European Commission

Methodology and tools for metadata governance and management for EU Institutions

Document Metadata

Property	Value
Date	2015-09-11
Status	Accepted
Version	1.00

This study was prepared for the ISA Programme by:

PwC EU Services (under SC118DI07171)

Disclaimer:

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

The ever-increasing volume of information exchanged within and between different organisations at both national and EU level requires setting up solutions that facilitate its automatic processing. Whilst technological developments offer various means to automate the exchange of information, technological developments alone cannot guarantee a greater interoperability between information systems. Thus, a fundamental aspect in the interoperability of public administrations is the need for common data standards (such as data models and reference data).

In this vein, this study, commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme, focuses on a mechanism to be put in place for the coordination of data standards governance and management at the EU level. The output of the tasks performed in the study can be summarised as following:

1. A specification of the requirements for managing the lifecycle of data standards in the EU institutions;
2. A proposed governance structure for data standards, including principles and roles and responsibilities;
3. A proposal for tools and best practices for managing and consuming data standards. In particular, this report contains a useful tools chain for managing data standards.

The study started with identifying the needs of **stakeholders from European public administrations and EU institutions**. Existing solutions for data standards management, governance, and tools have been analyses such as the Inter-institutional Metadata Maintenance Committee (IMMC) process, the INSPIRE maintenance and implementation group, Eurostat, the Dutch Knowledge and Exploitation Centre Official Government Publications, and the Local Government Inform in the UK.

Based on the description of the existing governance models and the requirements, **specifications for data standards governance are provided** by enlisting a number of activities that are considered important based on the feedback received by various stakeholders.

The study also defines the **need and specification** for data standards management. A number of **design principles** and **process specifications** are given. They should be implemented and **tailored to the specific scope** and context of each organisation.

The study considers the **standards and supporting tools** by making a **mapping** between stakeholder requirements and the functionalities offered by these tools. Following is the proposed use of tools and recommendations for further analysis on the integration and the re-use of these tools.

The study concludes by **proposing a number of next steps** for the implementation of data standards governance and management. By setting up a proper governance and management of data standards and by employing the right tools, public administrations greatly enhance their potential for coordination and interoperability and ultimately the possibilities for sharing and re-use of data standards thanks to:

- Reduced semantic interoperability conflicts;
- An increased quality and traceability of the information/data exchanged;
- A greater re-use of data standards;
- A reduced risk of duplication;
- An increased trust towards the information to be exchanged; and
- Savings derived from the re-use of already existing information.

The next steps should focus on the implementation and tailoring of the proposed governance and management models and tools. The proposed specifications are generic and should **be tailored to the specific context** and scope of each organisation. Emphasis should also be put on raise awareness and key stakeholder engagement. Implementation experiences should be captured in **lessons learned** and shared for **re-use**. The study demonstrates that existing tools are readily available and already used within the EU institutions to support the data standards lifecycle management process.

1. INTRODUCTION

This report is commissioned by the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission in the context of its Action 1.1 on semantic interoperability. It contains the requirements and the specifications for the governance and management of data standards as well as for tools for managing those that may be used by EU institutions and Member States.

1.1. Context: the need for coordination

The ever-increasing volume of information exchanged within and between different organisations at both national and EU level requires setting up solutions that facilitate its automatic processing. Whilst technological developments offer various means to automate the exchange of information, technological developments alone cannot guarantee a greater interoperability between information systems. A fundamental aspect is the need for common **data standards**[1]: primarily data models and reference data, which can be defined as follows:

A **data model** includes formal data names, comprehensive data definitions, proper data structures and precise data integrity rules [1]. It is a collection of entities, their properties and the relationships among them, which aims at formally representing a domain, a concept or a real-world thing. In practice, data models drive the design and development of information systems, as they can express the different types of information managed by an organisation.

Reference data is any data used to organise or categorise other data, or to relate data to information both within and beyond the boundaries of the enterprise. It usually consists of codes and descriptions or definitions [1]. Reference data typically consists of a small, discrete set of values that are not updated as part of business transactions but are usually used to impose consistent classification. Reference data normally has a low update frequency. Reference data is relevant across more than one business systems belonging to different organisations and sectors [2].

To make sure that different entities use the same data standards, stakeholders should invest time and effort in *coordinating* among each other. Uncoordinated exchanges among public administrations may lead, among other things, to

- Limited re-use of data standards, such as data models and reference data, that already exist because people are not aware of their existence;
- Use of competing standards;
- Use of different formalisms for encoding data standards and incompatible licensing rules;

[1] A data standard is a data standard specification that describes or defines other data [ISO111179]. Data standards indicate how compound objects are put together [NISO]. It can consist of among others data models, reference data, and identifier schemas. [10]

- Ad-hoc development of data standards that do not follow a structured process and methodology.

Better coordination requires data standards governance and data standards management, which we define as follows:

The governance of data standards comprises well-defined roles and responsibilities, cohesive policies and principles, and decision-making processes that define, govern and regulate the lifecycle of data standards.

The management of data standards is the good practice of adopting policies, processes, and systems to plan, perform, evaluate, and improve the use and re-use of data standards.

By setting up proper governance, management and tools for data standards, public administrations greatly enhance their potential for coordination and interoperability and ultimately the possibilities for sharing and re-use of data standards thanks to:

- an increased quality and traceability of the information exchanged;
- a greater re-use of data standards;
- a reduced risk of duplication;
- an increased trust towards the information to be exchanged; and
- savings derived from the re-use of already existing information.

Guidelines on how to forge agreements[2] for data standards already exist, but they do not provide details on how to set up the governance and management for this information.

The purpose of this study is therefore to provide guidelines for the setting up of governance, management and related tools for data standards both at the EU level and the Member States' level, taking stock of best practices and lessons learned from already functioning initiatives.

[2] European Commission. ISA Programme. Process and methodology for developing semantic agreements. 8 February 2013. https://joinup.ec.europa.eu/community/core_vocabularies/document/process-and-methodology-developing-semantic-agreements

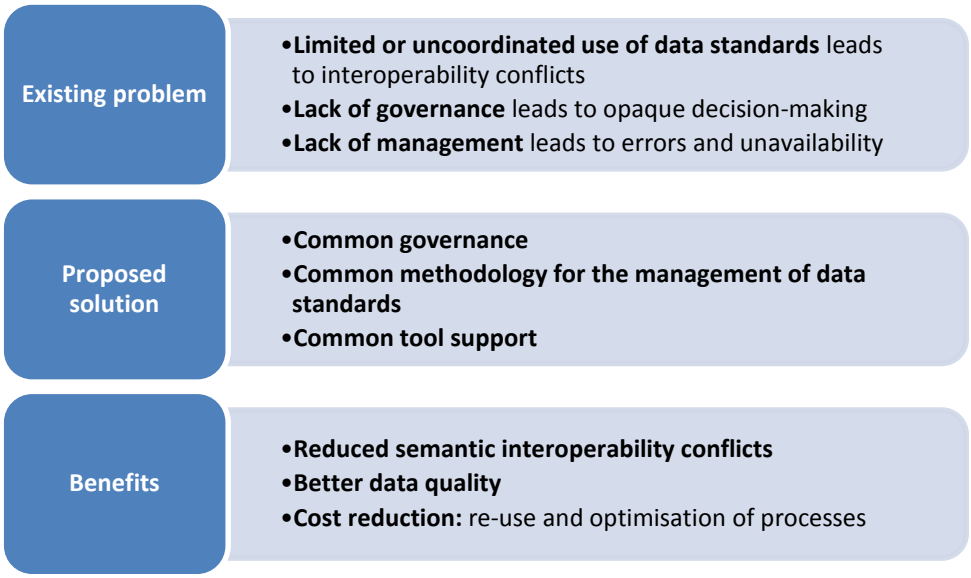


Figure 1 – Business case for a common approach to governance and management of data standards

1.2. Scope of this report

The scope of this report entails:

- **Local, inter-institutional and trans-European exchanges**: the study applies to information exchanges that take place at three levels, as depicted in Figure 2: local (i.e., within an EU institution), inter-institutional (i.e., between EU institutions) and trans-European (i.e., between Member States and the EU institutions).
- **High-level specifications**: the study provides high-level solutions that can be applied within a given public administration as well as across various public administrations and domains. Implementation in individual cases needs to be tailored to a specific organisational and technical environment. Therefore, a more in-depth implementation guide will be necessary using principles laid down in this document.
- **Data standards**: the study focuses on the governance, management and tools for data standards, only including reference data and data models. Metadata other than data standards, such as descriptive metadata (i.e., the description of documents, services and other resources that may be created, kept and shared across a network), are excluded from the scope of this report.
- **Tools**: the study investigates and tests a number of tools used at the momen of writing by public administrations for the governance and management of data standards. A complete market analysis of existing tools to support the governance and management of data standards is out of scope of this work.

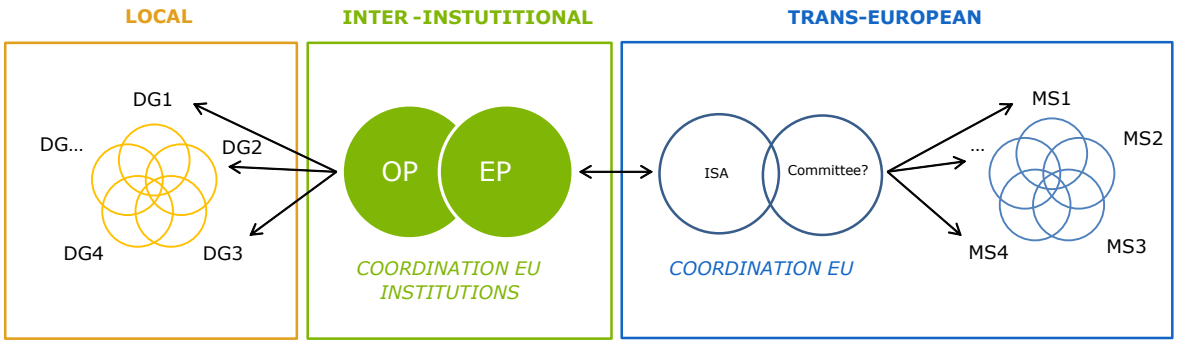


Figure 2 **– Levels of governance and management of data standards in the scope of this report**

1.3. Target audience

The main audience of this report is represented by the staff of the EU Institutions and consultative bodies as well as staff from national public administrations involved in the governance and management of data standards, and tasked to organise and operate the governance structures and maintenance activities.

This document can be of guidance to all sorts of situations where EU institutions have to coordinate work on data standards with public administrations in the Member States. The latter is often a consequence of an EU initiative or an EU legal act.

1.4. Approach

In order to define processes and identify tools for the management and governance of data models, the following three-step approach was used:

1. First, we identified explicit **stakeholder requests** that emerged in the course of interviews with stakeholders listed in Table 1 [3]. This was then complemented by additional **stakeholder needs** that were identified during three pilots with the European Commission Directorates-General for Competition (DG COMP) [4] and Maritime Affairs and Fisheries (DG MARE) [5], and the Publications Office (PO) [6].
2. Second, we identified **existing solutions** for the management and governance of data standards and associated tools.
3. On the basis of this input, we elaborated **specifications for the governance of data standards** (chapter 2)**, the management of data standards** (chapter 3)**, and data standards' tools** (chapter 4). The data standards' governance specification consists of a description of scope, organisation structure, decision mechanisms, etc. The data standards' management specification consists of a process description. The data standards' tooling section consists of a list of standards and tools, and a set of requirements against which the analysis of tools was performed.

Figure 3 – Requirements gathering and specification approach

The table below lists the stakeholder groups for which the proposed methodology and tools for data standards governance and management may be relevant.

Table 1 - Overview of stakeholder groups in the context of this report

Stakeholder groups	Example of stakeholder organisations
Standardisation organisations	• CEN • UN/CEFACT • OASIS • …
National public administrations	• KoSIT (Koordinierungsstelle für IT-Standards), Germany • CISE – Centre for Semantic Interoperability, Spain • Lithuanian Spatial Information Portal (LSIP), Lithuania • Knowledge and Exploitation Centre Official Government Publications (KOOP) , The Netherlands • Local Government Inform (LG Inform / LG Inform Plus), United Kingdom • …
European Parliament	• DG ITEC • …
Council of the European Union	• Archives of the Council of the EU • …
European Commission	• **Publications Office (pilot)** • **European Commission - DG MARE (pilot)** • **European Commission - DG COMP (pilot)** • European Commission - EUROSTAT • European Commission – JRC • …
Other European Union institutions	• Court of Justice of the European Union • European Court of Auditors • European Central Bank • European Economic and Social Committee • Committee of the Regions • …

2. GOVERNANCE OF DATA STANDARDS

This section elicits the stakeholders' requests and needs and formulates the specifications for data standards governance. We defined data standards governance as the set of roles and responsibilities, cohesive policies and principles, and decision-making processes that define, govern and regulate the lifecycle of data standards.

2.1. Stakeholder requests and needs

The table below lists the stakeholders' requests and needs for data standards governance.

Table 2 – Governance of data standards: stakeholders' requests and needs

ID	Stakeholders' requests and needs
ORGANISATION	
G1	**Involve direct stakeholders in the governance process**: direct stakeholders should be involved in the governance process of data standards to ensure that the interests of the stakeholders are taken into account.
G2	**Involve operational staff in functional meetings:** representatives from the operational level should participate in functional-level meetings.
SCOPE	
G3	**Local and inter-institutional governance:** the mechanism for governance should encompass both local and inter-institutional data exchange.
G4	**Identify a core set of data standards** to be standardised first.
G5	**Mappings** should be considered as a solution of last resort. It is recommended to try as much as possible to come up with a common agreement. Only if it is not possible to reach such an agreement, the governance should consider mapping as a solution of last resort.
DECISION MECHANISM	
G6	**The mandate should be clear** The governance mechanism should clearly state the mandate of the governance body with regard to taking decisions on: • Changes to reference data; • Intellectual property rights linked to reference data; and • Enforcement, i.e., implementation of reference data specifications in systems.
G7	**Decision process should take into account time constraints** Decision processes should be linked to time constraints which are dependent on the nature of the decision to be taken.
G8	**Describe how agreements are reached** The decision-making processes should describe how agreements are reached – e.g., via a qualified majority or via consensus building.
ENFORCEMENT POLICY	
G9	**Legislation should be formulated at high level** and should not specify details like the values in a code list or the elements of a data model.

ID	Stakeholders' requests and needs
G10	**Comply or explain approach**: it is recommended to deploy a "comply-or-explain" enforcement policy for the implementation of standards and specifications for data standards, irrespective of whether the implementation is realised through procurement or in-house development.
G11	**Increase awareness on benefits of sharing and re-use**: the solution should foresee to increase the awareness among stakeholders on sharing and re-use benefits by means of clear arguments aligned with specific business cases.
G12	**Take into account legal instruments:** the information which is exchanged between Member States and the European Commission is often specified in EU legislation. When including data standards governance in the decision-making process, efficiency and speed should be taken into account.
PROCESS FOR CONTINUOUS IMPROVEMENT	
G13	**Decisions should be documented** Specific decision-making processes which are dependent on the context in which a decision is required should be developed, documented and shared with all relevant stakeholders.
G14	**Licensing framework**: the governance should also take care of decisions related to the licensing framework.
G15	**Adaptations** to the needs of the users should be delivered timely. At the same time, it is necessary to guarantee stability.
G16	**Quality Assurance** The reference data management and governance methodology should take into account quality controls when: • designing reference data; • updating or importing reference data; and • propagating reference data to production systems.
G17	**Risk mitigation** Risks related to the propagation of changes to reference data into operational systems, should be mitigated by governance processes.

2.2. Specifications for the governance of data standards

Based on the analysis of existing governance models as described in *Annex II: Existing solutions for the governance of data standards* and the requirements identified above, the next paragraphs provide specifications for the governance of data standards. The following section does not aim to provide an exhaustive list of best practices that are necessary to be applied for the correct functioning of a data standards governance mechanism. Rather, the section aims to extrapolate general best practices from concrete examples coming from the day-to-day work of a limited number of stakeholders.

2.2.1. *Determine the scope*

When setting up a mechanism for data standards governance it is necessary to determine the scope in which it is applied. The scope comprises, among others, the following aspects:

The domain or sector: In certain cases, it may be limited to one specific domain, such as food safety, defence, healthcare, or finance. In other cases, it encompasses a variety of domains such is already the case with regard to the governance of data standards in the context of the European Union decision-making process. When identifying the policy domain the following elements should be clearly identified: the topics covered; who will be impacted by changes to the data standards; the existence of data standards harmonisation efforts for the same instances; the consequences derived from compliance or lack of compliance.

The governance levels: In this report, we have considered the local, inter-institutional and trans-European levels.

The level of abstraction: Within one domain or across domains, it is possible to define the extent to which data standards are being specified. Consider for instance the following levels of abstraction depicted in Figure 4.

- **Core specifications**: context-neutral data standards that define the fundamental characteristics. The data standards can be applied in several contexts. One example here are the Core Person[3], Business[4], Location[5], and Public Service[6] Vocabularies developed by the ISA Programme.
- **Domain specifications**: domain-specific data standards that cover a domain to a larger extent. One example here is the HL7 Reference Information Model[7] in the healthcare domain, or the Universal Business Library (UBL) of OASIS[8].
- **Information exchange specifications**: data standards specifications that are specific to one context of information exchange. Examples here are the good practices in the electronic invoicing in public procurement[9] developed by the European Commission and the Statistical Data and Metadata eXchange[10].

[3] Core Person Vocabulary: https://joinup.ec.europa.eu/asset/core_person/
[4] Core Business Vocabulary: https://joinup.ec.europa.eu/asset/core_business/
[5] Core Location Vocabulary: https://joinup.ec.europa.eu/asset/core_location/
[6] Core Public Service Vocabulary: https://joinup.ec.europa.eu/asset/core_public_service/
[7] HL7 Reference Information Model: http://www.hl7.org/implement/standards/rim.cfm
[8] OASIS UBL: https://www.oasis-open.org/committees/ubl/
[9] http://ec.europa.eu/enterprise/sectors/ict/files/invoicing/activity_2_public_procurement_final_en.pdf
[10] Statistical Data and Metadata eXchange: http://sdmx.org/

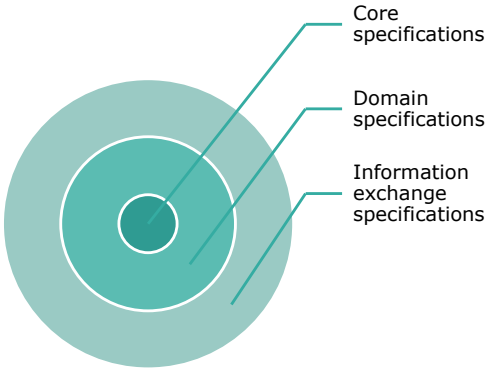


Figure 4 – The level of abstraction

Scope criteria: There must be a clear set of scope criteria that determine whether a specification for data standards should be placed under common governance, as this may require considerable coordination costs, but can also entail considerable benefits of interoperability. Some important scope criteria that have been identified as highly relevant are:

- **The level of information exchange**: whether the data standard(s) in question will be used exclusively within a given organisation or by two or more institutions/entities for exchanging information;
- **Maturity level**: the stability/maturity of the data standard(s) that an organisation wants to share. For example, if from the beginning an organisation knows that a given data standard is not mature enough and that it will probably still change in the future, it does not make sense to share it with a wider audience, which will start relying on something that is actually not finalised;
- **Potential for use and re-use**: A very important criterion is the potential for re-use of a data standard. For example, a reference dataset that may be re-used across various sectors and stakeholders has a greater potential than a sector-specific data model.
- **Commitment for maintenance**: Fundamental is also the degree of commitment to maintain and keep data standards up to date. For example, the Publications Office has expressed in different fora its willingness to maintain certain data standards, notably the named authority lists[11].
- **Commitment for use**: How strong is the commitment of organisations to actually use the specification for data standards?

2.2.2. *Set up a governance structure*

Once the decision to set up a governance mechanism for data standards has been taken, it is necessary to put in place the overall structure that it should have. The governance should include permanent members, temporary representatives and a secretariat taking care of logistical and coordination matters.

Based on our prior analysis, we concluded that that the optimal solution is to have three-level governance corresponding to:

[11] Named Authority Lists: http://publications.europa.eu/mdr/authority/index.html

- **Steering committee**: composed of representatives of the institutions and public administrations that set the strategic directions. It should include representatives from both the business and the architecture side. This level is driven by people who have the means and vision to take decisions on scope and goals. They meet periodically to review the progress made and to intervene sporadically to solve conflicts and nominate members.
- **Governance committee**: composed of the main stakeholders. It takes decisions on the organisational support required to the operational team. For example, this is the role of the Interinstitutional Metadata Maintenance Committee (IMMC)[12]. It oversees the compliance of the operational team and assumes the responsibility to develop, disseminate and enforce the required procedures.
- **Operational team**: It is composed of a single team that carries out the day-to-day work. It deals with various aspects of governance and management of data standards from collection and creation to administration. It is the level where management is executed. An example is the Metadata Registry of the Publications Office. Another example are the ISA Core Vocabularies.

2.2.2.1.Required expertise for the management of data standards

For the governance of data standards, a number of essential competences can be distinguished. The following areas of expertise should be included in the operational team that is responsible for the management and governance of the data standards.

- **Domain expertise**: knowledge about the semantics of the data for which the data standard(s) is (are) used and the applications in which the data is used.

  This expertise ensures that the team has a good understanding of the functionality that the data standard(s) is (are) supporting. This allows the team to identify potential problems that could be generated by changes in data models, schemata and reference data.

- **Information management expertise**: knowledge about theory and practice of information management, e.g., information and library science.

  This expertise ensures that approaches to definitions of data standards' elements and the expression of relationships between data standards' elements – e.g., hierarchies in controlled vocabularies – are sound and based on best practices in the domain of information science.

- **Technical expertise**: knowledge about the technical approaches to be used for the technical implementation in the environment in which the data standards are used.

  This expertise ensures that the implementation conforms to the technical environment, e.g., using the protocols, schema and mark-up languages, such as XML and RDF, and tools.

[12] Interinstitutional Metadata Maintenance Committee: http://publications.europa.eu/mdr/core-metadata/

- **Documentation and publication expertise**: knowledge about the documentation rules and publication processes used in the environment in which the data standards are used.

  This expertise ensures that the data standards and changes are documented in the format and language that are appropriate for the users of the data standards, and that the data standards are published in the formats (human- and machine-readable) that allow easy integration in applications and services.

- **Standardisation expertise**: knowledge about standardisation rules and procedures if the data standards approaches are intended to be submitted to standards bodies for regional, national or international standardisation.

  This expertise ensures that submission to the appropriate standards body conforms to the format and procedures used in the standardisation process.

2.2.3. ***Define decision mechanisms***

Decision mechanisms prescribe how and when to perform tasks related to the governance of data standards. They are fundamental in achieving the established goals without having to constantly intervene on daily operations.

Decision mechanisms should enable to take decisions such as:

- Whether a data standard must be placed under local or inter-institutional governance;
- How to change and improve the management process(-es);
- Whether a change request to a data standard must be accepted or rejected (based on an impact analysis; cost-benefit analysis, risk analysis);
- Whether an accepted change request will be released immediately or in a scheduled release;
- Where to store a data standard and with which access restrictions (define roles and responsibilities);
- Whether a data standard can be published under an open licence, or what is the proper licence to be used;
- Whether a data standard should be aligned with others;
- Which policy is followed to encourage or mandate the re-use of a data standard;
- Which method is used for documenting data standards;
- Whether a data standard should be deprecated; and
- Which standards and tools to use in the management process (-es).

These decisions can be taken using different modalities:

- Consensus: a decision is taken only when there is a full consensus;
- Majority vote: a decision is taken upon majority vote;
- Veto: stakeholders are informed and can raise a strong objection; or
- Endorsement: asking stakeholders to endorse it after creation or update.

Furthermore, in the context of the European Union, special mentioning should be made with regard to the Comitology procedure. In this context, two procedures are particularly relevant for this study:

- The advisory procedure; and
- The examination procedure

Details of the Comitology procedure are given in Table 3.

Table 3 – Comitology procedure

Illustration: Comitology procedure[13]
When the governance of data standards involves the EU and Member States and there is a legal instrument requiring uniform conditions for the implementation of data standards, then implementing powers can be conferred to the European Commission. In this case, Member States can control the Commission's exercise of implementing powers. The rules and general principles concerning mechanisms for control by Member States of the Commission's exercise of implementing powers[14] are set up in Regulation (EU) No 182/2011[15] of the European Parliament and of the Council of 16 February 2011. For the purposes of such control, committees composed of the representatives of the Member States and chaired by the Commission are set up. The primary role of these Committees is to provide an opinion on the draft measures that the Commission intends to adopt. These opinions can be more or less binding upon the Commission according to the procedure which has been foreseen by the legislator. One of the following two procedures is foreseen: • The advisory procedure: here, the Commission shall take the utmost account of the committee's opinion. • The examination procedure: here, implementing acts cannot be adopted by the Commission if they are not in accordance with the opinion of the committee, except in very exceptional circumstances, where they may apply for a limited period of time. In addition, specific procedures are foreseen for measures to apply immediately on imperative grounds of urgency (Article 8). In this case, the Commission adopts an implementing act of immediate application, without its prior submission to a committee.

2.2.4. ***Define procedures to handle requests***

To make sure that the needs of the requestors are taken into account, the governance of data standards should establish clear procedures to be followed depending on the case into question. For example, it may be that a requestor submits a request to update a schema. Such a request may have an important impact on several information systems which are implementing the schema and therefore should be carefully assessed. Here, timing may be less relevant than the analysis on the impact that such a request might have. Conversely, a requestor may submit a request for a deprecation and update of a code where the urgency outweighs the impact that such

[13] See also https://myintracomm.ec.europa.eu/corp/sg/en/comitology/implementing/pages/tools.aspx for further information

[14] OJ L 55, 28.2.2011, p. 13–18.

[15] http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32011R0182

a modification may have. Therefore, when deciding which procedure to apply the governance mechanism should take into account aspects such as:

- The **justification** behind a given request: is there a real need for taking into account such a request? It may be that the request is made on needs that have not really been thought through and therefore the implementation may be postponed or abandoned;
- The **urgency**: does the request need to be implemented as quickly as possible because otherwise several systems will be "blocked", or stakeholders will be using an outdated version?
- The **impact** of the request in terms of information systems as well as stakeholders involved: it may be that a request for a change of a schema would require an update by several entities and therefore would also impact several systems. In this case, the assessment of the impact should be carried out into details.

An example of good practice in this context comes from the Publications Office and the ISA Programme. The Publications Office is currently compiling sets of standard requests in order to know already in advance how to treat them based on which category they fall in. This approach may save time and help those analysing the various requests in their daily job. The ISA Programme has also defined a change management process for the Core Vocabularies[16] [7], which process formalises how changes to the specifications are managed and how new releases are published.

2.2.5. *Ensure that modifications are communicated promptly to relevant stakeholders*

Once the governance mechanism finally takes a decision, it is necessary to ensure that all relevant stakeholders are informed, so that not only can they adapt their systems but they can also provide feedback. Therefore, the governance mechanism should establish communication channels through which stakeholders are kept up to date. Depending on the target group and on the way they usually communicate, different solutions may be envisaged including for example: mailing lists, RSS feeds and announcements provided during the plenaries.

2.2.6. *Set up registry as the authoritative source*

When setting up a governance mechanism for data standards, it is fundamental to make available an authoritative source on which the data standards are housed. In most cases, the authoritative source is a repository or a file server that is accessible online. It should allow anybody to access code lists, concept schemes, data structure definitions, etc. The existence of an authoritative source increases the confidence of potential re-users because it ensures that everybody has access to the same information as well as the confidence over the quality of the data standards.

2.2.7. *Establish enforcement approach*

The governance mechanism should also establish which enforcement regime should be applied to promote the sharing and re-use of data standards. Enforcement policy

[16] Still to be published.

embraces a wide spectrum of activities, going from the drafting of public procurement to the implementation of data standards.

With regard to public procurement, it is worthwhile mentioning the best practices and guidelines prepared by the European Commission[17].

With regard to the different typologies of enforcement policies, the most common ones can be summarised as follows:

- **Legal requirement**: implementation is enforced by law, either by secondary legislation, council conclusions, or by referring to standards, via comitology procedure (as it is the case for state aids and the implementation of the Inspire Directive). Here, an important requirement is to make sure that specific data standards are not included in the legal instrument. Otherwise, every time that there is a need for an update, it will be necessary to go through the legislative process, which would make it a heavy process not serving the users' needs. Details like the values in a code list or the elements of a data model should be specified as part of the implementation documentation and made available from an authoritative source to which the legislation can refer.
- **Comply-or-explain**: implementation is not enforced by law, but public administrations have to comply with the use of a particular specification or standard for metadata. Otherwise, they should explain why the specification or standard in question does not fit their needs. In certain cases it may even be requested to contribute to upgrade the model.
- **Voluntary**: implementation is encouraged via information campaigns. What is crucial in this case is that stakeholders share the same goals and are aware of the advantages that an effective and efficient use of the data standard may provide. There are several actions that can be undertaken to make sure that this happens.

2.2.8. *Establish a licensing framework*

In order to make the data standards available for sharing and re-use, the governance should establish the licensing framework under which the data standard(s) can be exchanged and re-used.

To make sure that data standards are used by a critical mass, it is recommended to use licences that are as open as possible with protection against misrepresentation.

In addition, in order to increase legal certainty and help potential users, it is also recommended to make sure that information related to licensing frameworks is properly conveyed and easily accessible.

Examples of such licences are Creative Commons CC Zero (CC0)[18], Open Data Commons Public Domain Dedication and Licence (PDDL),[19] Creative Commons Attribution 4.0 (CC-BY-4.0)[20] and the ISA open metadata license[21].

[17] European Commission, Guide for the procurement of standards-based ICT — Elements of Good Practice, SWD(2013) 224 final, Brussels, 25.6.2013.
[18] http://creativecommons.org/publicdomain/zero/1.0/#sthash.dXmnPsbW.dpuf.
[19] http://www.opendatacommons.org/odc-public-domain-dedication-and-licence/#sthash.cSWz1Guq.dpuf.
[20] http://opendefinition.org/licenses/cc-by/#sthash.Hg8dbSEy.dpuf.
[21] https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11

2.2.9. *Set up quality controls*

In order to ensure that data standards are acceptable for publication and use, it is indispensable to apply quality assurance and quality control. The governance should take into account the following aspects:

- **Accuracy**: Data standards should enable instance metadata to describe the resources accurately, e.g., a data model needs to include all properties and attributes necessary for the applications that use the instance metadata; a controlled vocabulary needs to include all terms necessary.
- **Trustworthiness**: Data standards should be made available from an authoritative and reliable source to enhance their potential for re-use and therefore interoperability. If data standards is retrieved from an external source, such as an international standards body, this provenance information needs to be provided so that anybody wanting to use it can check the origin of the metadata itself.
- **Integrity**: Data standards should be protected against unauthorised alteration.
- **Timeliness**: Data standards should be kept up-to-date and promptly available when users want to access or use them. The frequency with which changes are applied should find the right balance between stability and flexibility. A main challenge is to make sure that the governance procedure put in place allows the processing of requests fast enough for users to actually be able to use the data standard(s) when needed.
- **Completeness**: Data standards should be created and maintained in conformance with an agreed standard, respecting common rules for identifiers, names and descriptions. This is an example of something that can be checked relatively easily by tools.
- **Validity**: Data standards may have restricted validity, for example in specific time periods or geographical areas. This information needs to be readily available to users.
- **Accessibility**: Data standards should be easily accessible, understandable and usable, for consumption by both humans and machines.

In addition, control processes should be in place in order to validate and guarantee the quality of the data standards. Consistency and completeness of data standards may be imposed by the tools for change management or checked before publication through automated checks (e.g., whether the data standards conform to common standards, or whether newer versions have later dates of modification) and human intervention, e.g., peer review.

3. MANAGEMENT OF DATA STANDARDS

This chapter provides an overview of the requirements and specifications for the management of data standards. First, it elicits stakeholder requirements and needs. Then, it identifies good practices from existing standards, which are enlisted in *Annex III. Existing methodologies for data standards management*. Finally, it elicits design principles and describe the lifecycle management processes for data standards.

3.1. Stakeholder requests and needs

The table below lists the stakeholder requests and needs for the management of data standards

Table 4 –Management of data standards: stakeholder requests and needs

ID	Stakeholders requests and needs
M1	**Access rights management** As certain parts of a data standard can contain sensitive information, for example in the case of the data models for the Marine and Maritime environment, processes have to be specified for: • managing access rights for the data standards; and • managing the access request procedure.
M2	**Provide unambiguous guidelines** for the use of data standards.
M3	**Documentation of data standards** Data standards should be documented in both human- and machine-readable formats. Moreover, every version of a standard should be made easily accessible for all stakeholders.
M4	**Documentation of management processes** Processes for the management and governance of reference data should be documented in a clear, accurate and complete way.
M5	**Continuous improvement** The management processes should allow for continuous improvement for the purpose of remaining responsive to the needs of the users.
M6	**Alignment with external bodies** The change management of data standards should be aligned with procedures from external standards bodies, especially when standards are re-used.
M7	**Data Model Lifecycle Support** The proposed management processes should support the lifecycle of data models, which includes: • Designing a data model; • Assuring the quality of the data model; • Managing access requests to the data model; • Requesting a change to the data model; • Updating a data model; and • Documenting and publishing a data model.

M8	**Reference Data Lifecycle Support** The proposed management processes should support the lifecycle of reference data, which includes: • Designing reference data; • Managing reference data changes; • Sharing and re-using reference data; • Harmonising reference data; and • Deploying reference data.

3.2. Specifications for the management of data standards

Data standards will evolve over time, either because of changes in the environment (e.g., emergence of new subject areas) or because of changes in functionality that must be supported (e.g., new services).

The lifecycle of the data standards is organised in six phases:

- **Design a data standard:** to support a new service or application, data standards need to be designed, implemented and subsequently used in applications to support interoperability;
- **Manage change of a data standard:** while requirements and functions of applications evolve, data standards need to change to support changing applications;
- **Harmonise data standards:** in order to foster interoperability, internally governed data standards should be harmonized with external data standards. Cross-references may be created at schema and value level. For example, at a value level, a local list of languages in a reference data set could be linked to the ISO639 list of languages. On the schema level, for example, the "language" concept of the ISA Core vocabularies could be mapped with matching concepts of other vocabularies, such as KoSIT and dcterms[22].
- **Release a data standard:** after changes have been applied to data standards, the resources and associated documentation need to be released to the stakeholders;
- **Deploy a data standard:** when a new version of a data standard has been released, the changes need to propagate to the operational systems used by the stakeholders:
- **Retire a data standard:** when applications are no longer supported or migrate to new data models or reference data collections, the data standard is no longer relevant and may be decommissioned.

The following sections describe the high-level processes that are included in the management of these six lifecycle stages of data standards. Although there are different levels of governance of data standards as described in section 2.2.1, the processes described below are generic and are applicable to all.

The management processes described in this section provide a generalised view on the steps to be taken in managing data standards. In practice, application of the

[22] The mappings of ISA Core Vocabularies with related vocabularies are expressed and visualised via http://mapping.semic.eu.

approach in individual cases will require tailoring of the processes to the organisational and technical environment of such cases.

3.2.1. *Design data standards*

Release	Document and publish the changed data standard
Goal	To design or update a common data standard taking into account the needs and interests of all stakeholders.
Preconditions	The decision for designing or updating a data standard has been taken by the appropriate authority and a governance structure has been put in place, including the following roles as defined in the "primary actors" section.
Success End Condition	A new version of a data standard that takes into account the needs of all stakeholders.
Primary Actors	• Steering Committee: - Endorsement Group (EG): a group of third parties that have an interest in the data standard • Governance Committee: - Activity Leader (AL) • Operational Team: - Secretariat - Expert Pool (EP) - Domain Model Working Group (DM-WG) - Data Entity Subgroups (DESG) - Review Group (RG)
Secondary Actors	The wider user community – their requirements will be elicited and taken into account during the design of the data standard.
Frequency	• Ad-hoc; or • Periodic.
Trigger	Decision to develop a data standard by an appropriate authority.

Flow:

Step	Description	Actor
1	Set up a work-plan for development, review and endorsement	Governance Committee
2	Establish working environment and culture	Governance Committee
3	Identify and invite potential candidate experts	Governance Committee
6	Invite members of the working group (WG) (Operational Team)	Governance Committee
7	Invite and appoint chair and editor (part of the Operational Team)	Governance Committee
8	Set up work plan for the DM-WG	Governance Committee
9	Publish drafts of the Domain Model, seeking feedback from the RG on each occasion	Operational team
10	Consider and process comments	Operational team

11	Issue the last call WG draft of the domain model and invite the RG to review	Operational team
12	Review last call draft	Operational team and community
13	Publish draft domain model	Operational team
14	Publish final domain model	Operational team
15	Endorsement by the stakeholders	Stakeholders

3.2.2. ***Manage change of data standards***

Request change	Request a change to a data standard
Goal	To create a change request for a desired modification to a data standard (data model or reference dataset).
Preconditions	• The data standard has been designed and published. • The data standard has been (or is being) implemented in a production system. • An authoritative source is available where stakeholders can access the data standard.
Success End Condition	The creation of a change request, which triggers the "design" phase discussed in section 3.2.1.
Failed End Condition	Decision not to create a change request
Primary Actor	• Governance Committee – receives feedback and decides on creation of change requests • Operational Team – performs analysis
Secondary Actors	Stakeholders – submit feedback
Frequency	• Ad hoc: when receiving feedback from users or when (new) legal obligations arise; or • Periodic: when carrying out periodic reviews of data standards to ensure conformance to re-used standards.
Trigger	• User feedback; • Periodic review; • Legal obligation; • Release of a new version of a re-used standard.

Flow:

Step	Description	Actor
1	Receive request	Governance Committee
2	Initial evaluation	Operational Team
3	Accept request	Governance Committee
4	Propose solution with impact analysis and roll-out plan	Operational Team
5	Review proposal	Governance Committee, Stakeholders
6	Accept proposed solution	Governance Committee

3.2.2.1. Implement change of data standards

	Create, modify or delete an item in an existing data standard
Goal	To transpose into the data standard, under the remit of the Governance Committee, the accepted change requests, leading to a new release of the data standard and the documentation that accompanies it.
Preconditions	Accepted change request
Success End Condition	The data standard is successfully updated.
Failed End Condition	The change request is not incorporated.
Primary Actors	• Operational Team – handles the change request and develops the solution • Governance Committee – accepts the solution
Secondary Actors	Stakeholders – are involved in the process to make sure that the solution meets their requirement
Frequency	• Ad-hoc; or • Periodic.
Trigger	When a change request has been accepted and the stakeholders have been informed of an upcoming change, optionally the "Harmonise" phase is executed followed by the "Release" phase.
Comment(s)	Activities in this phase may consider the incorporation of individual changes in data standards, or group changes together into pooled releases, depending on the urgency and impact of the changes. Activities and frequency are different for changes to data models and changes to reference datasets.

Flow:

Step	Description	Actor
1	Plan change	Operational Team
2	Apply changes	Operational Team
3	Test solution	Operational Team, Stakeholders
4	Prepare documentation	Operational Team
5	Accept change	Governance Committee

Data models and reference data can either be managed in an XML-based environment using XML Schema Definitions, or in a Linked Data environment using RDF-based formats. The management workflow for these types of data standards is outlined in the table below. The steps are further described in the text following the table.

Table 5 – Manage changes in data standards

Step	XSD[23] model (data model)	XSD code list (reference data)	RDF schema[24] (data model)	SKOS[25] vocabulary (reference data)
1	Determine whether the element is already defined in an existing XML schema. If so, import if possible.	Determine whether a code is already available in an existing code list. If so, use the same code if possible.	Determine whether the element (class, property) is already defined in an existing RDF namespace. If so, re-use; if not continue.	Determine whether the concept is already available in an existing SKOS concept scheme. If so, re-use; if not continue.
2	Identify XSD where element needs to be added, changed or deleted.	Identify XSD where code needs to be added, changed or deleted.	Identify existing namespace for new element.	Identify existing SKOS concept scheme for new concept.
3	Create new or modified element	Create new code in context of code list	Mint URI and create definition	Mint URI and create definition
4	Add new element, make change to existing element, or delete element.	Add new code, change meaning of existing term, or delete term.	Add element to namespace.	Add new concept in concept scheme.

Step 1:

In the development of data models and reference data, standard schemas and vocabularies should be re-used as much as possible; when local schemas and vocabularies are used, map those to standard elements as much as possible.

Step 2:

In all cases, it needs to be determined in which file the element needs to be added, changed or deleted. If the data standards are part of an XML Schema Definition, it is the XSD to be amended; if the data standards are managed as an RDF schema, it is the RDF namespace; if it is a controlled vocabulary expressed using SKOS, it is the SKOS concept scheme.

Step 3:

For an element in an XSD model, the element needs to be defined with its element name and structural definition. For a code to be included in an XSD-based code list, the name, attributes and location in a hierarchy need to be defined.

[23] W3C. W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. http://www.w3.org/TR/xmlschema11-1/
[24] W3C. RDF Schema 1.1. http://www.w3.org/TR/rdf-schema/
[25] W3C. SKOS Simple Knowledge Organization System Reference. http://www.w3.org/TR/skos-reference/

For elements (class, property) in an RDF schema and for a concept in a SKOS concept scheme, a URI needs to be minted in the context of the RDF namespace or SKOS concept scheme, together with an unambiguous definition of the element or concept.

Step 4:

In this step the element that was prepared in the previous step is incorporated in or deleted from the existing schema.

In XSD, a new element or code is added; a change in an existing element or code overwrites the old version; a deletion is simply removed from the schema definition.

In RDF namespaces and SKOS concept schemes, a change in semantics can only be applied if this does not affect existing applications. In general, semantic meaning may be broadened (as existing data remains valid) but never narrowed (which could make existing data invalid).

Deletion of elements or vocabulary terms should be avoided unless it can be verified with complete certainty that such items are not used anywhere; otherwise, items should be annotated, e.g., with an status property (e.g., adms:status) with value "Deprecated".

3.2.2.2. Update cycles

There are differences in the requirements for the periodicity of changes for data models on the one hand and reference data on the other hand. These differences are linked to the different needs for stability versus flexibility.

Data models are strongly linked to the interoperability of applications. Therefore, changes in a data model have a direct effect on the applications that are based on it. In many cases, software systems will need to be rebuilt importing the new model and upgrading the functionality before they can interoperate with others. In practice, changes in data models will be relatively infrequent (less than annual) and changes will be accompanied by a strongly managed implementation plan aligned with a software upgrade cycle.

Reference data is usually more loosely linked to the basic functionality of applications. Changing or adding a code in a code list will not have a disruptive effect on the existing functionality. These types of changes may also occur with a higher frequency (one or more times per year) than model changes, and are usually easier to propagate through a network.

3.2.2.3. Versioning

As part of the lifecycle, the change management process will lead to the creation of a set of versions of data standards.

While the latest version of a data model or reference data collection is clearly the most important resource to be re-used as this supports the functionality at that particular point in time, it is also necessary that previous versions are still available

for inspection. This makes it possible to determine what functionality was available in the past. In relation to that, it is also important that the documentation of previous versions as well as change logs are kept available.

Identification of versions of data standards can be done by time-stamping the versions, by assigning version numbers or by combining those two approaches.

3.2.3. *Harmonise data standards*

Harmonise	Create links to internationally standardised or widely used data standards
Goal	To establish equivalence links between data standards under remit of the Governance Committee and other data standards, either harmonising common local ones with internationally standardised or otherwise widely used data standards, or providing information to enable mapping from local data standards to common ones.
Preconditions	A minimum of two systems that are implementing different data standards need to interoperate.
Success End Condition	Mappings are created between data standards.
Failed End Condition	Equivalence links and mapping specifications are not available.
Primary Actors	• Governance Committee – decide which external data standards to use for harmonising • Operational Team – creates the links to external data standards and prepares mapping specifications
Secondary Actors	• Stakeholders – receive mapping specifications • Owners of external data standards – may be contacted to create appropriate links
Frequency	Ad-hoc
Comment(s)	It needs to be decided on the level of the Governance Committee to which external data standards links are established.

Flow:

Step	Description	Actor
1	Identify external data standards to be linked to	Governance Committee
2	Include links to external resources, if necessary contact owners of external resources	Operational Team
3	Create mapping specifications	Operational Team
4	Apply mappings from local data standards to common resources from the external data standards	Stakeholders

If appropriate, RDF classes or properties should be linked to other items in the namespace (e.g., to express sub-class or sub-property relationships) or to items in other namespaces (e.g., to indicate equivalent classes or properties); SKOS concepts can be linked to other concepts in the concept scheme (e.g., to link the concept to

broader or narrower terms) or to concept in other concept schemes (e.g., similar concepts).

3.2.4. *Release a data standard*

Release	Document and publish the changed data standard
Goal	To document a new version of the data standard and to publish it on the authoritative source.
Preconditions	The Governance Committee has signed off the publication of a new version of the data standard.
Success End Condition	The release of the data standard is published and documented on the authoritative source, including, the public documentation.
Primary Actors	• Operational Team – prepares and issues the release of the data standard and its documentation • Governance Committee – oversees the release and the announcement
Secondary Actors	Stakeholders – are informed of the release
Frequency	• Ad-hoc; or • Periodic.
Trigger	Completion of an updated (and optionally harmonised) version of the data standard.
Comment(s)	Data standards should be documented in human- and machine-readable formats. The machine-readable documentation should follow a standard vocabulary, such as ADMS. In addition to the machine-readable data, it is helpful to provide guidance documentation, for example outlining which standards and methods have to be used in specific cases. Models and model elements, as well as the items in controlled vocabularies should be assigned URIs and those should be maintained persistently. Descriptions of the metadata should follow unambiguous guidelines, in order to facilitate search and retrieval. Wherever possible, metadata should be made available under an open licence on an open platform, such as Joinup. However, if some parts of documentation are sensitive, those should be protected by appropriate access control. All the resources managed should be published in such a way that they can be re-used easily by other systems, for example as plugins, via web-services, via API, or using a dedicated client. It is important to make the sharing and the accessing of the shared model and reference data easy because sharing is the basis for interoperability.

Flow:

Step	Description	Actor
1	Issue new release documentation	Operational Team
2	Move new release of the data standard to the production environment	Operational Team
3	Notify stakeholders of the release and the roll-out plan	Governance Committee

3.2.5. *Deploy data standards*

Deploy	Roll out changed data standards to distributed systems used by stakeholders
Goal	To effectively implement the changes in data standards in the operational systems used by stakeholders, while protecting the live environment of their systems through planning, testing, building and implementing a grouped set of changes.
Preconditions	• New version of data standard(s) published on authoritative source • Roll-out plan established
Success End Condition	Changes successfully implemented in all systems that use the data standard(s)
Primary Actors	• Operational Team – provide project management and support for the roll-out. • Stakeholders – execute the roll-out in their systems
Secondary Actors	Governance Committee – oversees the roll-out
Frequency	• Ad-hoc; or • Periodic.
Trigger	Release of an updated version of the data standards
Comment(s)	For releases with low impact (e.g., regular releases of reference data collections), roll-out might be done using a fixed script, while for releases with higher impact (e.g., restructuring in a data model), a detailed implementation plan needs to be developed and agreed with stakeholders.

Flow:

Step	Description	Actor
1	Monitor roll-out	Operational Team
2	Apply changes in local systems	Stakeholders
3	Report progress to Governance Committee and Stakeholders	Operational Team

For **changes in data models**, two situations can occur:

- Changes are not backward compatible. This situation arises when there are fundamental rearrangements in the data model or changes in existing elements.
- Changes are backward compatible. This situation arises when minor changes to the data model are made, such as addition of new elements that do not affect existing model elements.

In case changes are not backward compatible and cannot work with the software that used a previous version of the model or schema, the propagation of these changes need to be accompanied by a software upgrade process. Especially in cases where multiple software vendors are involved, such upgrades need to be carefully planned

and executed with ample time for testing and verification. To avoid disruption of the operational system, testing and verification should be conducted in a separate test and acceptance environment.

For changes that are backward compatible, the process does not rely on all systems in the operational environment installing the changes at the same time. Existing systems can continue to operate unchanged, but before they upgrade they will not be able to access functionality that is provided by the new model elements. This means that, in the environment of interconnected systems, the new functionality will become available gradually over a certain period of time. To maintain interoperability, two conditions need to be met:

- Systems that still operate with the old version of the model need to be able to ignore the additional model elements in the new version of the model; and
- Systems that have already upgraded to the new version of the model need to be able to process data using both versions of the model.

Even in the case of backward compatibility, it is recommended to organise the upgrade across the network as a well-planned and well-communicated project so that all communication partners are aware of the status of the propagation of the new functionality across the network at all times during the transition period.

The way that **changes in reference data** affect interoperability, and therefore the way those changes propagate, depends on the technical implementation.

If the reference data is implemented as an enumerated list of string values in an XML schema, changes in reference data are in fact changes in the metadata model and therefore the approaches described in the previous section apply.

Otherwise, if reference data is implemented in a Linked Data approach, for example as a SKOS Concept Scheme, every item in the collection is identified by a URI. If implemented this way, changes in reference data are generally non-disruptive. Using the example of a collection of references to organisations that participate in a network, the following changes may occur:

- Addition of an organisation to the network: the addition of a new item in the reference data can be done without disruption as long as systems ignore items that they do not recognise. The new item will be identified by a new URI that enables all systems in the network to access the new item and its characteristics whenever they need it.
- Deletion of an organisation from the network: in general it is not a good idea to delete items that are no longer needed. As long as a certain item is used as a reference in instance data, physically deleting the item from the reference collection would make that instance data invalid. As discussed in the previous section, a better approach is to give the item that is no longer needed a status of ‘deprecated’ or ‘withdrawn’ so that further use is discouraged.
- Amendment of the information of a particular organisation, such as names, addresses, etc.: if the reference data is implemented in SKOS, such changes do not affect the interoperability as these characteristics are properties of the organisation that continues to be persistently identified by its URI. However, if some of the characteristics that have changed are being used, for example

for indexing, systems that refer to the item may need to re-ingest the data for the item to be able to update the indexes.

3.2.6. *Retire a data standard*

Retire	Delete or deprecate a data standard or part(s) of it
Goal	To mark a data standard or part(s) of it as no longer relevant for applications.
Preconditions	The data standard is no longer relevant.
Success End Condition	The data standard is marked as deprecated.
Failed End Condition	N/A
Primary Actors	• Governance Committee – decides on retirement of the data standard or part(s) of it • Operational Team – takes actions to delete or deprecate
Frequency	Ad-hoc
Trigger	Observation that a data model or a reference dataset is no longer in use.
Comment(s)	Before retiring a data standard, a complete impact analysis should be done to verify that indeed the metadata is no longer used in production environments. In general, it is recommended not to physically delete a data standard but to mark it as deprecated.

Flow:

Step	Description	Actor
1	Assess the impact of deprecation	Operational team
2	Review for approval	Governance Committee
3	Approach all consumers of the data standard	Operational team
4	Clearly mark data standard as deprecated	Operational team
5	Ensure backwards compatibility	Operational team

4. SPECIFICATIONS FOR TOOLS FOR THE MANAGEMENT OF DATA STANDARDS

In this chapter, we propose a set of existing tools for supporting the management of data standards. As there exist tenths of tools for this purpose, we have taken into account the following scope criteria in order to create this list:

- **Tools that implement standards**: tools that are based on standards are more likely to reduce ICT vendor lock-in [8][26];
- **Tools that are already used by public administrations**: tools, either commercial or open-source ones, that are already used, have a proven value and can be more beneficial because they are standards within an existing ecosystem; and
- **Open-source tools**: although there is no policy that mandates the use of open-source software tools, it is often recommended because it can allow contributions by the public sector to be used by others. The European Interoperability Framework (EIF), for example, recommends openness in developing software systems *allowing European public administrations to generate results that can be interconnected, re-used and shared, which also improves efficiency* [9][27].

During the pilots with DG COMP [4], DG MARE [5], and the Publications Office of the EU [6], we had the opportunity to try many of the enlisted tools. However, we acknowledge that the list is not complete, as a complete market analysis was not in the scope of this work. Public administrations will evaluate and select appropriate tools depending on their own contextual criteria.

The tools mentioned in this chapter should support stakeholder requests and needs, but also existing standards for data standards management. An overview of such standards for data standards representation is provided in annex III and IV. Before looking at standards, we should note that, in the context of this report, the following categories of tools should be considered:

- **Tools for data modelling:** to support the design and creation of data models;
- **Tools for editing reference data**: to create, edit, harmonise and document reference data;
- **Tools for change management**: to manage changes and new releases of data standards;
- **Tools for deployment**: to support the deployment of data standards in production systems and/or their retirement; and
- **Tools for publication:** to support the documentation of data standards and its publication on authoritative sources.

[26] European Commission, Staff Working Document. (2013). *Guide for the procurement of standards-based ICT — Elements of Good Practice. Against lock-in: building open ICT systems by making better use of standards in public procurement*
https://ec.europa.eu/digital-agenda/node/67038

[27] European Commission, ISA Programme (2010). European Interoperability Framework. http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

4.1. Stakeholder requests and needs

The first step towards identifying sets of tools which could support the management and governance of data models, consisted of eliciting the stakeholder requirements for such tools. Through pilots with DG MARE [5] and DG COMP [4], the requirements as enlisted in Table 6 were identified.

Table 6 - Requirements for data standard management tools

ID	Requirements	Scope	Benefits
Modelling and visualisation			
T1a	Support the collaborative design of data models or reference data.	**DM**	Allowing collaboration by delegating responsibilities for the development of data standards, will increase the quality of the data models and time to market. Visualising data models, publishing documentation and guidelines and searching for classes and properties, helps the implementers of data models to optimise their development time.
T1b	Support search functionalities at the level of classes and properties via a graphical user interface.	**DM, RD**	
T1c	Support multilingual representations of data models or reference data.	**DM, RD**	
T1d	Support ordering concepts in a hierarchical concept scheme.	**RD**	
T1e	Support the publication of visual and browsable representations of data models and reference data	**DM, RD**	
Accessibility and interoperability			
T2a	Support import functionalities of data models or reference data in standard formats.	**DM, RD**	Achieve interoperability. Allowing integration of tools, creates the opportunity to automate the data standards management lifecycle.
T2b	Support export functionalities of data models or reference data in standard formats.	**DM, RD**	
T2c	Support the identification of resources (classes, properties, concepts, code list values, etc.) using persistent and resolvable HTTP URIs.	**DM, RD**	
T2d	Support file-based read-access over HTTP for reference data	**RD**	
Access restriction			

T3	Support granular access control management for reference data and data model repositories.	**DM, RD**	Allowing increased control of the data standards development process, leads to increased quality.
Change and release management			
T4a	Support the collection of change requests and the documentation of impact analyses and release notes.	**DM, RD**	Achieve interoperability between solutions of reference data consumers. Allowing increased control of the reference data development process, leads to increased quality of the reference data.
T4b	Support versioning and workflow functionalities in order to manage changes to reference data and data models in a collaborative and controlled manner.	**DM, RD**	
T4c	Support the (semi-)automatic notification of upcoming changes to users of reference data and data models	**DM, RD**	
Reference data deployment			
T5a	Support deploying reference data as a service.	**RD**	Fetching from authoritative sources a limited amount of reference data, leads to a faster access to and processing of reference data.
T5b	Support a publish-subscribe mechanism for reference data consumers.	**RD**	
Harmonisation of data standards			
T6	Support mapping functionalities for mapping reference data and data model terms with other sources.	**DM, RD**	Achieve harmonisation.

4.2. Tools for data standards management

In order to identify a set of tools that are able to support the governance and management specifications of the previous sections, the requirements enlisted in the previous section were checked against the tools enlisted in Table 7.

The mapping between the functionalities of the tools and the stakeholder requirements is detailed in the spreadsheet below.

Analysis of too

Table 7 – Tools for governance and management of data standards

Tool	Licence	Description
Free or Open Source		
Re3Gistry[28]	EUPL	The *Re3Gistry* is developed and used by the Joint Research Centre (JRC) of the European Commission to publish the INSPIRE registry.
VocBench[29]	Open source	*VocBench* is developed and used by the Food and Agriculture Organisation of the United Nations (FAO). The tool is used to manage the FAO's multilingual agricultural thesaurus "AGROVOC". Moreover, the tool is adopted by a growing number of user communities, including FAO's Fisheries and Aquaculture Department, the data.fao.org project, the European Commission Publications Office, the European Environment Agency, the Italian Senate and others.
Skosmos[30]	Open source	Skosmos is a web-based tool providing services for accessing controlled vocabularies. Currently, Skosmos is used by several organisations, including Finto, the FAO, the Rhineland-Palatinate spatial data initiative classifications and the University of Oslo Library thesauri.
Joinup[31]	Open Source	Joinup was developed under the ISA programme. It is a platform for eGovernment professionals to share their experiences and solutions for public administrations, in the domain of semantic interoperability.
ArgoUML[32]	Open source	*ArgoUML* is a Windows-based UML modelling tool
Umbrello[33]	Open source	*Umbrello* is a Windows- and Linux-based UML modelling tool
Open ModelSphere[34]	GPL	*Open ModelSphere* is a Windows-based UML modelling tool
Draw.io[35]	Free SaaS	Draw.io is a web-based UML modelling tool, which is also available as a plugin for Confluence and JIRA.

[28] https://joinup.ec.europa.eu/software/re3gistry/description
[29] http://vocbench.uniroma2.it/
[30] http://skosmos.org/
[31] https://joinup.ec.europa.eu/
[32] http://argouml.tigris.org/
[33] https://umbrello.kde.org/
[34] http://www.modelsphere.com/org/

[35] https://www.draw.io/

Tool	Licence	Description
Apache Subversion (SVN)[36]	Open Source	*SVN* is a software versioning and revision control system, used to maintain current and historical version of files. The tool is used by European Commission as a source control system for internal software development projects.
CDMMD[37]	Open Source	The *Core Data Model Mapping Directory (CDMMD)* facilitates the creation and publication between the Core Vocabularies and other (core) data model initiatives.
Commercial		
Talend[38]	Commercial	*Talend* is a master data management tool, used by DG MARE for reference data management in the fisheries domain. Talend is also used by over 1700 enterprise customers, such as Allianz, BNP Paribas Real Estate and Groupon.
Collibra Reference Data Accelerator[39]	Commercial	The Collibra *Reference Data Accelerator* is used by organisations in different sectors for managing reference data. Collibra's customers are active in the public, healthcare and industrial sectors.
Software AG webMethods OneData[40]	Commercial	Software AG's *webMethods OneData* is a tool to reconcile, cleanse and synchronize any master data. The tool is used in public and private sectors, including the U.K. Ministry of Defence.
Orchestra EBX5[41]	Commercial	Orchestra *EBX5* is a tool for master and reference data management that supports a wide range of master data management needs: from data modelling to distribution.
PoolParty[42]	Commercial	PoolParty *Thesaurus Server* is a web-based tool to create domain-specific vocabularies or thesauri. PoolParty is used in different organisations, including the European Commission, The World Bank and Roche.

36 http://subversion.apache.org/
37 http://mapping.semic.eu/
38 https://www.talend.com/
39 https://www.collibra.com/data-governance-center/reference-data-accelerator/
40 http://www.softwareag.com/corporate/products/webmethods_integration/integration/products/mdm/overview/default.asp
41 http://www.orchestranetworks.com/product/

42 https://www.poolparty.biz/

Tool	Licence	Description
Confluence and JIRA[43]	Commercial	Confluence and JIRA are both online tools that support the execution of projects and the communication between its stakeholders. Confluence is an online collaboration platform, where team members can create, organise and discuss work. JIRA is an issue tracker designed for software teams. The tools are integrated and used by EC as a platform for communication, collaboration, and project management.
IBM InfoSphere[44]	Commercial	InfoSphere Master Data Management Reference Data Management (MDM RDM) Hub provides key functions including: • Role-based user interface with security and access control including integration with LDAP • Management of reference data sets and values • Management of mappings and relationships between reference data sets • Importing and exporting of reference data in CSV and XML format through both batch and user interface • Versioning support for reference data sets and mappings • Change process controlled through configurable lifecycle management • Hierarchy management
MagicDraw[45]	Commercial	*MagicDraw* is a UML business process, architecture, software and system modelling tool with teamwork support.
SAP PowerDesigner[46]	Commercial	SAP Sybase PowerDesigner is a graphical enterprise modelling solution supporting standard methodologies and notations and providing automated code reverse engineering and generation through customizable templates.
Enterprise Architect[47]	Commercial	Enterprise Architect is a largely adopted modelling tool that provides full life cycle modelling for: • Business and IT systems • Software and Systems Engineering • Real-time and embedded development

[43] https://www.atlassian.com/software/confluence , https://www.atlassian.com/software/jira
[44] http://www-01.ibm.com/software/data/infosphere/
[45] http://www.nomagic.com/
[46] http://www.powerdesigner.de/en/
[47] www.sparxsystems.com/products/ea/

Tool	Licence	Description
Coreon[48]	Commercial	Coreon provides flexible tools for creating, exploring, and maintaining taxonomies with terminology management. The tools were used by ESTeam AB (esteam.se) during the European LISE project.
TopBraid Reference Data Manager[49]	Commercial	TopBraid RDM is a tool that supports the governance and provisioning of reference data including the curation of reference datasets (code-lists) and enriching them with comprehensive metadata.
Internally developed solutions		
REDDA[50]	DG COMP	*REDDA* is a web-based tool which is developed and used by DG Competition to manage reference data for their State Aid Notification systems.
CS/RD 2	DG TAXUD	The CS/RD application is developed by DG TAXUD to manage reference data that is used across different Member States of the EU. The tool offers the possibility for the users of the application to consult and maintain this reference data, as well as distribution functionalities to operational systems.

4.2.1. *Example of a tool chain for the governance and management of data standards*

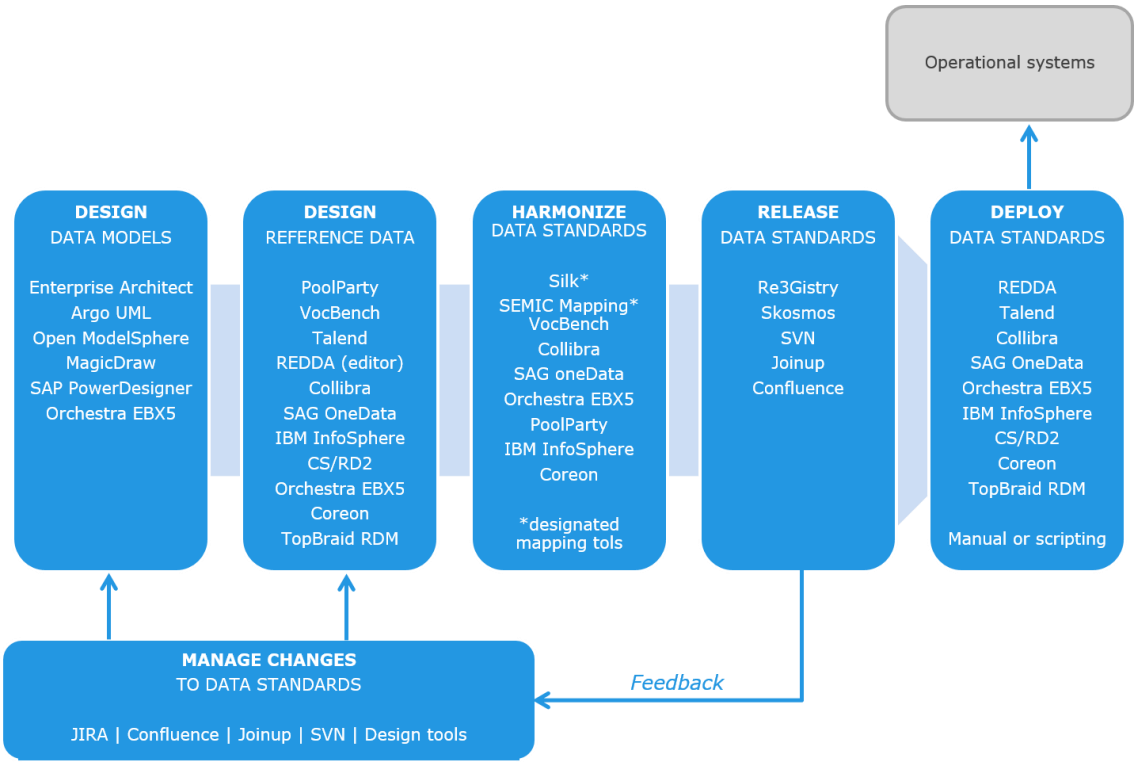
This section proposes a set of tools that can be combined and used together for managing data standards. All requirements can be supported by existing tools, some of which are already being used within the European Commission. Some tools can support the full lifecycle (i.e., all management process as in the case of the big commercial solutions), while others need to be combined as they do not support all steps. In order to identify which set of tools can be used to support the management and governance of data standards, the tools are categorised in a tool chain following the management processes described in section 3.2 and as displayed in Figure 5.

Many of the tools listed below are already used by public administrations, including the European Commission. In some cases, these tools are even used in conjunction already. We believe that the mentioned tools are sufficient to cover the immediate needs and requests for the management of data standards. Therefore, it would not be necessary to define a new set of specific functional requirements to build a new tool. No single open source tool exists or is likely to be build that can cater for all requirements. A collection of different tools needs to be deployed. It is important that

[48] https://www.coreon.com/
[49] http://www.topquadrant.com/docs/5.0.0/1802798.html
[50] https://joinup.ec.europa.eu/asset/rd/asset_release/genis-reference-data-interoperability-solution

these tools are interoperable; the proposed standards for metadata representation in annex IV may contribute to achieving this.

Figure 5 – Tool chain for managing data standards



5. CONCLUSION AND NEXT STEPS

5.1. Conclusion

The ever-increasing volume of information exchanged within and between different organisations at both national and EU level requires setting up solutions that facilitate its automatic processing. Whilst technological developments offer various means to automate the exchange of information, technological developments alone cannot guarantee a greater interoperability between information systems. Thus, a fundamental aspect in the interoperability of public administrations is the need for common data standards (such as data models and reference data).

As explained in this report, by setting up a proper governance and management of data standards and by employing the right tools, public administrations greatly enhance their potential for coordination and interoperability and ultimately the possibilities for sharing and re-use of data standards thanks to:

- Reduced semantic interoperability conflicts;
- An increased quality and traceability of the information/data exchanged;
- A greater re-use of data standards;
- A reduced risk of duplication;
- An increased trust towards the information to be exchanged; and
- Savings derived from the re-use of already existing information.

In this vein, this report, commissioned by the ISA Programme described a model for the coordination of governance and management of data standards for the EU institutions and public administrations in the Member States. It is based on the results of a survey on data standards requirements and existing solutions in the EU institutions and the Member States. The report builds on previous work carried out in collaboration with the Directorate-General for Competition and the Directorate-General for Maritime Affairs and Fisheries. Both Directorate-Generals intend to set up mechanisms for governing data standards. The Publications Office of the EU was also involved in this task, as it is experienced in the governance and management of data standards in the context of the legal decision-making process of the EU institutions.

During the preparation of this report, existing practices, solutions and tools for the governance and management of data standards were identified based on stakeholder requests complemented with findings from the study and implementation experience gained in the context of ISA Action 1.1. Such existing solutions for the governance of data standards include the Inter-institutional Metadata Maintenance Committee (IMMC) process and the Metadata Registry of the Publications Office, the INSPIRE maintenance and implementation group and the Dutch Knowledge and Exploitation Centre Official Government Publications (KOOP).

Based on the description of existing governance models, the requirements, and the feedback received from stakeholders, the specifications for governance and management of data standards have been provided by enlisting a number of phases which together compose the lifecycle management process for data standards.

The output of the tasks performed in the study can be summarised as following:

1. A specification of the requirements for managing the lifecycle of data standards in the EU institutions;
2. A proposed governance structure for data standards, including principles and roles and responsibilities;
3. A proposal for tools and best practices for managing and consuming data standards. In particular, this report contains a useful tools chain for managing data standards.

5.2. Next steps

We outline three directions of future work related to the governance and management of data standards:

1. Raise awareness of public administrations about the importance of proper governance an management of data standards;
2. Support the adoption and implementation of governance and management practices for data standards;
3. Develop new or integrate existing tools in order to support the governance and management of data standards.

Raise awareness

As explained earlier, a number of stakeholders from EU Member States and EU Institutions have contributed to this work. We observed that one of the main roadblocks to the adoption of proper governance and management practices is in fact the lack of awareness. Stakeholders reported that it is often hard to persuade top management within public administration to invest in such activities as their knowledge about the importance and the benefits of such activities is quite limited. Hence, emphasis should be put on raise awareness and promotion activities, which will disseminate within organisations the benefits of governance and management of data standards, and all key stakeholders should be engaged timely in the process. For example, senior management within an organisation would be definitely interested to know that the coordinated re-use of reference data can reduce software development and maintenance costs, as well as the potential data exchange conflicts between different systems that partake in cross-border and/or cross-sector information exchanges.

Implementation support

The next steps should be for public administrations to implement and tailor the proposed generic governance and management models and deploy tools to support them. The applied methodology should fit the context of the public administration that adopts it and should match the scope of use. One aspect to adapt to a local context deals with the adoption of the lifecycle management phases. While none of the phases should be removed from the lifecycle management process, there can be various degrees of formalism for their implementation. For example, the management of a generic, context-neutral data model that defines a rather small number of concepts will less likely require the implementation of a high formal way

to manage change requests or the retire phase, compared with a domain data model defining a large number of concepts. Another aspect that should be taken into consideration when adopting the management methodology deals with selecting the appropriate tools to support the processes. For example, when the management of change requests is very important, a tool such as JIRA can offer all the necessary benefits, while a simple exchange of emails between the stakeholders might better suit a situation when the management of change requests is less vital. It all depends on the context and scope of use.

In order to support further the implementation of governance and management practices for data standards, the Community of Practice on data standards, supported by the ISA Programme, is currently working on a set of common implementation guidelines, based on existing good practices and lessons learned coming both from EU public administrations and from third countries.

Tool support

This report indicates that existing tools are readily available and already used within the EU institutions and the Member States to support the governance and management data standards. Commercial solutions usually support most, if not all, management processes, which open source ones are usually more process-specific and have to be combined in order to support the full lifecycle. We believe however that the market covers quite well the needs of governance and management of data standards, and only minor context-specific customisations are required, e.g. to support more advanced access management.

We observed though that many of those tools are not based on open standards, or do not provide open APIs and interfaces, which results in a data and vendor lock-in. In this vein, the need for strengthening their integration using the open standards listed in this report is imperative.

BIBLIOGRAPHY

[1] DAMA International, The DAMA Guide to the Data Management Body of Knowledge (DAMA DMBOK), Bradley Beach: Technics Publications, 2010.

[2] J. Jordan and C. Ellen, "Business need, data and business intelligence," *Journal of Digital Asset Management*, pp. 10-20, 2009.

[3] European Commission, ISA Programme, "D4.1 – Metadata management requirements and existing solutions in EU Institutions and Member States," European Commission, Brussels, 2013.

[4] European Commission, ISA Programme, "D4.3. Report on the implementation of a Metadata Management pilot for DG COMP," European Commission, Brussels, 2014.

[5] European Commission, ISA Programme, "D4.4. Report on the implementation of a Metadata Management pilot for the Common Information Sharing Environment (CISE)," European Commission, Brussels, 2014.

[6] European Commission, ISA Programme, "D5.4. Metadata alignment pilot in the EU institutions and MSs," European Commission, Brussels, 2014.

[7] European Commission, ISA Programme, "Description of a change management release and publication process," European Commission, Brussels, 2015.

[8] European Commission, Staff Working Document, "Guide for the procurement of standards-based ICT — Elements of Good Practice," 2013.

[9] European Commission, ISA Programme, "European Interoperability Framework (EIF) for European public services," European Commission, Brussels, 2010.

[10] ISA Programme - European Commission, "An introduction to the European Interoperability Reference Architecture v0.9.0 (EIRA)," 31 August 2015. [Online]. Available: https://joinup.ec.europa.eu/asset/eia/asset_release/eira-release-v090-beta.

[11] European Commission, ISA Programme, "D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC," 2012. [Online]. Available: https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris/.

[12] G. Shukair, N. Loutas, V. Peristeras and S. Sklarß, "Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema," *Computers in Industry*, no. 64, pp. 10 - 18, 2013.

[13] P. Archer, M. Dekkers, S. Goedertier and N. Loutas, "Study on business models for Linked Open Government Data (BM4LOGD)," European Commission, ISA Programme, 2013.

[14] S. Bechhofer and A. Miles, "SKOS Simple Knowledge Organization System Reference," W3C, 2009.

[15] International Organization for Standardization, "ISO/IEC 11179-1:2004 - Metadata registries," 2009.

[16] OASIS, "Universal Business Language v2.0," 12 December 2006. [Online]. Available: http://docs.oasis-open.org/ubl/os-UBL-2.0/UBL-2.0.html. [Accessed 29 November 2013].

[17] OASIS, "OASIS Universal Business Language (UBL) TC," 2013. [Online]. Available: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl. [Accessed 29 November 2013].

[18] G. Hohpe and B. Woolf, Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions, Addison-Wesley, 2003.

[19] "Official Journal of the European Union," 27 June 2013. [Online]. Available: http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2013:175:SOM:EN:HTML. [Accessed 02 December 2013].

[20] W3C, "Linked data," 2013. [Online]. Available: http://www.w3.org/standards/semanticweb/data. [Accessed 02 December 2013].

[21] European Commission, "Official documents - Employment, Social Affairs & Inclusion - European Commission," 2013. [Online]. Available: http://ec.europa.eu/social/main.jsp?catId=868&langId=en. [Accessed 17 December 2013].

[22] European Commission, ISA Programme, "D6.1.2 – Report for documenting and reusing data models and reference data Business Case," Brussels, 2013.

[23] A. Coates and M. Watts, "Code List Representation (Genericode) Version 1.0," OASIS, 2007.

[24] P. De Leenheer, A. de Moor and S. Christiaens, "Business Semantics Management at the Flemish Public Administration," 2010.

[25] Object Management Group, Inc., "Unified Modeling Language™ (UML®)," October 2012. [Online]. Available: http://www.omg.org/spec/UML/. [Accessed 20 November 2013].

[26] Government On-Line Metadata Working Group, “Records Management Application Profile,” Government of Canada, Canada, 2006.

[27] U. B. L. H. O. H. M. K. M. G. L. R. F. S. Y. R. A. S. Fabian Büttner, “Model-driven Standardization of Public Authority Data Interchange,” 2013.

[28] ECN, “Joint Statement of the Council and the Commission on the Functioning of the Network of Competition Authorities,” [Online]. Available: http://ec.europa.eu/competition/ecn/joint_statement_en.pdf. [Accessed 23 12 2013].

[29] IBM, “Reference Data Management: IBM Redbooks Solution Guide,” International Business Machines Corporation, New York, 2013.

[30] W.-J. Chen, J. Baldwin, T. Dunn, M. Grasselt, H. Shabbar, D. Mandelstein, I. Milman, E. A. O'Neill, S. Pandit, R. Tamlyn and F. Xu, A Practical Guide to Managing Reference Data with IBM InfoSphere Master Data Management Reference Data Management Hub, International Business Machines Corporation, 2013.

[31] M. Mosley, M. Brackett, S. Earley and D. Henderson, The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide), New Jersey: Technics Publications, LLC, 2009.

[32] International Organisation for Standardisation, ISO 19135:2005 Geographic information -- Procedures for item registration, Geneva, 2005.

[33] OMG, “Business Process Modelling Notation 2.0 (BPMN),” 01 2011. [Online]. Available: http://www.omg.org/spec/BPMN/.

[34] European Union, *Regulation no. 182/2011 laying down the rules and general principles concerning mechanisms for control by Member States of the Commission's exercise of implementing powers,* 2011.

[35] A. Miles and S. Bechhofer, “SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant,” 18 August 2009. [Online]. Available: http://www.w3.org/TR/skos-reference/skos-xl.html. [Accessed 3 April 2014].

[36] Bizer, Heath and Berners-Lee, “Linked Data - The Story So Far,” *International Journal on Semantic Web and Information Systems,* pp. 1-22, 2009.

[37] T. Berners-Lee, “Linked Data,” 27 July 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html. [Accessed 02 December 2013].

[38] United Nations - Centre for Trade Faciliation and Electronic Business, "Core Components Technical Specification - Version 3.0," 2009.

[39] Publications Office of the European Union, "Proposal for metadata governance on interinstitutional level," 2011.

[40] D. Roy, "National Information Exchange Model (NIEM): Technical Introduction to NIEM".

[41] Spanish Ministry of Finance and Public Administration, *Decision of 19 February 2013 of the secretary of state for public administration approving the technical interoperability standard for the reuse of information resources.*

[42] European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions,* Brussels, 2011.

[43] NIEM, "NIEM Tools Catalog," 2013. [Online]. Available: https://www.niem.gov/tools-catalog/Pages/tools.aspx. [Accessed 25 11 2013].

[44] European Commission, "Evaluation of the 2004 action plan for electronic public procurement," European Commission, Brussels, 2010.

[45] Digitaliseringsstyrelsen, "About OIOXML," 30 May 2012. [Online]. Available: http://www.digst.dk/Servicemenu/English/IT-Architecture-and-Standards/Standardisation/Standardisation-creating-digital-Denmark/About-OIOXML. [Accessed 22 November 2013].

[46] UN/CEFACT, *Core Components Business Document Assembly Technical Specification,* 2012.

[47] UN/CEFACT, "Core Components Technical Specification - Version 3.0," 2009.

[48] UN/CEFACT, "Core Components Business Document Assembly - Technical Specification - Version 1.0," 2012.

[49] UN/CEFACT, "XML Naming and Design Rules Technical Specification - Version 3.0," 2009.

[50] UN/CEFACT, "UML Profile for Core Components (UPCC)," 2008.

[51] Portal Administración Electrónica, "Technical standards for interoperability," 2013. [Online]. Available: http://administracionelectronica.gob.es/pae_Home/pae_Organizacion/pae_DGMAPIAE.html?idioma=en. [Accessed 09 December 2013].

[52] European Commission, “INSPIRE Directive,” 2013. [Online]. Available: http://inspire.jrc.ec.europa.eu/. [Accessed 10 December 2013].

[53] Interactive Instruments, “Beyond service interfaces - OGC encoding standards in INSPIRE: GML and SLD/SE,” Bonn, 2011.

[54] UN/CEFACT, “Standard Business Document Header - Technical Specification,” European Commission, 2004.

[55] Council of the European Union, “Council Decision 2009/316/JHA of 6 April 2009 on the establishment of the European Criminal Records Information System (ECRIS) in application of Article 11 of Framework Decision 2009/315/JHA,” *Official Journal L 093,* pp. 33-48, 2009.

[56] European Commission , “Commission Regulation (EU) No 1213/2010 of 16 December 2010 establishing common rules concerning the interconnection of national electronic registers on road transport undertakings Text with EEA relevance,” *Official Journal of the European Commission,* pp. 21-29, 2010.

[57] CooP, “Final Report of Work Package 5: Specifications of Common Data Formats and Semantics,” 2014.

[58] European Commission, “Commission Decision of 12 December 2011 on the reuse of Commission documents (2011/833/EU),” *Official Journal of the European Union,* pp. 39-42, 2011.

[59] European Community, “European Union Public Licence v.1.1,” 2007. [Online]. Available: https://joinup.ec.europa.eu/system/files/EN/EUPL%20v.1.1%20-%20Licence.pdf. [Accessed 11 02 2014].

[60] P. Uhrowczik, “Data dictionary/directories,” *IBM Systems Journal,* pp. 332-350, 1973.

[61] International Organization for Standardization, “ISO/IEC 11179-1:2004 - Metadata registries,” 2009.

[62] International Organization for Standardization, ISO/IEC 11179-6:2005 - Metadata registries, part 6: Registration, 2005.

[63] M. Dekkers and S. Goedertier, “Metadata for Public Sector Administration,” NISO/DCMI, 2013.

[64] National Information Standards Organization, “Understanding Metadata,” 2004.

[65] European Commission, ISA Programme, D4.2. Methodology and tools for Metadata Governance and Management for EU Institutions and Member States, Brussels, 2014.

[66] European Commission, ISA Programme, *eGovernment Core Vocabularies: The SEMIC.EU approach,* Brussels: European Commission, 2011.

[67] European Commission, ISA Programme, "ISA Open Metadata Licence v1.1," Brussels, 2013.

[68] European Commission, ISA Programme, "Metadata Management Survey Results," 2012.

[69] European Commission, ISA Programme, "Process and methodology for developing semantic agreements," 2013.

[70] European Commission, ISA Programme, "Towards harmonised governance and management of data models and reference data - Business case," Brussels, 2013.

[71] European Commission, ISA Programme, "Towards Open Government Metadata," Brussels, 2011.

[72] R. S. Smith, M. Lutz, A. Perego and R. Sgaolin, "Building a missing item in INSPIRE: The Re3Gistry," European Commission - Joint Research Center, Ispra, 2013.

[73] European Commission - Joint Research Centre, "Re3gistry software v0.3," European Commission, Ispra, 2014.

ANNEX I: GLOSSARY

The table below provides common definitions used throughout this report.

Table 1 – Glossary

Term / Acronym	Description
Data model	A data model is a collection of entities, their properties and the relationships among them, which aims at formally representing a domain, a concept or a real-world thing.
Domain	Domain is a specific subject matter area that has government body i.e. ministry or department responsible for that domain e.g., the Ministry of Agriculture, the Ministry of Finance.
Domain model	A domain model is a conceptual view of a system or an information exchange that identifies the entities involved and their relationships.
ICT	Information and Communication Technology
Interoperability	According the ISA Decision, interoperability means the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the *exchange of data* between their respective ICT systems.
IT	Information Technology
Metadata	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 or information about information. (National Information Standards Organization, 2004)
Metadata governance	Metadata governance comprises well-defined roles and responsibilities, cohesive policies and principles, and decision-making processes that define, govern and regulate the lifecycle of metadata.
Metadata management	We define metadata management as the good practice of adopting policies, processes, and systems to plan, perform, evaluate, and improve the use and re-use of data models and reference data.
Metadata alignment	Metadata alignment is the harmonisation of data standards either by forging a wide consensus on the use of a common specification for data standards or through the creation of mappings between terms of two or more specifications.
Reference data	Reference data is small, discrete sets of values that are not updated as part of business transactions, but are usually used to impose consistent classification. Reference data normally has a low update frequency. Reference data is relevant across more than one business systems belonging to different organisations and sectors.
Data standards	A data standard is a structural metadata specification that describes or defines other data [ISO111179]. Data standards indicate how compound objects are put together [NISO]. It can consist of, among others, data models, reference data, and identifier schemas. [10]

Term / Acronym	Description
Stakeholder	Anyone who is materially affected by the proposed solution. Effectively solving any complex problem involves satisfying the needs of a diverse group of stakeholders. Stakeholders will typically have different perspectives on the problem, and different needs that must be addressed.
Stakeholder request	An explicitly formulated statement by a stakeholder about a requirement the federation must satisfy.
Stakeholder need	A business or operational problem (opportunity) that must be fulfilled in order to justify purchase or use of the federation. Also known as goal or objective.

ANNEX II: EXISTING SOLUTIONS FOR THE GOVERNANCE OF DATA STANDARDS

This section contains an overview of existing solutions for metadata governance. These solutions serve as inspiration for the specifications described in section 2.2.

A. The Interoperability Solutions for European Public Administrations (ISA) Committee

The European Commission is assisted in the implementation of the Interoperability Solutions for European Public Administrations (ISA) Programme by the **ISA Committee**, which represents the Member States. Furthermore, the **ISA Coordination Group**, nominated by the **ISA Committee**, ensures continuity and consistency at working level.

In the past, the ISA Coordination Group has endorsed data standards such as the Core Vocabularies[51]. This governance body may be useful for taking high-level decisions on voluntary, trans-European harmonisation initiatives on data standards.

B. Inter-Institutional Metadata Maintenance Committee (IMMC)

The Inter-Institutional Metadata Maintenance Committee (IMMC)[52] is responsible for the decisions related to key reference data and data models used in the legal decision-making process of EU institutions and the EU Open Data Portal (ODP)[53].

The IMMC is part of a three-level organisational structure, consisting of an inter-institutional steering committee, a metadata maintenance committee and a metadata registry. The governance methodology applied by the IMMC meets most requirements for inter-institutional governance. However, currently it is limited to inter-institutional governance in the area of the legal decision-making processes of the EU and open data. Given its inter-institutional character, it does not offer the possibility for trans-European governance. Nevertheless, its structure and main principle can be re-used by other entities that want to set up a trans-European or local data standards governance mechanism.

C. INSPIRE Maintenance and Implementation Group (MIG)

The purpose of the INSPIRE Directive is "to lay down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Community". Maintenance and evolution of INSPIRE is governed by the INSPIRE Maintenance and Implementation Framework (MIF)[54].

The central role in the governance of data standards management is the INSPIRE Maintenance and Implementation Group (MIG) which is responsible for strategy related to the implementation of INSPIRE. It is chaired by The Joint Research Centre

[51] Joinup (30 May 2012), ISA Member State representatives endorse key specifications for e-Government interoperability, https://joinup.ec.europa.eu/node/48837

[52] Publications Office. Proposal for metadata governance on interinstitutional level. http://publications.europa.eu/mdr/resource/core-metadata/IMMC_reu3_adoption_anx3.pdf_A-2011-764293.pdf

[53] European Union Open Data Portal. https://open-data.europa.eu/

[54] European Commission. INSPIRE. Maintenance and Implementation. http://inspire.ec.europa.eu/index.cfm/pageid/5160/list/mif

(JRC) of the European Commission and composed of two representatives per country. The INSPIRE Regulatory Committee in which the Member States are represented advises the European Commission on the adoption of the Implementing Rules. Any decisions that require a change in the INSPIRE Regulation are formally taken by the European Commission, the European Parliament and the Council under the Comitology procedure (see Table 3 in section 2.2.2.1).

The MIG is complemented by a pool of experts drawn from the stakeholder community. The experts in this pool are called upon when MIG sub-groups are formed to address specific implementation or maintenance issues, but will also provide the opportunity to reach out to experts involved or interested in particular aspects of INSPIRE implementation or maintenance.

D. Germany: IT Planning Council and KoSIT

Since 2009, article 91c of the Basic Law (Grundgesetz), the Constitution of Germany establishes the basis for cooperation between the federal level (Bund) and the state level (Länder) on the implementation and interoperability of IT solutions.

In 2010, the IT Planning Council (IT-Planungsrat)[55] was established to coordinate the collaboration between the federal and state levels. The members of the council are the federal state secretary for IT and representatives from the states. In addition, three representatives from local government and the responsible person on the federal level for data protection and freedom of information participate in an advisory role.

Under responsibility of the IT-Planungsrat, KoSIT[56], the Coordination Agency for IT Standards, takes care of the coordination of development and implementation of standards for data exchange. KoSIT manages the XÖV framework (XML in der öffentlichen Verwaltung – XML in public administration) and provides access to several tools, guidelines and XML schemas with code lists, data types and core components.

All organisations involved in e-Government in Germany can submit requirements for standards to KoSIT. KoSIT submits proposals to the IT-Planungsrat which, in its annual meeting, decides on standardisation proposals.

[55] IT-Planungsrat. http://www.it-planungsrat.de/DE/Home/home_node.html
[56] IT-Planungsrat. Koordinierungsstelle für IT-Standards. http://www.it-planungsrat.de/DE/Organisation/KoSIT/KoSIT_node.html

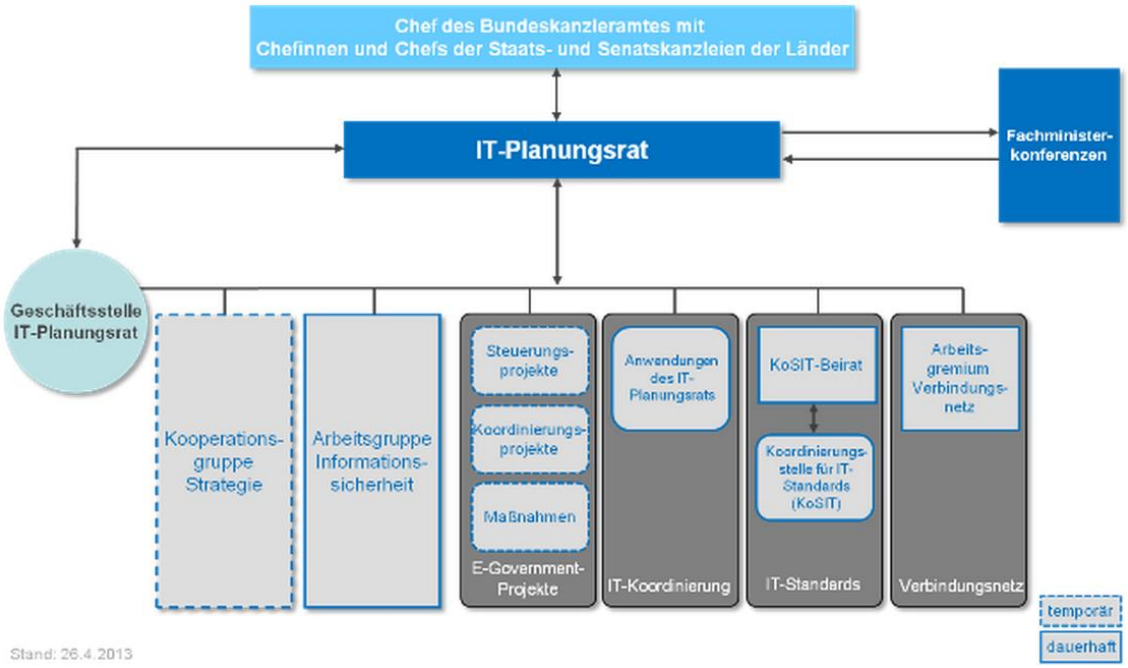


Figure 6 – Organisational structure of IT-Planungsrat

E. The Netherlands: Knowledge and Exploitation Centre Official Government Publications (KOOP)

The Knowledge and Exploitation Centre Official Government Publications (KOOP) is an autonomous unit under the Ministry of the Interior and Kingdom Relations of The Netherlands. KOOP develops and maintains products and managed services for all levels of government, including central government and provinces, water authorities and municipalities.

KOOP was initially set up to realise the conversion of the three official gazettes (Staatscourant, Staatsblad and Tractatenblad) into electronic publications with the objective to offer the official promulgation of legislation, decrees and treaties exclusively on Internet. These publications are now available at www.overheid.nl.

One of the products developed and maintained by KOOP is the Government Web Metadata Standard OWMS. This national standard, based on the Dublin Core, specifies the metadata properties to be used to provide structured descriptions of unstructured governmental information on the Web, enabling searching, discovering and presenting such information.

OWMS consists of agreements concerning[57]:

- Properties (descriptors) for describing government information;
- Lists of values to be used for the properties; and
- Syntax of the values to be used for the properties.

[57] For more information: http://standaarden.overheid.nl/owms

The governance structure is as follows[58]:

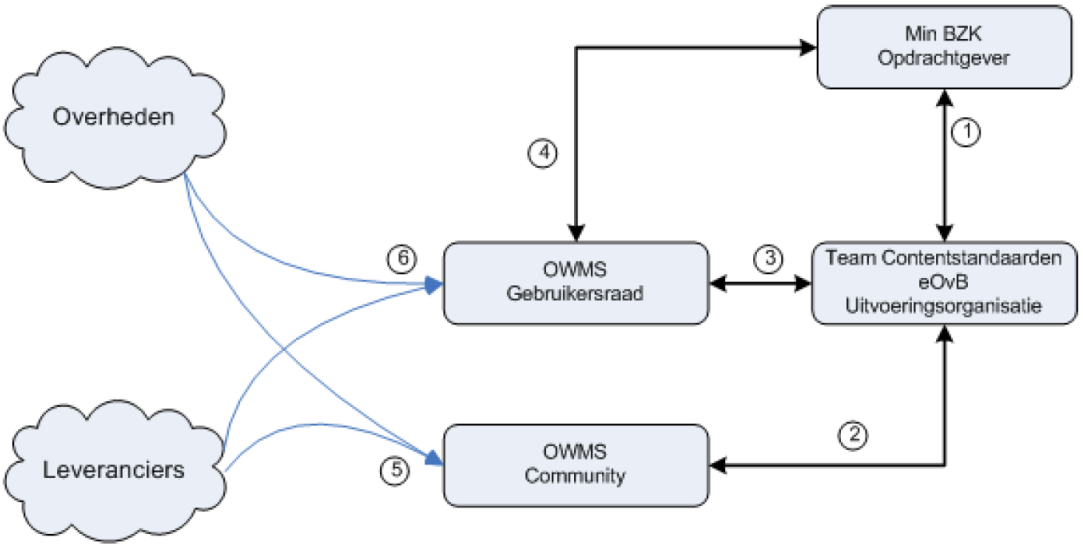


Figure 7 – Governance structure OWMS

1. The Ministry of the Interior and Kingdom Relations (MinBZK) instructs the Team Content Standards (Contentstandaarden) to develop and publish OWMS.
2. Members of the OWMS Community submit change requests to the Team Content Standards. The Team takes the request into consideration and produces a change proposal if the request is feasible and within the scope of its charter.
3. The Team Content Standards submits change proposals to the OWMS User Council (Gebruikersraad) and implements the proposals that are agreed by the User Council.
4. If changes concern normative specifications and would lead to a new version of OWMS, the User Council does not take the decision, but advises the Ministry which then decides on the changes.
5. Anyone with an interest in OWMS can become a member of the OWMS Community.
6. Membership of the OWMS User Council is open to all Government agencies (Overheden) and vendors (Leveranciers), who apply OWMS in their products and services, subject to consultation with the Ministry.

F. ISO11179-6 Metadata Registration

A general standard for the registration of metadata items is ISO/IEC 11179. As part of the six-part standard, ISO/IEC 11179-6:2005[59] specifies the procedure by which Administered Items required in various application areas could be registered and

[58] E-Overheid voor Burgers. Beheerplan OWMS. http://standaarden.overheid.nl/owms/beheer/BeheerplanOWMSv1.0.pdf

[59] ISO/IEC 11179-6:2005. Information technology -- Metadata registries (MDR) -- Part 6: Registration. http://www.iso.org/iso/catalogue_detail.htm?csnumber=35348

assigned an internationally unique identifier. This procedure includes organisations such as the Registration Authority, the Responsible Organisation, and the Submitting Organisation. It also includes roles such as the Registrar, Steward, and Submitter.

G. Data Management Body of Knowledge (DMBOK)

The Data Management Body of Knowledge (DMBOK)[60] is a general methodology for data management. The DM-BOK guide strives to adopt a generally accepted view of data management. It provides standard definitions for data management functions, roles, deliverables and other common terminology. The DM-BOK devotes an entire chapter to Reference and Master Data Management.

In terms of governance, it defines a number of Reference Data Management processes. In terms of Governance Structure, it defines a number of operational roles including the Data Architect, Business Analyst, Data Stewart, and Application Architect. It attributes all decision power onto the role of a Data Governance Council.

[60] DAMA Data Management Body of Knowledge (DAMA DMBOK). 1st Edition 2009. http://www.dama.org/i4a/pages/?pageid=3364

ANNEX III. EXISTING METHODOLOGIES FOR DATA STANDARDS MANAGEMENT

A. Standards for management

For the management of data standards, and in particular the registration of metadata in registries, several standards exist. A general standard for the registration of metadata items is ISO/IEC 11179. There are also domain-specific standards; an example is ISO 19135 for geographic information.

ISO/IEC 11179[61] specifies the kind and quality of metadata necessary to describe data, and it specifies the management and administration of that metadata in a metadata registry (MDR). It applies to the formulation of data representations, concepts, meanings, and relationships between them to be shared among people and machines, independent of the organization that produces the data. It does not apply to the physical representation of data as bits and bytes at the machine level. As part of the six-part standard, **ISO/IEC 11179-6:2005**[62] specifies the procedure by which Administered Items required in various application areas could be registered and assigned an internationally unique identifier. For each Administered Item to be registered, ISO/IEC 11179-6:2005 defines the type of information that is specified, the conditions that are met, and the procedure that is followed.

ISO 19135:2005[63] specifies procedures to be followed in establishing, maintaining and publishing registers of unique, unambiguous and permanent identifiers, and meanings that are assigned to items of geographic information. In order to accomplish this purpose, ISO 19135:2005 specifies elements of information that are necessary to provide identification and meaning to the registered items and to manage the registration of these items.

The Data Management Association's guide to the **Data Management Body of Knowledge (DMBOK)** recommends that changes to controlled vocabularies and their reference data sets be conducted by following a **change request process**:

1. Create and receive a change request;
2. Identify the related stakeholders and understand their interests;
3. Identify and evaluate the impacts of the proposed change;
4. Decide to accept or reject the change, or recommend a decision to management or governance;
5. Review and approve or deny the recommendation;
6. Communicate the decision to stakeholders prior to making the change;
7. Update the data;
8. Inform stakeholder that the change has been made.

[61] ISO/IEC 11179-1:2004. Information technology -- Metadata registries (MDR). http://www.iso.org/iso/catalogue_detail.htm?csnumber=35343

[62] ISO/IEC 11179-6:2005. Information technology -- Metadata registries (MDR) -- Part 6: Registration. http://www.iso.org/iso/catalogue_detail.htm?csnumber=35348

[63] ISO 19135:2005. Geographic information -- Procedures for item registration. http://www.iso.org/iso/catalogue_detail.htm?csnumber=32553

ITIL [64]is the abbreviation for the guideline **IT Infrastructure Library**. The main focus of the development was on mutual best practices for all British government data centres to ensure comparable services. Today, **ITIL** is the **worldwide de-facto-standard** for service management and contains broad and publicly available professional documentation on how to plan, deliver and support IT service features. In the meantime ITIL is already 20 years old and is now at its fourth release of the publications. The core publications are:

- Service Strategy
- Service Design
- Service Transition
- Service Operation
- Continual Service Improvement

These core publications describe 26 processes starting from the strategic orientation of the IT to the continual improvement of Services.

ITIL is a systematic approach to the delivery of quality IT services. It provides a basic vocabulary and a number of processes that are relevant in managing the lifecycle of IT services such as change management, release management, and service validation and testing.

B. Standards for description

For the description of metadata resources as interoperability assets, the **Asset Description Metadata Schema (ADMS)**[65] was developed by the ISA programme in 2011 and 2012. ADMS was subsequently published by the World Wide Web Consortium (W3C) as a W3C Note[66] in 2013.

The Asset Description Metadata Schema (ADMS) is a vocabulary to describe interoperability assets making it possible for ICT developers to explore and search for interoperability assets.

[64] http://www.itil-officialsite.com/
[65] European Commission. Joinup. The Asset Description Metadata Schema (ADMS). https://joinup.ec.europa.eu/asset/adms/home
[66] W3C. Asset Description Metadata Schema (ADMS). W3C Working Group Note 01 August 2013. http://www.w3.org/TR/vocab-adms/

ANNEX IV. EXISTING STANDARDS FOR REPRESENTATION OF DATA STANDARDS

This section lists a number of metadata standards that should be supported by metadata tools:

- Standard exchange formats for reference data;
- Standard exchange formats for data models;
- Standards for documenting metadata specifications.

A. Unified Modelling Language

The Unified Modelling Language (UML) is a standard by the Object Management Group (OMG) that can be used for data modelling. UML allows capturing the fundamental characteristics of the classes, properties and relations. Its primary purpose is to enable *humans* to understand the meaning of the data model. It is not used as a physical data model for information exchange per se.

UML has the following characteristics:

- **Graphical representation:** UML has become a de-facto standard for the graphical representation of a data model in the form of a class diagram.
- **XML Exchange format:** UML model scan be serialised and exchanged with other tools using the XML Metadata Interchange (XMI) – even though XMI conformance and interoperability is a known weak point of UML tools[67].
- **Local data elements:** In the UML language attributes and associations are local data elements that are encapsulated within the classes in which they are defined. This encapsulation prevents attributes and associations from being re-used independently from the classes in which they are defined. Unlike properties in RDF Schema, UML attributes and relationships are not global data elements.
- **UML profiles:** The use of UML profiles allows extending the UML language with a number of method-specific stereotypes, tagged values, and constraints. UML profiles are useful to adhere to a specific design patterns, and use model-driven development aids for the generation of XML and RDF Schemas.

B. XML Schema

An XML schema is a description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntax constraints imposed by XML itself. There are several different languages available for specifying an XML schema such as XSD, WXS and Schematron. Each language has its strengths and weaknesses.

Schema-validity assessment has three aspects:

1. Determining local schema-validity, that is whether an element or an attribute information item satisfies the constraints embodied in the relevant components of an XSD schema;

[67] See also the work of the OMG Model Interchange Working Group (MIWG) http://www.omgwiki.org/model-interchange/doku.php

2. Determining an overall validation outcome for the item by combining local schema-validity with the results of schema-validity assessments of its descendants, if any; and
3. Determining the appropriate augmentations to the infoset (and, if desired, exposing them to downstream applications in some way, to record this outcome).

As mentioned in chapter 3.2.2, data standards can be managed in an XML-based environment using XML Schema Definitions, or in a Linked Data environment using RDF-based formats.

C. RDF

In the Resource Description Framework (RDF), data is organised in graphs around subject-predicate-object statements (called triples). RDF has come to be used as a general method for conceptual description or modelling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.

RDF has, among others, the following characteristics:

- Flexible data integration
- Global data elements
- Uniform Resource Identifiers
- Multiple formats

D. Simple Knowledge Organisation System (SKOS)

SKOS[68], the Simple Knowledge Organisation System, is a common data model for sharing controlled vocabularies such as code lists, thesauri, and taxonomies via the Web in a machine-readable format. SKOS is a W3C Recommendation and commonly used in open-source and proprietary tools. In the Core Vocabularies[69] specifications of the ISA Programme, SKOS is the recommended vocabulary for the representation of code lists. The Publications Office already uses SKOS as the official format of EuroVoc, the EU's multilingual thesaurus, and the Named Authority Lists.

SKOS provides a standard way to represent knowledge organization systems using the Resource Description Framework[70] (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.

Using RDF also allows knowledge organization systems to be used in distributed, decentralised metadata applications. Decentralised metadata is becoming a typical scenario, where service providers want to add value to metadata harvested from multiple sources.

SKOS represents the terms in a controlled vocabulary as instances of the class skos:Concepts. SKOS also defines properties for multi-lingual labels (skos:prefLabel), associated codes (skos:notation), and definitions (skos:definition). The publication of controlled vocabularies represented in SKOS on the Web brings the following advantages:

1. **De-referencing**: the principles of Linked Data requires each term in the controlled vocabulary to be identified by a corresponding term URI based on the HTTP protocol. The term "Taxonomy" in the "Asset Type" scheme has for example the following term URI: <http://purl.org/adms/assettype/Taxonomy>. This means that when someone else encounters such an URI, he can look up its meaning by entering the URI in the address bar in his browser. This is called *de-referencing*. This is a simple yet powerful feature of the Web.
2. **Machine-readability**: In the example of "Taxonomy", the user can use the term URI to retrieve both a machine-readable and human-readable file containing definitions, labels, and related concepts for this term expressed in SKOS. Well-known thesauri such as EuroVoc have been defined using an ontology that extends SKOS.
3. **Multilingualism**: SKOS allows associating labels and definitions in multiple languages to any concept. This means that we can associate the labels "taxonomie"@FR, "Taxonomie"@DE, or "taxonomia"@PT to the concept identified with URI http://purl.org/net/mediatypes/application/OWL+XML to include the French, German, and Portuguese labels.

[68] http://www.w3.org/2004/02/skos/vocabs

[69] https://joinup.ec.europa.eu/system/files/project/Core_Vocabularies-Business_Location_Person-Specification-v1.00_0.pdf

[70] http://www.w3.org/RDF/

4. **Metadata alignment**: SKOS provides mapping properties like skos:closeMatch, skos:exactMatch, skos:broadMatch, skos:narrowMatch and skos:relatedMatch. These properties are used to state mapping alignment links between SKOS concepts in different concept schemes, where the links are inherent to the meaning of the linked concepts.

    a. The properties skos:broadMatch and skos:narrowMatch are used to state a hierarchical mapping link between two concepts.

    b. The property skos:relatedMatch is used to state an **associative mapping link between two concepts**.

    c. The property skos:closeMatch is used to link two concepts that are sufficiently similar that they can be used interchangeably in some information retrieval applications. In order to avoid possibilities of "compound errors" when combining mappings across more than two concept schemes, **skos:closeMatch is not declared to be a transitive property**.

    d. The property skos:exactMatch is used to link two concepts, indicating a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications. **skos:exactMatch is a transitive property**, and is a sub-property of skos:closeMatch.

SKOS is an extensible vocabulary. One popular extension is the *SKOS eXtension for Labels* (SKOS-XL)[71], which extends SKOS with additional support for describing and linking lexical entities and allows to track historical changes in the names of elements in a code list.

E. GeneriCode

The OASIS Code List Representation format, GeneriCode[72], is a single model and XML format (with a W3C XML Schema) that can encode a broad range of code list information. The XML format is designed to support interchange or distribution of machine-readable code list information between systems.

F. HTTP URIs

In order to facilitate its sharing and re-use across systems and organisations, data standards need to have persistent unique identifiers. As we are experiencing the era of the Web of Data, it is recommended that such identifiers come in the form of HTTP URIs. The ISA Programme as well as W3C have created good practices and guidelines for the design and management of well-formed, persistent URIs [11], e.g., see ISA's 10 Rules for Persistent URIs[73].

[71] SKOS-XL: http://www.w3.org/TR/skos-reference/skos-xl.html
72 http://docs.oasis-open.org/codelist/ns/genericode/1.0/
73 https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris/

G. Asset Description Metadata Schema (ADMS)

The Asset Description Metadata Schema (ADMS) is a common vocabulary for descriptive metadata, used to describe interoperability solutions [12]. ADMS is currently a W3C Working Group Note[74].

74 http://www.w3.org/TR/vocab-adms/