informatics (DIGIT), with the participation of the Publications Office (OP) and Eurostat (ESTAT), will be launched in mid-2020. It will provide consultancy and support for data governance and data management projects.

([2]) This includes DGs, services and executive agencies.

1. INTRODUCTION

1.1. Context and scope

The 2016 Communication on data, information and knowledge management at the Commission ([3]) emphasised the need to improve information retrieval and delivery and to maximise the use of data for better policy making. To this end, the Commission's data strategy set out a corporate commitment to transform the Commission into a data-driven organisation, enabled by a data ecosystem governed by corporate data governance and data policies. It helps the Commission implement its digital strategy for a digitally transformed, user-friendly and data-driven administration ([4]). It highlights the importance of putting data creation, collection, acquisition, access, use, processing, sharing and preservation at the centre of the Commission's digitalisation process.

In 2018, the Commission adopted a package of measures ([5]) to strengthen its corporate governance structure with the Corporate Management Board and its supporting bodies, including the Information Management Steering Board (IMSB). The IMSB helps steer data governance and data policies, and recognises them as crucial for achieving the Commission's vision of unlocking the power of data to improve policy making.

Data governance and data policies aim to provide guidance, assurance and support that will transform the Commission into a data-driven organisation by:

- defining clear roles and responsibilities; and
- introducing common principles, guidance and working practices that provide the foundation for harmonised and coordinated data management across the organisation.

The scope of data governance and data policies includes data assets that are owned, used or reused by the Commission and its executive agencies regardless of their level of information confidentiality ([6]). This includes data used for policymaking, administrative data and personal data.

For third-party data assets where the Commission is not the owner but the user, processor or controller, e.g. data collected from Member States due to a legal obligation or data licensed from third parties ([7]), the Commission remains responsible for their management for as long as it is using them both for their primary purpose and for secondary purposes (i.e. reuse).

The data governance and data policies laid out in this document are applicable to DGs, services and executive agencies. Decentralised agencies may also consider following them.

1.2. What are data governance and data policies?

Data governance entails defining, implementing and monitoring strategies, policies and shared decision-making over the management and use of data assets. It is performed by Commission staff with established data-related roles.

Data policies are a set of broad, high level principles ([8]) which form the guiding framework in which data assets in the Commission can be managed ([9]). More specifically, data policies govern data management, data interoperability and standards, data quality, data protection and information security.

([3]) C(2016) 6626 final.

([4]) https://ec.europa.eu/info/publications/EC-Digital-Strategy_en

([5]) Governance in the Commission.

([6]) Levels of information confidentiality as defined in C(2019) 1903 final.

([7]) In the case of acquired data assets the Commission obtains IP ownership, e.g. via a service contract, whereas when data is licensed, the Commission obtains no IP ownership but only the rights of use.

([8]) Where needed, data policies will be complemented by detailed guidelines and processes. Such guidelines and processes may be developed either at corporate level, interservice level or local level. This remains outside the scope of this document

([9]) See definition of 'data policy' in the European Interoperability Reference Architecture.

A **data asset** is any collection of data, any data set or any information that is 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 ([10]), static or dynamic, raw or curated. Data assets are in digital formats.

1.3. Why invest in data governance and data policies?

To ensure that the Commission can build on the available data and information as effectively and efficiently as possible when developing its policies, it must invest in data governance and data policies today. To implement the 'whole of government approach' and the 'need to share' principle of the von der Leyen Commission, obstacles to internally sharing, combining and reusing the data assets will need to be removed.

Corporate data governance and data policies will improve the traceability, accessibility, and preservation of data across the entire Commission, which will increase the transparency of the evidence that underpins policymaking. This will allow the Commission to deliver on its commitments under both the Interinstitutional Agreement on Better Law-Making and the better regulation policy, which will strengthen Interinstitutional cooperation and trust in the EU policy-making process.

Data governance and data policies, accompanied by advanced analytical models, will help the Commission improve its evidence-based policymaking and enhance its internal processes. Trusted and reliable analytics and artificial intelligence require accessible, trusted, unbiased, interoperable, secure and high-quality data.

Data governance and data policies help the Commission comply with relevant regulatory and legal requirements — notably those linked to data and document management, access to data and documents (including open data), data protection, intellectual property and information security — thereby reducing associated risks.

Data governance and data policies will create medium to long-term efficiencies in data management resources, due to the optimisation of data creation, collection, acquisition, access, use, processing, sharing, preservation and deletion, and to better data quality.

([10]) 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.

1.4. How to implement data governance and data policies?

Implementing data governance and data policies will require action and investment at both corporate and local level. It should be coordinated with the implementation of domain specific data policies, e.g. for statistical or geospatial data, and the implementation of the Commission's digital strategy to maximise synergies.

The Commission is steadily improving the process underpinning its data strategy. In 2019 it started with the definition and set-up of the framework. From 2020 onwards it will proceed with iterative development, progress monitoring and other improvements (see figure below).

Corporate actions will be organised to help DGs/services apply data governance and data policies.

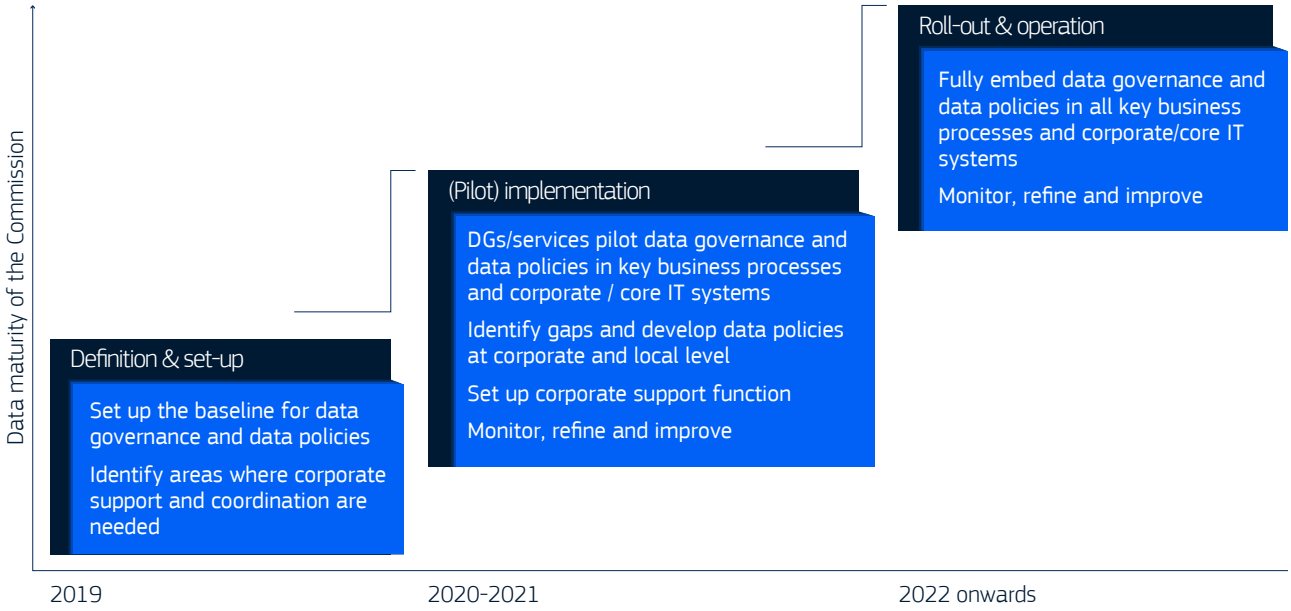


Figure 1: Data governance and data policies implementation at the Commission

1.5. Guiding principles

Data governance and data polices at the Commission will be:

- Embedded: the focus should be on their seamless integration into existing business processes and their simplification, whenever possible, which remains the responsibility of DGs/services. Data governance and data policies should not introduce additional bureaucratic burden in existing business processes. Any costs introduced by data governance and data policies will be neutralised in the medium term from the added value they provide.
- Sustainable: data governance and data policies should not be seen as a project with a pre-determined end date. They are a continuous improvement process undertaken by the Commission and sponsored by the highest level of management.
- Measurable: being able to measure progress and plan ahead is crucial for continuous improvement. The progress and impact of data governance and data policies on data quality, people, business processes and IT systems will be measured regularly to maintain the commitment of all stakeholders.
- Accountable and responsible: good governance makes clear to all stakeholders their own roles and responsibilities as well as those of others; it communicates the relevant policies clearly and encourages behaviours and actions that are consistent with these. A culture of collaboration and shared responsibility for data-related matters, from quality through to protection and security, will be nurtured, going beyond compliance to rules and requirements.
- Transparency-oriented: To increase trust in its policymaking process, data governance and data policies, the Commission should seek to enable other EU institutions and agencies, Member States' administrations and third parties to access and reuse Commission data assets, in particular those used for policymaking.
- Principle-based: Corporate data governance and data policies should focus on laying down principles and providing guidance, rather than on specifying detailed processes. This will allow DGs/services to organise themselves in the way that best suits their internal organisation, while ensuring a common level of coordination and harmonisation across the Commission. These principles will therefore be reflected in the processes to be specified and implemented by the DGs/services for managing their data assets. These principles may be reviewed or extended as the implementation of data governance and data policies in the Commission progresses.
- Commission-wide and comprehensive: Data governance and data policies may be implemented locally but need to be coordinated across the organisation to be effective. They govern primarily the way people interact with data assets, and are implemented through changes in business processes, IT systems and staff. At the same time, the full lifecycle of data assets should be managed. Data-related risks, especially those linked to data protection, intellectual property, and information security will also be managed.
- Proportionate: Some DGs/services are more 'data intensive' than others. Activities and investments to implement data governance and data policies will be proportionate to the size of each DG/service, their data management needs (including type of data managed and information confidentiality level) and capabilities, and the relative importance of data management for their operations. Less 'data intensive' DGs/services are encouraged to collaborate with others, e.g. as part of families, to benefit from synergies.
- 'Comply-or-explain': Unless required by a binding instrument, such as a regulation or a Commission decision, or specifically mentioned as 'optional', data policies are implemented on a 'comply-or-explain' basis. DGs/services are expected to implement the principles and requirements introduced by the data policies. Otherwise, they will need to provide a formal justification to the right level of governance.

2. DATA GOVERNANCE ROLES AND RESPONSIBILITIES

This section specifies the corporate roles (i.e. those of boards and groups) and the individual roles of those involved in data governance and their responsibilities ([10]). There are three levels of data governance ([11]):

- Strategic, which defines the long-term vision, gives direction, oversees progress, takes strategic decisions, and acts as the highest point of reference for issues and matters related to data governance and data policies.
- Managerial, which is accountable and responsible for developing and implementing data policies at corporate level and local level. It monitors progress, reports to the strategic level and refers to them any issues and matters that are beyond its decision-making power or mandate.
- Operational, where data policies are actually implemented and most decisions about data are taken. Whenever necessary, issues are escalated to the managerial level for resolution.

Figure 2 provides a general overview, but is not an exhaustive mapping of all data governance or Commission-level management committees, groups and networks. Additional layers may be possible, depending on the domain's complexity and the organisational set-up of a DG/service.

All knowledge and practices relating to data management and data quality will be actively shared within and across groups at all three levels (strategic, managerial or operational) by means of corporate collaboration solutions and practices.

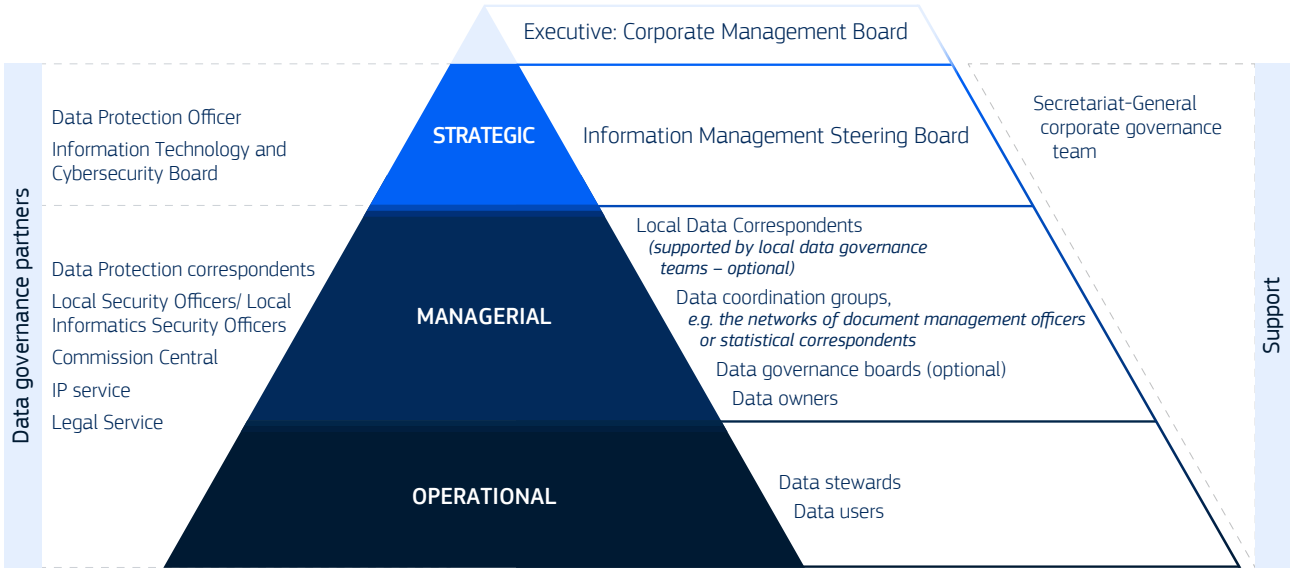


Figure 2: Commission data governance levels and roles

([10]) Principles listed in these document are aligned with the principles put forward by the European Commission Digital Strategy.

([11]) Inspired from R. S. Seiner, '*Non-Invasive Data Governance™: The Path of Least Resistance and Greatest Success*', first edition. Technics Publications: 2014.

2.1. Boards and groups

2.1.1. Strategic level - Information Management Steering Board

The Information Management Steering Board (IMSB) is a permanent subgroup of the Corporate Management Board. It supports the Corporate Management Board in overseeing the implementation of the Commission's strategy on data, information and knowledge management. The IMSB is supported in its role by the Information Management Team. With regard to data governance and data policies, the IMSB:

- steers and approves the corporate data strategy, data governance and data policies;
- provides an opinion on local data policies, while leaving room for DGs/services to develop and implement their own approaches tailored to their specific needs;
- prioritises corporate actions and seeks to enhance collaboration and synergies in the implementation of the strategy;
- makes strategic decisions that cannot be made at the managerial and operational levels;
- monitors the implementation of data policies, and ensures cooperation with the data governance partners, where necessary;
- actively communicates and promotes the value of proper data governance and management throughout the organisation, giving visibility to related projects and outcomes;
- calls on the managerial level, DGs/services that are not members of the IMSB, data governance partners and existing interservice groups to provide contributions to and/or to implement data policies and actions;
- provides advice on data policies to the Information Technology and Cybersecurity Board; and
- considers the human and financial resources necessary for developing, implementing and running data governance and data policies.

2.1.2. Managerial level — Data coordination groups

To enhance (interservice or interinstitutional) coordination on data matters and to facilitate the implementation of data policies in the Commission a number of committees, networks and communities have been established. These are represented in corporate data governance as data coordination groups.

A data coordination group is a formal or informal network, body, committee or community of practice that deals with data-related matters.

Each data coordination group has a clear mandate. The mandate is approved by the DGs/services requesting its formation. Some data coordination groups may be permanent while others may be active for a specific period of time. The IMSB will have the overview of existing groups and should be consulted on the establishment of new groups.

In general, a data coordination group may be organised around:

- a data (or reference data or metadata) policy domain (for which it may be assigned as domain owner), such as the Statistical Correspondents, the Commission interservice group on geographic information (COGI), the HR family data coordinators, and the interservice working group for calls and code management;
- a specific data or information policy, such as the document management officers (DMO), the data protection correspondents, the local security officers, the local intellectual property rights (IPR) and reuse officers or the knowledge management network;
- an activity linked to data or metadata standardisation, such as the interinstitutional metadata management committee (IMMC) and the interinstitutional metadata formats committee (IMFC);
- a data competency, for example in the form a data lab or community of practice, where people who share a common interest, competency or skill come together to share knowledge and expertise, and co-create solutions which can be used also by others in the Commission;

Data coordination groups:

- develop, implement, monitor and raise awareness on data policies in their data (management) domain and/or policy area, in alignment with corporate data governance and data policies;
- develop, implement and promote the use of data common metadata, standardised controlled vocabularies and data standards for their data domain or data asset;
- provide guidance, report and escalate to the strategic level, as needed;
- connect the business and policy needs with the data assets, identify and address data gaps and needs at corporate, interservice or local level;
- collaborate with other groups and data governance partners to contribute to or implement specific data policies and actions; and
- provide support and share knowledge with data owners, data stewards and other Commission staff working in their data domain and/or policy area, or competency.

2.1.3. Managerial level - Data governance boards

Data governance boards coordinate data-related matters within a (family of) DG(s)/service(s). The establishment of data governance boards is optional ([12]). Their responsibilities can be carried out by other existing roles or entities in a DG/service.

A data governance board should be chaired by a (senior) manager and may comprise representatives from directorates or units dealing with policy, data, IT, security, legal matters and procurement. This includes the Local Data Correspondent, the Local Security Officer, the Local Informatics Security Officer, the Data Protection Coordinator, the Document Management Officer, the Statistical Correspondent and the Information Resource Manager.

Data governance boards:

- support the proper implementation of corporate data governance and policies in their DGs/services;
- ensure that their DGs/services develop and implement local data policies, building upon existing corporate guidance;
- monitor the maturity of their DG/service, in terms of governance, management and use;
- promote data governance, management, use (including but not limited to reporting, analytics and artificial intelligence applications) and literacy at all levels within their DG/service; and
- advise on human and financial resources required to support the development, implementation and operation of data governance and data policies.

2.2. Individual roles

2.2.1. Managerial level — Local data correspondent

Local data correspondents (LDCs) serve as the single point of contact for data management in their DG/service. LDCs have been designated in all DGs/services. Their role can be delegated to or fulfilled by more than one person, depending on the size and needs of a DG/service. Local data governance teams may be set up by DGs/services to implement data governance and data policies, supporting the LDCs. They:

- contribute to the proper implementation of corporate data governance and data policies in their DGs/services;
- coordinate the development and local implementation of data policies, and align local data policies (if any) with the corporate ones;
- represent their DG/service in the LDCs network coordinated by the Secretariat-General;
- coordinate the update of their DG's/service's data assets (covering internal data, open data and licenced/purchased

([12]) Some DGs/services have already set up such boards.

data) in the Commission's data catalogue, following well-defined and, where possible, automated processes;
- remain informed, promote and, if relevant, participate in data standardisation activities related to the data assets of their DG/service;
- liaise with the strategic, managerial and operational levels, and coordinate with data governance partners;
- raise awareness of corporate and local data governance and data policies in data coordination groups, data stewards, data owners, data analysts, and data governance partners in their DG/service;
- raise awareness of the added value of proper data management in the management and staff of their DG/service;
- monitor and report progress to their senior management; and
- contribute to the (bi-)annual work programmes of the IMSB and action plans per area, as needed.

2.2.2. Managerial level — Data owner

Data owners are managers or staff who have been assigned by the strategic level as being responsible and accountable for a data domain or data asset. By data owners we refer to the business owners of data assets and not to the owners or providers of IT systems used for creating, collecting, storing, processing, disseminating or archiving data. Every data asset must have a designated data owner. Likewise, the owner(s) of a data domain may be designated ([13]).

Data owners can be supported by one or more data stewards. Data owners may coordinate with each other to share knowledge and resolve common problems. Data owners:
- are accountable for the quality of their data asset(s);
- are accountable, together with the owners of IT systems that store the data assets, for the proper implementation of corporate and local data policies for their data domain or data asset, notably access to data, data protection, information security and IT security ([14]);
- ensure that no unnecessary data access restrictions are put on their data asset(s), and that their data asset(s) are appropriately classified according to applicable information security and confidentiality policies;
- decide on data access requests received by data users, should they be responsible for data assets that are personal, restricted, confidential or EU (top) secret ([15]); and
- ensure that proper data licensing and reuse conditions accompany their data asset(s).

Commission information security policies introduce additional responsibilities for data owners ([16]).

2.2.3. Operational level — Data steward

Data stewards are subject matter experts in a data domain or data asset. They are selected and appointed by the data owner, whom they support. Data stewards may also be working outside the boundaries of their organisational entity. In some of their tasks, especially those linked to data modelling and communication of data requirements to IT, data stewards may collaborate with specialised data architects.

Data stewards:
- manage data assets coherently through their entire lifecycle (creation, collection, or acquisition, access, use, sharing, preservation or deletion) to maintain their quality, integrity and consistency and to avoid duplication;
- define, together with data owners, data quality rules which define the business requirements of what is considered as good quality data, and are responsible for maintaining quality;
- facilitate reuse and enable accessibility to the data asset(s) for which they are responsible (for internal or external purposes, e.g. open data), and also process internal or external data access requests;

([13]) A data domain is a logical representation of a data asset category that has been designated and named, for example HR, financial or procurement data. A data domain comprises one or more data assets.

([14]) Commission Decision (EU, Euratom) 2017/46.

([15]) Levels of information confidentiality as defined in the Security Notice information assessment and classification C(2019) 1903 final

([16]) Commission Decision (EU, Euratom) 2017/46.

- maintain quality metadata of the data asset(s) for which they are responsible on local and corporate data catalogues, including the EU Open Data Portal;
- support and share knowledge with data owners, other data stewards and Commission staff within data coordination groups or beyond;
- help define and implement data definitions, common metadata, standardised controlled vocabularies and data standards for the data domain or data asset(s) for which they are responsible;
- are aware of data protection, intellectual property and information security policies ([17]), liaising with the Data Protection Coordinator, the Local Security Officer and the Local Informatics Security Officer of their DG/service, as well as with the owners and/or providers of IT systems used for data creation, collection, access, use, sharing, preservation or deletion;
- collect business and policy needs and feedback on quality for the data asset(s) for which they are responsible, and address gaps; and
- organise and contribute to communication, promotion and outreach activities to increase awareness on and use of the data asset(s) for which they are responsible.

2.2.4. Operational level — Data user

This includes staff or contractors who are (re)using data assets to develop products such as management reports, business intelligence dashboards, analytical and artificial intelligence models. There are different types of data users including:

- policy staff who use data and data products for policy- and decision-making;
- data analysts who work with data, e.g. analyse, integrate or visualise it, to extract insights relevant for a policy domain or internal operation;
- data or information architects who define conceptual, logical or physical models and architectures for representing, managing, exchanging or using data;
- data engineers who query, cleanse, integrate and prepare data for analysis, and define data architectures;
- data scientists who analyse (big) data using statistical analysis, advanced analytical techniques, including data mining and information retrieval, machine learning and artificial intelligence in order to interpret data, identify trends, figures and other relevant information; and
- software developers namely Commission staff or contractors who implement data architectures, build data management pipelines, develop interfaces to enable the access to data in IT systems and implement data policies in IT systems.

Data users have no formal responsibilities, but they should, where relevant:

- be aware of and respect in their daily work corporate and local data policies, notably those related to quality, interoperability, data protection, intellectual property and information security;
- if necessary, become accredited prior to working with Commission data assets;
- inform the data owner and the data steward of the intended use of the data asset;
- consult the Commission's data catalogue for reusable data assets;
- report to the data steward or data owner any issues that they may discover regarding the quality, reliability and integrity of the data asset;
- communicate to data stewards the quality requirements based on how they intend to use the data asset, and inform the data steward of quality issues discovered; and
- use corporate platforms and tools for data management and analysis, such as the Commission's data platform.

Commission information security policies introduce additional responsibilities for data users (referred to as 'end users').

([17]) In the case of document management, there is a partial overlap between the role of data steward and that of document management correspondent.

2.3. Data governance partners

This includes roles that may be involved in data governance and policies depending on the type of activity, such as:

- The Information Technology and Cybersecurity Board (ITCB) which ensures that resources and investments in IT are used efficiently and that business needs are supported by efficient, secure and resilient communication and information systems, in compliance with personal data protection principles. It oversees the implementation of the Commission's digital strategy. The Secretariat-General corporate governance team ensures alignment between data governance and data policies and the work of the ITCB.
- The Commission's Data Protection Officer, who may be consulted on topics concerning the management and processing of personal data. The data protection correspondents' network, a type of a data coordination group, provides support to DGs/services on this.
- The local security officers, who should be consulted on topics concerning information security; and the local informatics security officers / information resource managers who should be consulted on topics concerning the security of information systems and of operational procedures.
- The Commission's Central IP Service that should be consulted on questions regarding licensing of data and information, reuse conditions, and other IP-related matters. The IPR correspondents' network, a type of a data coordination group, provides support to DGs/services on this ([18]).
- The Commission's Legal Service that should be consulted on any legal issue concerning data-related clauses in contracts and other legal documents.
- ESTAT that should be consulted on the development and production of other (than European) statistics and on the usage of geospatial information in the Commission.
- The DMO network that should be consulted on topics concerning the medium- and long-term preservation of data in information systems.

2.4. Support roles

2.4.1. Secretariat-General corporate governance team

There is a corporate coordination team for data and information management in the Secretariat-General that reports to the chair of the IMSB. This team:

- supports the activities of IMSB and the LDCs network by overseeing the implementation of corporate data governance and data policies, actively monitoring ongoing activities and coordinating communication;
- provides the secretariat for the LDCs network and the IMSB, and coordinates the communication with the data governance partners on corporate matters; and
- collects feedback from the strategic, managerial and operational levels for maintaining, aligning and improving corporate data policies.

([18]) Comprised of local IP and reuse officers who: provide initial support to enquiries related to IPR and reuse, raise IP awareness within their DG/service help monitor IPR compliance and facilitate the identification of IPR-related risks in the daily work of their DG/service; and collaborate with the Central IP Service in establishing and maintaining the Dynamic IP Inventory and Management System for assets under the responsibility of their DG/service.

3. DATA POLICIES

This section introduces a set of core principles which form the guiding framework for data policies in the Commission ([19]). These principles may be reviewed or extended while the implementation of data governance and data policies in the Commission is progressing.

3.1. Data management

This section introduces corporate principles for managing data assets in the Commission which apply across data domains.

Different domains, including, but not limited to, documents ([20]), procurement and grants data, financial data, statistical data ([21]), geospatial information, health data, trade data, audit and compliance data, research data and security data, have specific policies, detailed processes and guidance, which are complementary to what is defined in this document.

Corporate principles for the creation, collection or acquisition of data assets:

(1) Data assets will be created or collected only once. DGs/services first explore whether necessary data or information already exists before proceeding with its creation, collection or acquisition (including transfer) from individual data subjects, Member States or any other third party. Likewise, data assets will be reported only once. Should a new inventory be launched, it must take into account existing inventories.

(2) In case of acquisition of data assets against payment of fees ([22]), where this is reasonable and proportionate from a financial point of view, contractual arrangements should be made to ensure that the data asset licensed to a DG/service can be made available, under agreed conditions, to other DGs/services and possibly other EU institutions and agencies. The resulting additional licence fees should in such cases be weighed against a dataset's importance in the policymaking context.

(3) In case of licences with non-commercial providers such as NGOs, public administrations or universities, the lead Commission service acquiring the data asset should try to obtain rights of use which cover other DGs/services and possibly other EU institutions and agencies, in particular if the licence is cost-free.

(4) The needs of the intended acquisitions will be coordinated by the lead Commission service, helping DGs/services to leverage their position in the negotiations, in particular with commercial data providers. Such coordination would also allow services to benefit from synergies and avoid redundant or double acquisitions. A corporate process for data acquisitions will be specified.

([19]) Whenever the implementation of a principle calls for further action or future work, 'will' is used in the formulation of the principle.

([20]) https://ec.europa.eu/info/about-european-commission/service-standards-and-principles/transparency/freedom-information/access-documents/information-and-document-management/archival-policy/document-management-and-archival-policy_en

([21]) Commission Decision 2012/504/EU defining the role and responsibilities of ESTAT within the internal organisation of the Commission, as regards the development, production and dissemination of statistics. With regard to statistical data and in particular European statistics as defined in Regulation (EC) No 223/2009 and produced by Eurostat, the data governance and data policies set out in this document complement, where appropriate, the existing legal or other formal rules applicable to such data, but do not collide with them.

([22]) There are two distinct cases of acquired data assets (data acquisitions): a) where there is transfer of ownership and thus the Commission obtains the IP ownership of the data, e.g. via a service contract (rare), or b) where data is merely licensed to the Commission and thus the Commission does not obtain the IP ownership but only the rights to use the data (more common).

(5) Contractors or beneficiaries of projects which are (co-)funded by the Commission and support, among others, evaluation, impact assessments and reporting obligations, e.g. through procurement or grants, and which entail data collection and processing, may submit a data management plan in line with the Better Regulation requirements and other local requirements ([23]).

(6) Harmonised terms and conditions for data management to be included in contracts, grants and other agreements will be created. These can be in the form of standard contract clauses or templates. A data coordination group will be set up for this purpose and will include the Commission's Central IP Service and the Legal Service.

Corporate principles for accessing, using and sharing data assets:

(7) A Commission-wide list of data domains ([24]) will be created on the basis of the Commission's data inventory. Owners will be defined for each data domain and each data asset. A data asset may be categorised under several data domains mentioned earlier in this section. Certain data assets within those domains may constitute Commission master data or reference data and will need to be managed accordingly.

(8) Corporate tools, platforms and services for creating, managing and visualising data, such as the Commission's data platform, will be provided and used by DGs/services, unless specific requirements are not met. As per the Commission's digital strategy, the design of new IT tools, platforms and services will take into account data policies and reuse considerations.

(9) Duplication of reference and master data should be avoided.

(10) Priority will be given to the reuse of master data, for example staff data, procurement data, vendor data or financial data, across business processes and systems. The IMSB will endorse a reference list of Commission master data. Owners of master data will coordinate their management.

(11) The 'need to share' (or 'share by default') principle is the norm. Data assets will be shared by default as widely as possible across the Commission. Restrictions need only be applied on the basis of the information confidentiality levels defined in C(2019) 1903, data protection, notably linked to the legal principle of purpose limitation, intellectual property or other legal provisions. Data assets used for policy proposals, and not falling under any of the aforementioned exceptions can be shared by default with the European Parliament and the Council in the context of the Interinstitutional Agreement on Better Law-Making when these documents are released.

(12) Commission data assets are made publically available under the conditions of free, full, open and timely access via the EU Open Data Portal to facilitate reuse for commercial and non-commercial purposes, within the framework defined by relevant legislation and policies. Data assets are made open under the Creative Commons type BY licence. As per the Better Regulation guidelines ([25]), all data assets collected and used for the preparation of impact assessments and

([23]) As a minimum, a data management plan should define: what data will be collected, processed and/or generated; which methodology and standards will be applied; the handling and processing of data during and after the end of the project, and whether data will be shared with third parties.

([24]) EuroVoc and the data theme vocabulary, managed by the OP, can be used as a basis.

([25]) Better Regulation guidelines.

related Commission proposals and implementing plans should be published on the EU Open Data Portal ([26]), unless they include confidential elements or are otherwise protected by rules on data protection or intellectual property rights of third parties.

(13) Exceptions to making data assets open may be imposed for the reasons referred to in the Commission Decision 2011/833/EU, in Regulation (EC) No 1049/2001 and Directive (EU) 2019/1024. These reasons include: protecting public interest as regards: security, defence and military matters, international relations and the EU's financial, monetary or economic policy; privacy and the integrity of the individual; the commercial interests of natural or legal persons, intellectual property, court proceedings and legal advice, and for the purpose of inspections, investigations and audits; restricted, confidential or EU (top) secret data assets; personal and private data.

(14) The Commission will maintain (internally) a catalogue of data assets, starting with those of corporate or interservice interest. Data stewards, in coordination with data owners and local data correspondents, will register the data assets of their DGs/services in the Commission's data catalogue, and will keep the metadata records up to date.

(15) Data stewards will accompany each data asset with standardised metadata ([27]) which provide, as a minimum, information on the content, the owner and the way to access the data asset. Such metadata facilitate findability, access and (re)use of data assets. For high-value data assets accessible via the data catalogue, the metadata may also contain information about reuse conditions, applicable IPR and the legal basis (if any).

(16) Metadata of data assets will be shared internally by default, even if access to and use of the underlying data are restricted or confidential, except if the metadata itself is considered as restricted or confidential.

(17) Access to sensitive non-classified and classified data assets remains restricted. Processes for requesting access to such data assets will be defined by the data owners together with the local security officers or, for personal data, the data protection coordinators. This information will be made available in the Commission's data catalogue. Data users requesting access must use the data assets only as required for the performance of the purpose defined in the request ('need to know' basis and purpose limitation).

(18) Updates of data assets are made by data stewards or (IT) staff.

(19) Until specific guidelines are made available, DGs/services can store and analyse only publically available ([28]) data assets on public cloud platforms. For all other cases, the opinion of the local security officers, or, when personal data is involved, the data protection coordinators, of the DG/service in question should be requested.

(20) Unless covered by legal provisions under specific contracts, contractors working on behalf of DGs/services on projects that entail access to and use of Commission data assets on the Commission's data platform will need to be accredited, e.g. by signing a non-disclosure agreement or taking any other action deemed necessary, before access is granted.

(21) In collaboration with the IMSB, DGs/services will periodically review and optimise the use of data platforms, services and data acquisition activities to increase synergies and efficiencies, improve performance and increase use.

([26]) http://data.europa.eu/euodp/en/home

([27]) There are different standardised formats for metadata, often defined in legislation. Examples include the data catalogue application profile, SDMX for statistical data, and the INSPIRE metadata specification for geospatial data.

([28]) As defined in https://ec.europa.eu/transparency/regdoc/rep/3/2019/EN/C-2019-1903-F1-EN-MAIN-PART-1.PDF.

Corporate principles for preserving and deleting data assets:

(22) Every data asset has a primary (authentic) source. Any changes or updates to the primary source will be pushed to known secondary sources, such as in the case of master data management solutions, and vice versa, preferably using automations. Corporate IT systems and data platforms will be used for data storage and preservation, especially for master data and reference data.

(23) Data assets are stored, preserved or deleted in accordance with applicable legislation, regulations and guidelines. The corporate digital preservation strategy will provide guidelines and processes ([29]).

3.2. Data interoperability and standards

(24) Data assets will be identified using standardised persistent identifiers ([30]). Identifiers will also be used for referencing data assets, e.g. see Eurostat's guidelines for referencing statistical data, the document identifiers assigned by the OP and the guidelines for identifiers of geospatial data compiled by the interservice group on geographical information. Identifier schemes and collections will be managed by specialised data coordination groups, either at the corporate level, such as the Interinstitutional uniform resource identifier committee, or within a data domain.

(25) Data assets will be created, collected or acquired, managed and made available using data standards or commonly agreed specifications endorsed by the responsible data coordination group and/or the IMSB. They may be developed by a recognised standardisation organisation. Their reliability, change and release management is guaranteed by their owner. Data standards or commonly agreed specifications for representing data assets may be either generic or specific to a policy or data domain.

(26) DGs/services will use a business glossary, at corporate level, comprising definitions about data-related terms common across the Commission.

(27) IT systems hosting data assets provide access in standard or commonly agreed machine-readable formats to enable data exchange, distribution and portability (when needed), and implement corporate data governance roles and policies (as required).

(28) DGs/services must give priority to the reuse of master data and common reference data ([31]) in Commission IT systems to simplify data exchange and sharing. Owners of reference data will coordinate their management and ensure reliability in dedicated data coordination groups.

(29) Several data coordination groups in the Commission work on agreements on corporate data standards and the use of common reference data. The strategic and managerial levels actively seek coordination and look for opportunities for harmonisation or standardisation in domains and processes that have not been covered yet.

([29]) The long-term preservation of some data assets falls under the regulatory framework of documents and archives management (e-Domec). The long-term preservation of other types of data assets might fall under other regulatory frameworks, e.g. OP for publications, websites, etc.

([30]) A persistent identifier is a long-lasting reference to a document, file, web page, data asset, or physical or digital object.

([31]) Examples of reusable Commission reference data are available on EU Vocabularies.

3.3. Data quality

(30) Managing data effectively means managing quality, which is a collective responsibility. The main attributes of data quality include: accuracy, completeness, consistency, uniqueness, integrity, timeliness, provenance and data collection methods. Data quality is relative to the level of the reliability of information or the acceptable error rate for serving an intended purpose, including policymaking or internal operations. Data stewards are primarily responsible for managing and ensuring quality. Other roles at all levels of governance, notably data users, provide feedback to data stewards on quality matters and possible improvements.

(31) The quality of data assets should be defined, measurable and clearly indicated, particularly for high-value data assets. Data quality plans, specifying quality indicators and quality remediation actions, can help data stewards manage quality.

(32) Quality will be measured regularly and managed consistently across the data asset's lifecycle by data owners and data stewards. Domain-specific quality frameworks are required. Such frameworks have already been developed in certain data domains, for example the Quality Framework of the European Statistical System consisting of the European statistics Code of Practice and the Quality Assurance Framework, and the Reference Quality Framework applicable in DGs/services managing statistics and data, both developed by ESTAT ([32]), and the data quality campaigns of the DG for Development Aid and Cooperation (DEVCO).

(33) High-value data assets, such as master data and reference data, will have a change-and release-management plan, which should at a minimum indicate how feedback, e.g. on fixes and requests for change, is collected, what type of releases exists, and the planning for making new releases (versions) of the data asset available. Data owners will make available processes or mechanisms for allowing users to report quality issues and give feedback.

(34) Data quality rules will be embedded in IT systems and processes to reduce errors and improve quality from the point of data entry or creation through to data processing (quality by design).

3.4. Data protection and information security

(35) With regards to data protection, DGs/services have to comply with the legal requirements and operational obligations of Regulation (EU) 2018/1725. In addition, the Commission's Data Protection Officer has issued implementation principles and guidelines. The European Data Protection Supervisor (EDPS) has also issued guidance ([33]) on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725.

(36) With regards to information security, please refer to the principles and implementing guidelines provided by the Commission's information security team.

([32]) https://ec.europa.eu/eurostat/web/quality/overview

([33]) EDPS Guidelines on the concepts of controller, processor and joint controllership under Regulation (EU) 2018/1725.