Regulatory Reporting Community

Directorate-General for Informatics

D2 – Interoperability

The Importance of Metadata for Regulatory Reporting

Join our Regulatory Reporting Community of Practice here.

This study was carried out by Wavestone for the ISA[2] programme by:

Authors:

Patricia BACHMAIER

Esther BLEYS

Barbora KUDZMANAITE

Alessandro ZAMBONI

Disclaimer

Table of Contents

Introduction

The continuous increase of regulatory reporting requirements in the EU legislation, combined with the many stakeholders involved in the reporting process, contribute to an ever-growing amount of data. This is not only due to the pressing information needs coming from the public entities, but it is also the result of a lack of communication between the stakeholders involved: there is an information gap regarding where each data flow is going, and which data can be reused.

This significant amount of data exchanged has turned regulatory reporting into a complex process with numerous dependencies, difficult to sustain for the reporting entities and burdensome to manage for the regulatory entity. Hence, the absence of data management, coordination and governance across EU bodies is a missed opportunity for all the parties involved, in terms of effort, time and the taxpayer's money.

Indeed, **championing data governance** is already high up in the European Commission (EC)'s policy agenda. One of the key priorities within data governance is to improve metadata, as it is a precondition for data reuse. Additionally, such governance of (meta)data would be beneficial to the growing initiatives[1] around the creation of **data catalogues** (see the definition in section Terms and definitions) as it would bring more harmonisation and standardisation among metadata.

The Directive on open data and the reuse of public sector information[2] offers a clear example of the type of efforts the EC is carrying out in this line. The Directive encourages public bodies to make relevant data and metadata publicly available in open and machine-readable formats to ensure interoperability and ultimately facilitate reuse.

This paper aims to shed light on the importance of metadata, to indicate what is preventing it from being a core part of the regulatory reporting process, and to point towards the next areas to be investigated to make metadata a reality, paired with some practical examples of actions.

After a thorough desk research and consultations with colleagues from the Directorate-General for Environment (DG ENV) and Joint Research Centre (JRC), the final goal of this paper is, of course, to convince you, the reader, to contribute to the creation of metadata in whichever way possible, by inspiring the Regulatory Reporting Community of Practice members to continue investigating the different ways of making it a reality. This paper is published by the Regulatory Reporting Community of Practice, which you can read more about and join here.

The paper includes the following sections:

- **Section 1** – This section provides a definition of the concept of metadata and explains the different stakeholders involved in the provision of metadata. It also provides some context on the specific cases where metadata impacts the regulatory reporting process.

---

[1] Such as the European Data Portal (https://data.europa.eu/data/catalogues?locale=en) or the JRC Data Catalogue (https://data.jrc.ec.europa.eu/).

[2] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. Available at : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG

- **Section 2** – This section is dedicated to demonstrating the benefits of using metadata in regulatory reporting.
- **Section 3** – This section focuses on the challenges that hinder the management of metadata.
- **Section 4** – Finally, this section offers a set of recommendations for the creation and management of metadata in the EC. These recommendations highlight potential future work for the community to explore itself.

1. What is metadata?

Notably, while there is still no streamlined vocabulary to refer to metadata and related concepts within the EC landscape (to be further discussed in Section 3), there are existing definitions and models, such as the one provided by the Data Advisory Group and DCAT-AP that should be widely reused. Hence, for the purpose of this issue paper, metadata is referred to as[3]:

> *Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data. In short, metadata provides information about other information to make it easier to find and manage. Basic metadata elements include, for example, the author, title and year of publication*

An example of metadata in the music world could be the following: the name of song, the artist, the album, the year of publication, the type of music, etc.

Moreover, metadata provides information that makes sense of the three following types of information[4]:

- Data (e.g. documents, images, data sets);
- Concepts (e.g. classification schemes); and
- Real-world entities (e.g. people, organisations, places, paintings, products).

In the regulatory reporting stages (see the definition in section Terms and definitions), all three types of information are at use: a reporting entity will notify the name of its organisation (real-world entity), it will potentially share its data through an Excel sheet, a PDF (data) or submit it via an online system, and the regulatory entity may use a classification scheme (concepts) to structure the data according to its analysis needs.

Therefore, in the context of regulatory reporting requirements, metadata are critical to the way the data collected through regulatory reporting are managed, organised and used. When created and handled properly, metadata serve the clarity and consistency of collected information. Metadata also facilitate the discovery of relevant information and the search and retrieval of resources. However, in practice, the handling of metadata in regulatory reporting requirements is not as straightforward as it could be.

---

[3] Data Advisory Service (n.a) How to create metadata?, EU Open Data Community, Available at: https://webgate.ec.europa.eu/fpfis/wikis/display/EUODDVC/How+to+create+metadata
[4] Information retrieved from the European Data Portal, which is available at: https://data.europa.eu/en

There are already some tools which were developed within the EC, such as Reportnet 3.0[5], DECLARE[6], KOEL[7] and RCD[8], supporting the regulatory reporting process[9], with very positive impact on the quality of metadata created. However, while they are a first step towards a better management, use and reuse of regulatory reporting requirements metadata, these tools, and consequently the metadata they create, are not interoperable, which hinder their reuse.

On this basis, the importance of metadata for regulatory reporting is demonstrated and the challenges their creation and management are facing are further elaborated in the next sections of this paper (Sections 2 and 3).

---

[5] Reportnet 3.0 is a centralised e-Reporting platform for reporting environmental and climate data to the EEA. It aims to simplify and streamline the data flow steps across all environmental domains. It provides a framework of data standards, applications and interoperability mechanisms to exchange and share information within and between information systems. It is available here: https://reportnet.europa.eu/

[6] DECLARE is an instance of DSC (a web-based system supporting reporting obligations and data collection) and is used by DG ENV to support the reporting obligations and data collection for policies which are out of scope of EEA or other DG ENV partners IT solutions. It enables national competent authorities and economic operators to submit required data by the EU Regulation.

[7] The Knowledge Online on European Legislation (KOEL) application, created by DG FISMA, is a web-based application that provides support for the setting of regulatory requirements and contains information on all the existing regulatory reporting requirements across the financial acquis. Its purpose is to provide the necessary collection of all relevant legislations containing regulatory reporting obligations into a single repository with the aim to identify gaps, overlaps and inconsistencies in regulatory requirements and avoid duplication.

[8] The Regulatory Concept Dictionary (RCD) application, developed by DG FISMA, provides support for a greater standardisation of EU-level supervisory reporting requirements by automatically creating a glossary of concepts defined in all legal texts within the domain of DG FISMA and by setting up a dictionary of reporting obligations contained in these texts.

[9] For more information about these four tools, we invite you to have a look at the **Reuse Guide of IT tools supporting the regulatory reporting process within the European Commission**, which is available here on the Community page.

2. Why is metadata important to regulatory reporting?

Due to the nature of regulatory reporting which entails the collection of key information, metadata can offer many advantages to the stakeholders that take part in the process. Below are the key benefits:

Improvement of the harvested data's interoperability

Metadata contributes to the interoperability of the data collected, as it ensures its consistency and portability. By using a common dataset or data model (provided through the metadata), information is easily shared and cross-checked with different documents format (for instance, when comparing data subtracted from Word documents with data coming from Excel documents).

Support of the accessibility, findability and reuse of the data

By creating a common framework for the data (that is to say, metadata), data can be easily structured, found, shared and reused. In this way regulatory entities can avoid harvesting a relevant amount of data, saving efforts, time and financial resources to all parties involved. Metadata also improves the accessibility and findability of data, by creating awareness of the existing data, how it is structured, and hence how reusable it is. Notably, the Corporate Reference Data Management[10] and the Data Management Guidelines for DG GROW emphasise that the availability and accuracy of metadata is an important feature of data assets to facilitate data sharing[11]. Metadata provides information on the attributes of the data, i.e. the purpose for which the data is collected and its reuse conditions (or licensing). Data reuse depends not only on the content of the data, but also on its quality, granularity, and license.

Transparency

By making data accessible and findable, metadata also enhances the transparency of entities collecting data. Metadata documents and makes accessible information about the data, allowing users to verify how the data is used and managed, ensuring that it serves the purpose for which the data was originally collected.

Coherent regulatory reporting requirements

Metadata ensures that the regulatory reporting requirements are coherent with each other, by avoiding duplication or overlapping requirements in EU legislation (which would lead to unnecessary parallel data flows in different regulatory entities). In practice, a policy officer would be able to swiftly check if some specific metadata is already available and reuse it. In the case of open data, a policy officer would even do so without having to reach out to other departments or institutions.

---

[10] See Corporate Reference Data Management policy in the European Commission
https://op.europa.eu/en/web/eu-vocabularies/corporate-reference-data-management
[11] See the Data Management Guidelines for DG GROW and Corporate Reference Data Management policy in the European Commission.

Relevant and up-to-date data feed to the policy cycle

Metadata helps regulatory reporting provide relevant and up-to-date data to the policy cycle, by facilitating the data sharing and monitoring without requiring extra burden on regulatory entities: the metadata would be just there, ready to be consulted when key decisions are taking place in the European Parliament or the EC, for instance.

3. What are the challenges of metadata creation and management in the EC?

There are a number of key challenges to metadata creation, management, and reuse. For the purpose of this paper, they have been categorised into the four layers of interoperability[12] as put forward by the European Interoperability Framework (EIF)[13], due to data interoperability dependence on metadata. Each group of challenges is explored in more detail in this section.

Figure 1 summarises the four layers of challenges to the creation, management and reuse of metadata in the European Commission.

Figure 1 Summary of the challenges to metadata creation, management and reuse in the EC

Legal challenges
- The GDPR and confidential reporting requirements need to be respected.
- The collection of metadata is not always mandated in primary and/or secondary legislation.
- The terms used when designing a legislation and setting regulatory requirements are not consistent.

Organisational challenges
- Metadata is not a priority in EU institutions.
- Managerial commitment is missing.
- There is not enough coordination among entities.
- Sharing and reusing data catalogues is difficult.
- There is a lack of overview of reporting requirements.
- There is a lack of budget and human resources.
- A mission creep can make data management cumbersome.
- Expertise and training are needed.

Semantic challenges
- The collection of data needs to be relevant and coherent.
- It is difficult to understand the nuance between metadata for discovery and metadata for access and use.
- High (meta)data quality is difficult to achieve.
- There is a lack of taxonomy/harmonised vocabulary when setting regulatory requirements.

Technical challenges
- The poor quality of metadata hinder its reuse.
- There is a lack of reusable (semi) automated tools.

Legal challenges:

√ **Varying privacy levels and data sovereignty:** while metadata, by definition, should always be publicly available[14], the main challenge related to the legal aspect is the need to respect the General Data Protection Regulation[15] (GDPR) and confidential reporting requirements. In this regard, metadata needs to be ready to address the different levels of privacy complexity that come with the reuse of data collected from very different types of stakeholders. Metadata

---

[12] The four layers of interoperability can be described as follow:
  - Legal challenges relate to the legal provisions creating the legal basis for the creation and management of metadata.
  - Organisational challenges refer to the challenges linked to the organisational structure of the EC and its Agencies.
  - Semantic challenges have to do with aspects of semantic properties, concepts and terminology.
  - Operational challenges are the challenges identified at the technical level that hinder the creation and management of metadata.

[13] The EIF contains 43 recommendations to public administrations at national, regional and local level on how to develop and implement interoperable public services. The EIF is crucial to ensure that systems are compatible and interoperable with each other, and that they support the current and future business needs of users. More information is available at: https://ec.europa.eu/isa2/eif_en

[14] Corporate Reference Data Management policy in the European Commission. Section 3.1.17.

[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), available here: https://eur-lex.europa.eu/eli/reg/2016/679/oj

provides information on the reuse conditions of the data and should specify any derogation to data sharing.

√ **Different sources of request:** the collection of metadata is not always mandated in primary and/or secondary legislation. Therefore, the gathering of metadata depends on a voluntary approach, which results on a lack of harmonisation in the formatting of metadata since it is not a formalised process.

√ **Semantic legal issue:** there is a lack of consistency in the terms used when designing a legislation and setting regulatory requirements. Therefore, different regulatory reporting provisions refer to the same object but with different names/wordings. There are also cases where the terms employed in the regulatory requirements do not correspond to the typical terminology used in the field. This can thus lead to confusion and mismatches with regard to what exactly has to be reported.

Organisational challenges:

√ **Absence of metadata:** metadata is not a priority across the European institutions, which is mainly caused by the poor knowledge of its usefulness and benefits. Hence, there is no metadata to begin with, because efforts have been to put to work on it.

√ **Lack of managerial commitment** to the creation and maintenance of metadata, with all its consequences, e,g. budget, more human resources, motivation to review and maintain a metadata strategy or program, etc. Senior management is usually not aware of the importance of metadata, therefore it does not prioritise it, reward its creation, and may even consider it a cost, discouraging its creation. Moreover, the upper hierarchy is not always aligned with the practical needs of the metadata creation, specifically its costs, or does not recognise its importance, hence the challenges on investments are very common. This also leads to a situation where the time invested in data creation and management is not rewarded, reducing the motivation to continue these tasks.

√ **Lack of coordination:** work is generally carried out in silos across regulatory entities and the absence of backward compatibility between entities. It also results into to untapped potential of synergies between services. Some entities set requirements without checking which data is already being collected in other organisations, and how it was collected (under with which data sets, periodicity, granularity, format, etc.).

√ **Challenges to share and reuse data catalogues:** this aforementioned lack of coordination has other ramifications, specifically the difficulties in sharing and reusing data catalogues and the few integration of IT tools into reporting requirements.

√ **Lack of an overview of regulatory reporting requirements** across the extensive EU law impedes regulatory and reporting parties from having an accurate image of all the existing parallel data flows, and hence they are not aware of the size of the burden it is creating. As a result, it may prevent them from prioritising a solution, such as metadata, that they are not aware they need: "we don't know what we don't know".

√ **Lack of human resources allocated** can also be a challenge for initiating metadata creation work, as well as metadata management. The lack of a coordinated data and metadata

management process hinders the right allocation of the human resources needed to maintain metadata. For instance, when there is an employee turnover, the handover of this task is often left aside or even forgotten about. Would the organisation follow streamlined methods, turnover would not be an issue.

√ **Mission creep** can make metadata management cumbersome: new requirements or elements have to be added, while others have to be maintained, making metadata management more time-consuming for the people who have to fill in the information. If metadata records are extended for purposes other than those initially assigned to the information system, maintaining an overview or designing an information management strategy, these records can become unmanageable.

√ **Need for expertise and training**: high metadata quality needs to be developed by professionals who understand how to build it and manage it. when metadata sets developed by non-professional, often not aware of existing standards (e.g. for code-lists), this can lead to a waste of resources as they would design something that already exists. This is even more relevant for dynamic metadata which evolves constantly and require regular updates. Metadata management is not considered a continuous task. Emerging technologies, such as Artificial Intelligence, cannot automate fully the creation and maintenance of metadata, it requires dedicated projects and employees.

Semantic challenges

This section refers to challenges related to the semantic properties, concepts and terminology of metadata:

√ **Collection of coherent and relevant data** needed to create useful metadata. If the collection of data is in place, but it is not coherent or includes irrelevant data, it would lead to incoherent metadata across datasets, thus requiring great efforts to transform it into the right datasets.

√ **Difference between metadata for discovery** (describing the data) **and metadata for access and use** (making the data accessible and understandable). What should be described and with which level of granularity is often unclear, and also depends on whether the data is published for proof of work or for reuse. Also, metadata for access and use requires much more detail than metadata for discovery, often preventing metadata reuse.

√ **High quality metadata**: it is challenging to achieve as it relies on semantics and format, which are often under considered when the data is published for reuse, due to the absence of well-known corporate guidelines to manage metadata (naming rules, design decisions, definition rules, etc.). In some cases, documentation may be necessary to understand the metadata, as it may contain abbreviations, codes, etc.

√ **Lack of taxonomy/harmonised vocabulary**: the lack of consistency in the terms used when setting regulatory requirements leads to the fact that different regulatory reporting provisions refer to the same object but with different names/wordings. Consequently, the same object is reported differently.

Technical challenges

- √ **Poor quality of metadata available for reuse**. The available metadata has technical challenges for reuse, such as non-working links, complicated databases, etc. Poor quality of (meta)data can come from the fact that the person in charge of entering and/or maintaining the information might not see the purpose of doing so, and/or might not have all information required to fill the information asked. Consequently, this low quality creates a lack of trust in the database and can discourage people from using the (meta)data. Links between different systems have to be continuously maintained, because when broken, people will lose interest in reusing the data as they will not be able to find it.

- √ **Lack of reusable (semi) automated tools:** the fact that the metadata often has to be entered and maintained manually into the relevant systems adds time constraint to employees dealing with metadata maintenance. The use of automated tools would allow or facilitate the maintenance and management of metadata. These tools are expensive, hence, one of the barriers to investing in them is notably the lack of budget, itself linked to the lack of managerial commitment towards metadata management. They also require many components to be fully usable, complicating their deployment.

4. How to promote the creation and management of metadata in the EC?

The following section addresses three different topics that should be investigated through different workstreams: they complement each other and need to be tackled as a whole.

4.1 Looking for believers: raising awareness

To create and maintain metadata across public European entities **raising awareness** around the importance of metadata and its usefulness is the first step. All involved stakeholders need to be onboard with its creation and reuse, starting with an in-house awareness campaign, before going to the Member States and other less central stakeholders.

Specific steps:

- √ **Training** seems a relevant way to actively engage the appropriate regulatory entities;
- √ The use and sharing of **success stories** (in the EC or the Member States)**, good practices, and testimonials** of people reusing metadata could also be ways of demonstrating the usefulness and benefits of metadata;
- √ **Involving key senior management** in metadata processes organisation will be key in this workstream, in order to ensure human resources are allocated to the metadata related tasks, and to guarantee that metadata becomes a priority within the regulatory entities' agenda. While this is being addressed by the **EC Corporate Data Strategy**, more can still be done to increase awareness beyond the local data correspondents;
- √ Using **costs-benefits analysis** to demonstrate the benefits of metadata creation and management against their costs would be a mean to convince the senior management to further invest in them;
- √ **Incentives for those in charge of creating metadata** should also be generated to reward their efforts;
- √ **Ensure that data acquisition contracts include metadata provision** by providing tools for officers to be able to draft terms of reference which comprise metadata as an essential element and require the inclusion of the **link to the legislation** that calls for this metadata.
- √ **Assign responsibilities and accountability** for the management of taxonomies within each organisation to warrant its maintenance;
- √ **Encourage legal and IT departments to collaborate** when setting regulatory reporting requirements, to exploit the potential of technologies in the collection of data.
- √ Ensure that **EC services are reminded** on a regulatory basis to **update** their metadata and inventories (e.g. by the SecGen).

4.2 Finding common ground: relevant taxonomies

The next step to make metadata a reality is to create the framework (or structure, as preferred) that will form the skeleton around which the data will be distributed upon. More concretely, the regulatory stakeholders should create, or reuse, common taxonomies across the European Commission and other European institutions. The creation and use of a common taxonomy will have a relevant role to play in guiding the European Commission and the European Parliament in the setting of regulatory reporting, by reusing the same terminology (such as common vocabularies and ontologies[16]), allowing to link specific legal provisions and datasets across all stakeholders. Metadata could be used as a link between the data that has been reported and the regulatory provision that requires the reporting of the data.

Specific steps:

√ **Identify existing common taxonomies** across the European Commission and other European institutions before creating a new one. There are already some existing conceptual frameworks (DCAT-AP[17], for instance) that should be considered for its reuse in case it can save efforts and enhance metadata's quality. Data and metadata should be structured using encodings that are widely accepted in the target user community and aligned with organisational needs and observing methods, with preference given to reference data[18].

DCAT-AP offers collections of metadata about datasets or data services, it also provides a common specification for describing public sector datasets in Europe to enable the exchange of descriptions of datasets among data portals. Interestingly, this specification offers extensions for the identification of some specific type of metadata, such as geospatial metadata or statistical metadata, showing the potential to create more relevant extensions depending on the needs identified across stakeholders.

√ **Definition of a common ontology and use of common semantic standards:** For this to become a reality, there is a need to work on an agreement on the use of a common ontology/semantic standards and increase the visibility and reuse of existing solutions. Concepts should be managed jointly, and their evolution closely monitored within the EC. While it is not possible to create an all-encompassing ontology for ell European institutions, it would be useful to have common ontologies per big domains, and to develop mechanisms in place to translate them into other ontologies. Moreover, open data should be created and promoted as much as possible, to connect and link data from heterogeneous data sources by relying on a common metadata and common semantics.

An example of this work can be found on a DG SANTE's project, developed jointly with DG ECHO (Directorate-General for European Civil Protection and Humanitarian Aid Operations) and Member States: the project focuses on streamlining the terminologies and semantics used in different systems to make matching of data easier across systems, with the

---

[16] See Corporate Reference Data Management policy in the European Commission.
https://op.europa.eu/en/web/eu-vocabularies/corporate-reference-data-management
[17] https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en
[18] See the Data Management Guidelines for DG GROW.

project partners share the common terminologies and mechanisms they use to streamline reporting data.

Moreover, the SEMIC Action[19] of the EC aims to improve semantic interoperability in European government systems, by offering unified semantics that allow for the reuse of data, enhancing in this way, metadata creation. This resource should be investigated and reused if relevant.

√ **Identify and promote common models** that fits the needs of the regulatory entities. There are already existing models, catalogues, code-lists and data sets ready to be used in the regulatory landscape; a key step should be the investigation of these resources to identify which ones are most relevant to use and promote its uptake.

√ **Create an extension of DCAT-AP for Regulatory Reporting metadata**. Extensions of DCAT-AP have already been done for the representation of geographic metadata in the frame of the INSPIRE Directive (GeoDCAT-AP[20]) and such exercise could be investigated and replicated for Regulatory Reporting metadata.

√ **Ensuring data quality**: the Secretariat-General is working on a framework for ensuring data quality. All these initiatives may have great impact, if they are spread and share sufficiently to become a common place for other European institutions.

√ **Create a catalogue that identifies which data is available for reuse**. Based on this, it could be addressed how to create a metadata that is relevant within this regulatory ecosystem's circumstances. Such data catalogue would also increase data discoverability and data reuse: by analysing the metadata, end-users are able to assess whether the available data is fit for their purpose and thus to decide to reuse it.

√ **Reuse of code lists.** Code-lists are important to harmonise the meaning of the (meta)data being used and to ensure that semantics are explained at their source. Using existing code-lists, or (if necessary) developing new ones, that are properly managed could be useful for finding monitoring and reporting data. Such code-lists are also relevant for the analysis of practices and patterns in reporting data. The Publication Office, which notably manages code-lists from several DGs, is aware of which DGs are linked to which code-list. The DGs only have to update who is being responsible for the code-lists, therefore insuring management of information over time. Other services managed by the Publication Office, such as EuroVoc and the Authority tables, should also be further used. As the management and curation of the terms is done at source, no duplication is required.

---

[19] https://ec.europa.eu/isa2/actions/improving-semantic-interoperability-european-egovernment-systems_en
[20] https://inspire.ec.europa.eu/good-practice/geodcat-ap

4.3 Technologies/infrastructures' support: designing the right tools

Specific steps:

√ **Metadata tools:** designing specific tools that would assist in creating metadata, notably through semi-automatic generation, could incentivise the allocation of resources to metadata creation.

Some of these already exist in the data service within JRC. For instance, the Semantic Text Analysis tool from the JRC[21], which screens text to identify key words could be re used by other institutions to screen the legislation for data to assess whether the data already exists or could be collected in combination of another legislation.

√ **Increase interoperability between existing tools:** Making existing tools, such as Reportnet 3.0, KOEL and RCD (mentioned in Section 1) more interoperable would result in a unique database which would provide a complete overview of all regulatory reporting obligations and provisions in the legislations.

√ **Create a metadata finding tool:** Semi-automated tools could also be designed for finding metadata, by performing text mining or by suggesting keywords for example. This would support the reuse of metadata. Several geographic information (GI) tools already exist and could be taken as inspiration for other types of data.

√ **Ensure infrastructure stability:** it is important to ensure that the infrastructure supports people's needs and remain interoperable. Hence, metadata, data sets, data catalogues, the all need to be maintained and kept functioning. For example, JRC and Eurostat services need to remain stable and back-ward compatible, so as to ensure collaboration between the services relying on the same tools. This would help further breaking and discovering silos.

4.4 Bird's-eye view: understanding what is there

Taking some distance and adopting a higher view (or bird's-eye view) to connect all available resources, and putting them into context, would also help breaking silos.

Specific steps:

√ **Creating inventories/databases** that raise awareness of what is already in place and make it more available for reuse. For instance, a Modelling Inventory[22] was done in MIDAS[23], which revealed to be useful to have a global picture of the interactions between different models and to identify what could be improved. MIDAS describes how models have been used, for example to support Impact Assessments. It this case, MIDAS allows to harvest information

---

[21] https://publications.jrc.ec.europa.eu/repository/handle/JRC116152

[22] The MIDAS inventory contains descriptions of models previously or currently in use by the Commission in support of policies. It is designed to make it easier to meaningfully assess models for complex problems, to maximize their benefits and to communicate their strengths and weaknesses clearly.

[23] MIDAS is available here: https://web.jrc.ec.europa.eu/policy-model-inventory/

about an Impact Assessment, such as the proposal the Impact Assessment accompanies, which legislation it resulted in, and whether this legislation has been superseded.

√ **Text and data mining tools** could be used to provide a wider overview of the reporting requirements that exist in different DGs and identify possible overlaps, which would also facilitate users experience (tools for facilitating users experience when navigating metadata are falling short at the moment). By harvesting all sources of reporting data and metadata, patterns could be identified. Such exercise could also be applied to business processes[24], in order to identify and streamline them. While it requires time and efforts, such analysis would highlight the costs and benefits of using metadata and highlight its value.

√ **Revising existing legislation** in order to streamline regulatory reporting requirements. An example of such initiative can be found in the Commission's Action Plan to Streamline Environmental Reporting[25] which consists in a stepwise approach to revise all environmental legislations in order to streamline the reporting requirements and to make sure Member states only report things once. This streamlining is done at both semantic and process levels. Such exercise could be replicated in other domains.

√ **Monitoring the availability of metadata:** once aware of the existence and availability of certain data, it is important to monitor if the metadata related to this data is also available. If the metadata is not provided, the next step is to investigate the reasons behind this absence (whether it is a lack of awareness, of capacity, etc.) and provide the necessary tools to ensure this gap is filled.

√ **Identifying stakeholders:** It is important to know who is involved in the different parts of the data cycle. Having a clear idea of who to call, a network of experts and existing Data Advisory teams would facilitate reaching out for help or for information whenever needed.

A good practice that could be the creation of data contact points network at the JRC. This means they are the data people of their unit, making themselves available for data enquires across the JRC. This is a practice that could be replicated across other DGs.

4.5 Take action: what our Community of Practice could do

Considering these recommendations, there are some key initiatives that the Regulatory Reporting Community of Practice could carry out, in order to further promote the creation and management of metadata in the EC.

- Publicising existing investments in the design of semi-automatic tools to create and find metadata;
- Creating the bird's-eye view document to identify existing data flows, metadata available, inventories, etc.

---

[24] An example of interesting approach to business analysis can be found here: https://www.youtube.com/watch?app=desktop&v=_vS_b7cJn2A

[25] COM(2017)312, Actions to Streamline Environmental Reporting, available at: https://ec.europa.eu/environment/legal/reporting/pdf/action_plan_env_issues.pdf

- Connecting with other services/groups working on (meta)data in the EC, such as the Data Advisory Service at JRC and groups that have visibility on the Information Systems of the EC.
- Fostering a federation of data catalogues by applying a standard format[26], such as DCAT-AP and its extensions, when it comes to regulatory reporting;
- Increasing the visibility of the Regulatory Reporting Community and of its goals in relation to metadata.
- Preparing training and educational packages with simple and clear instructions, to be spread across European institutions (e.g. to demonstrate some of the uses of metadata, to learn how to document metadata, etc.).
- Working in parallel with the EC Corporate Data Strategy.
- Aligning on some common standards within the Community to document our metadata and datasets, considering there will be always be some sectorial discrepancies.
- Create awareness around foundational ontologies such as the Core vocabularies
- Develop mappings to pivot around the various standards
- Promoting a common corporate tool to manage our metadata.
- And of course, starting the execution of the specific actions suggested in the previous sections of the chapter 4.
- Develop guidelines for managing metadata

---

[26] For instance, the ISO 19115 standard is being used as part of the INSPIRE Directive and is targeted towards the geographical information, while in the open data community in the Commission, the DCAT standard has been selected for metadata.

Terms and definitions

Code-lists – Code lists are predefined, organised sets of items that describe one or more (statistical) concepts. They are the building blocks for defining indicators. They are used to build multi-dimensional tables. Their dimensions can either take all items of a code list or only a subset. Where appropriate, code lists are based on official classifications such as NACE, ISCO etc.[27]

Data catalogue – A data catalogue is an organized inventory of data assets in the organization. It uses metadata to help organizations manage their data. It also helps data professionals collect, organize, access, and enrich metadata to support data discovery and governance[28].

Dataset – A dataset is a collection of data grouped according to certain criteria. The collection of one or many data files is referred to as a 'resource'. This makes it easier for users to locate related data, and easier for data providers to manage and maintain it[29].

Metadata – Metadata is referred to as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is data that provides information about other resources, a resource being defined as an identifiable asset or means that fulfils a requirement. Metadata can be used both to interpret the data and to search for (discover) the data[30].

Regulatory reporting – Regulatory reporting is the provision of periodical structured or unstructured data (qualitative or quantitative) from concerned private and public organisations, to competent authorities (at EU or national level) as required by the requirements set in specific EU legislations. It is a process, which entails the following main stages: the setting of regulatory reporting requirements in EU legislation, data acquisition, data processing and data sharing. These stages involve both the European Commission and officers within its Agencies dealing with reported data, as well as the parties which will be submitting data[31].

---

[27] Eurostat (n.a.) Code lists (Dictionaries). Retrieved from: https://ec.europa.eu/eurostat/data/metadata/code-lists
[28] Oracle (n.a.) *What Is a Data Catalog and Why Do You Need One?* Retrieved from: https://www.oracle.com/big-data/what-is-a-data-catalog/
[29] Definition from EU Data advisory Service. Retrieved from: https://webgate.ec.europa.eu/fpfis/wikis/display/EUODDVC/How+to+create+metadata
[30] Jenn Riley (2017) Understanding Metadata: What is Metadata, and What is it For?: A Primer, NISO. Available at: http://www.niso.org/publications/understanding-metadata-2017
[31] Regulatory Reporting Community of Practice (2021) What is regulatory reporting? Retrieved from: https://webgate.ec.europa.eu/fpfis/wikis/pages/viewpage.action?pageId=569247529

Regulatory Reporting
Community