



European
Commission



WP4 Methodology

Data Semantics, Format and Quality

Document Metadata

| Property | Value |
|--------------|----------------------------|
| Release date | 2020-10-12 |
| Status | For review |
| Version | 0.04 |
| Authors | WP4 moderators and editors |

The latest version of this controlled document is stored on the [SDG Wiki](#).

| Disclaimer |
|--|
| <p>This report was prepared for the ISA² programme by PwC EU Services.</p> <p>The views expressed in this report are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.</p> <p>The European Commission does not guarantee the accuracy of the information included in this report, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission.</p> <p>All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.</p> |

Contents

| | |
|---|-----------|
| PHASE 1: SELECT EVIDENCE | 8 |
| PHASE 2: ANALYSE EXISTING EVIDENCE AND MODELS..... | 13 |
| PHASE 3: DRAFT DATA MODEL | 16 |
| PHASE 4: REVIEW DATA MODEL AND INCORPORATE COMMENTS | 22 |
| PHASE 5: SELECT CONTROLLED VOCABULARIES | 28 |
| PHASE 6: FINALISE DATA MODEL | 31 |
| PHASE 7: CREATE DISTRIBUTIONS AND PUBLISH DOCUMENTATION..... | 33 |

INTRODUCTION

This document describes the methodology prepared as part of WP4 for defining and agreeing on data models for the different types of evidence relevant for the SDG Regulation and the Once-Only Principle implementation. In order to provide a concrete and actionable methodology, the document has been kept as short and dense as possible, focussing on the essential activities and tools the respective stakeholders have to put in place when defining and agreeing on data models.

Important note for reviewing

Both content and form of the methodology have been developed for a web-based publication in the ASCII doc format on the [SDG Sandbox](#), to facilitate its usage and maintenance over time. In the context of the OOP first review cycle, the methodology has been transposed in a Word-format document. While reviewing, please focus on the following questions:

- Do the phases make sense (e.g. gaps, overlaps)?
- Do the steps make sense?
- Is the ownership of a specific step justified?
- Are the activities of a step complete and granular enough?
- Are the activities of a step well described?
- Can I provide additional rules, guidelines and tools?
- What example could best illustrate a specific step?

Following the review and the feedback from the reviewers, the methodology will be updated on the SDG sandbox.

Throughout the different steps, the methodology is illustrated with different examples. Please bare in mind that additional examples will be captured during a dedicated workshop on the 4th of November.

As described in the [review guidelines on the SDG wiki](#), this very document should be uploaded on the SDG wiki with comments, as a single version, on the same link, respecting the following rules:

1. Please add your comments to this document only and do not split them.
2. Please add comments only and avoid using track changes.
3. All track changes will be disregarded. Once you have completed the review, please append your alpha-2 country code to the document's title, for example: BE_(document title)
4. Please upload the document to the respective row in the table on the SDG wiki (see table 'to upload').

As indicated, reviewers have until the 6th of November to perform their review and share their comments if any.



STAKEHOLDERS' ROLES AND RESPONSIBILITIES

This chapter describes the stakeholders identified in the process of developing data models along with their role and responsibilities.

The shared goal of [developing a set of common data models \[...\] that best serves the interests of the SDG regulation and the Member States \(MS\) is broken down in different phases](#). These different phases are executed by distinct groups, which are described below.

Authority

Final decision owner regarding the results of development of the data models.

In the context of this Work Package, the European Commission is taking this role.

Working Group

The Working Group members contribute to the different deliverables and help others to meet the deadline of the SDG. Ideally, knowledge of the SDG is required and semantic awareness is recommended.

In the context of this Work Package, the Working Group is composed of representatives of Member States. Representatives attend the webinars and coordinate the work at the national level.

Domain experts

The domain experts can be divided per domain or [evidence type](#) (e.g. vital records, vehicles, etc.). They are the people who have the business experience from the domain. They know how the evidence is used, for which procedures, by whom and, most importantly, the information described within each type of evidence.

In the context of the SDG, one expert per domain should ideally be reachable for each country. Alternatively, a pool of 2-5 experts per domain would be enough to provide the expected input with the Working Group making sure that all the Member States have the possibility to monitor the quality of the work and the models proposed.

Editor(s)

The Editor(s) lead(s) the drafting of the deliverables and specifications (i.e. data model) by integrating and consolidating the input received from the Working Group. Specifically, the role of the editor(s) is threefold:

- To create a formal specification which is in line with the best practices in regards to data modeling and data standards reuse.
- To motivate and explain how every information request being discussed is either adopted in the formal specification, or not.
- To initiate the consensus making process around discussion topics.

Moderator

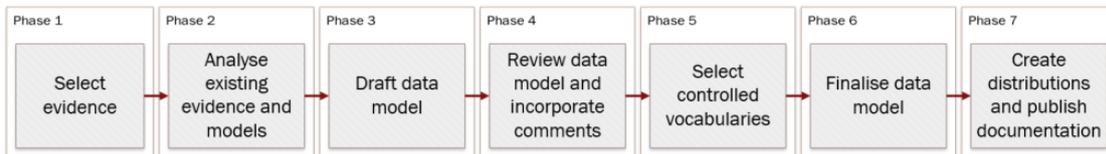
The Moderator works with the Rapporteur to ensure that the objectives, deliverables and deadlines of the Work Package are well defined and followed-up. Communicates with other Work Packages to ensure alignment.

Rapporteur

The Rapporteur collects input from the Working Group, ensures that the Working Group is aligned to the deadline of each deliverable in collaboration with the Moderator. Also, the Moderator communicates with other Work Packages to ensure alignment. The Rapporteur is drawn from the Working Group.

PROCESS

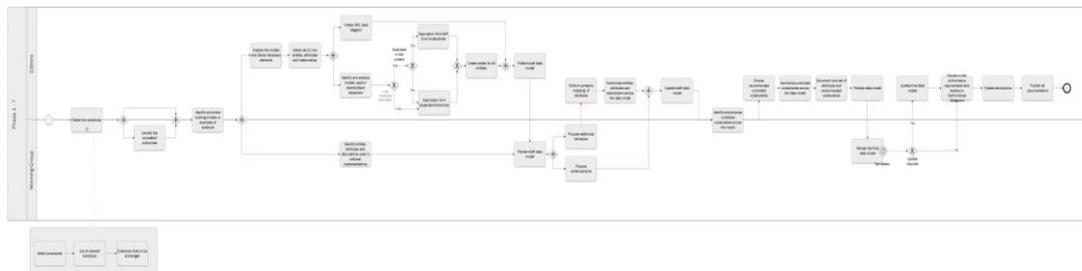
Key phases of the process:



Access each phase

- Phase 1: [Select evidence](#)
- Phase 2: [Analyse existing evidence and models](#)
- Phase 3: [Draft data model](#)
- Phase 4: [Review data model and incorporate comments](#)
- Phase 5: [Select controlled vocabularies](#)
- Phase 6: [Finalise data model](#)
- Phase 7: [Create distributions and publish documentation](#)

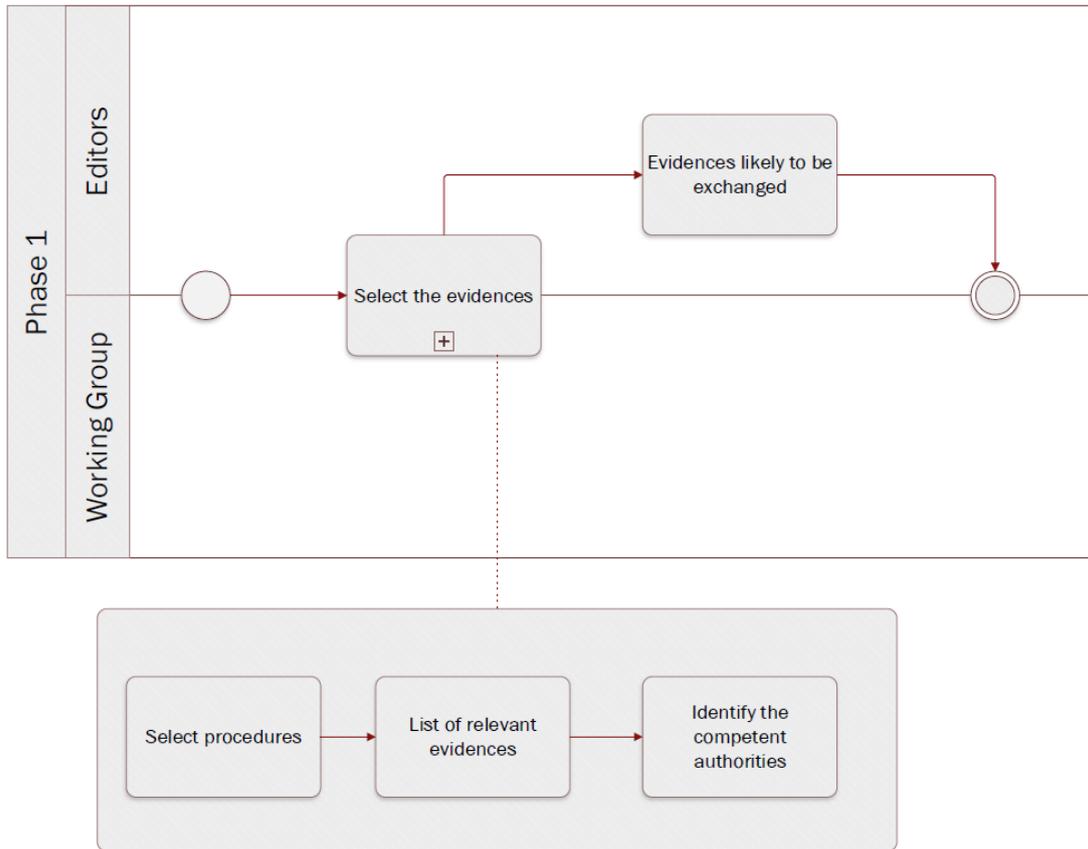
Detailed process



For each step, the key activities of every stakeholder groups involved are described. If you would like to know the more general roles and responsibilities of a

stakeholder group, please have a look at the first chapter defining roles and responsibilities.

PHASE 1: SELECT EVIDENCE



Process Phase 1



Step 1. Select the evidences

Key activities

- The Working Group members and the Editor(s) define together which evidences will be modelled.

Description

The process consists of three sub-steps:

- Step 1.1 Select procedures
- Step 1.2 List relevant evidences
- Step 1.3 Identify the competent authorities

Rules and Guidelines

According to Article 3 of the Regulation, evidence means “any document or data, including text or sound, visual or audio-visual recording, irrespective of the medium

used, required by a competent authority to prove facts or compliance with procedural requirements referred to in point (b) of Article 2(2) of the regulation”.

- In a world of paper-based documents, evidence takes the form of a written document issued by a competent authority including some key information in a more or less structured way.
- In a digital world made of databases and structured data exchange, the notion of evidence as a document is becoming obsolete. Digitally advanced Member States are exchanging and consuming pure data instead of documents to justify claims.

The notion of evidence as a document is gradually becoming obsolete. This evolution is fundamentally changing the way Member States store information and how they exchange evidences between public administrations.

Currently, both approaches coexist across Member States and are taken into account when developing data models.

In the evidence-based approach, civil servants are focusing on specific information (usually specific attributes) residing in an evidence, proving that a user fulfils specific criteria. In this sense, users are required to provide a set of evidences and the procedure is centred on provision of those evidences. In a criteria-based approach, the procedure moves away from evidence as such, and focuses on specific data about the user that proves the user fulfils a set of criteria to be met. In essence, the two approaches seek to answer the same question, but the question is framed differently.

Step 1.1 Select procedures

Key activities

- The Working Group members share their existing procedures.
- The Editor(s) analyse which procedures should be selected.

Description

The listing of available procedures will help to determine the scope of future exchange flows between competent authorities requesting and issuing evidences within the SDG OOP technical system. The Editors have to define whether a procedure exists in all countries for the latter to be considered in the next step. This should be done via a collaborative approach, via which the Member States are asked whether they perform the respective procedure.

The non-availability of a procedure in a country may be due to various reasons and automatically disqualifies the procedure for the next step. Additionally, the complexity of procedures can be assessed. Three categories are identified;

- A low level of complexity is assigned to procedures, which corresponds to one specific procedure conducted in a relatively similar way across Member States;

- A medium level of complexity is assigned to procedures, which corresponds to two separate procedures, or to one procedure which may involve separate options;
- A high level of complexity is assigned to procedures, which corresponds to more than two procedures, and which typically involves different competent authorities.

Rules and Guidelines

- The procedure is available in all Member States, and is relevant across borders;
- In each Member State, the management of the procedure is regulated by a competent authority;
- The procedure involves the use of at least one evidence.

Tool(s)

A collaborative tool, e.g. Confluence, GitHub.

Example(s)

The procedures in scope of the SDG OOP technical system are those listed under Annex II of the SDGR and Directives 2005/36/EC, 2006/123/EC, 2014/24/EU and 2014/25/EU.

In the Study on Data Mapping for the cross-border application of the Once-Only technical system SDG, the procedure *requesting proof of registration of birth* was identified as being available in all the member states surveyed. Additionally, it was associated with a low level of complexity.

Step 1.2 List relevant evidences

Key activities

The Editor(s) analyse which evidences should be selected, based on input from the Working Group.

Description

In order to understand what would be the actual types of evidence exchanged as part of the SDG OOP technical system. For each procedure, the list of evidences likely to be exchanged should be identified via input from the Member States. The relevance of cross-border evidence should be assessed as follows, i.e. whether the evidence required to complete the procedure is likely to be fetched from abroad.

- **Low:** there is no or little cross-border exchange of evidence (incl. case where no evidence is requested);
- **Medium:** there is a limited case for the exchange of evidence across-borders (only a few Member States request evidence that could be fetched across borders);
- **High:** there is a clear case for the exchange of evidence across-borders.

Tool(s)

A collaborative tool, e.g. Confluence, GitHub.

Example(s)

For instance, in the Study on Data Mapping for the cross border application of the Once-Only technical system SDG, the procedure *Applying for a tertiary education study financing, such as study grants and loans from a public body or institution* was identified as having a high level of cross-border evidence exchange.



Step 1.3 Identify the competent authorities

Key activities

- The Working Group members identify and map competent authorities with the selected evidences.
- The Editor(s) provide a means to collect the input from the Working Group members

Description

In order to further refine the scope of evidence that will be exchanged under the SDG OOP technical system, and by extension will be modelled, it remains to be seen which types of evidence are lawfully issued by a competent authority.

The competent authorities are most often public bodies at the national, regional and local levels – such as Ministries, National Social Security and Pensions systems, Regulatory Authorities, Local authorities and others. In some cases, the competent authority may be a private body or institution acting with a mandate from public authorities. For the latter, it is important to assess whether certain identified bodies can indeed be considered a competent authority (e.g. private universities, private health insurance companies).

An evidence that is not lawfully issued by a competent authority (e.g. evidence produced by the private sector/individuals such as invoice, leasing contract, sworn declaration) will be automatically considered out of scope.

Rules and Guidelines

- Note that the issuing authority is not necessarily also the data provider. For a birth certificate, e.g., the issuing authority might be the local municipality, while the data provider might be a state registry.

Tool(s)

A collaborative tool, e.g. Confluence, GitHub.

Example(s)

In the Study on Data Mapping for the cross border application of the Once-Only technical system SDG, it was identified that, for the procedure *requesting proof of registration of birth*, Member States either manage this procedure at the local or national level, most likely depending on the level of prior interconnection between the civil registries.

Step 2 Evidences likely to be exchanged

Key activities

- The Editor(s) propose which evidences should be modelled.
- The Working Group members validate which evidences should be modelled.

Description

In line with Article 14 of the SDG regulation, only the types of evidence that meet the following conditions will be exchanged under the SDG OOP technical system:

1. The procedure requires evidence beyond identification;
2. There are cases in which the evidence can be located abroad;
3. The evidence is lawfully issued by a competent authority (under SDGR); and
4. The evidence is available in an electronic format that allows for automated exchange.

First, when a procedure does not request any type of evidence other than identification, the procedure is not in scope of the technical system. Second, some types of evidence requested – other than identification – cannot originate from abroad due to their nature (e.g. proof of compliance of facilities with local sanitary or security legislation). Third, the evidence needs to be lawfully issued by a competent authority. Finally, following Article 14(2) of the Regulation, only evidence lawfully issued “in an electronic format that allows automated exchange” needs to be made available by Member States. Therefore, a type of evidence not issued in such a format by a given Member State will not be in scope of the technical system.

Tool(s)

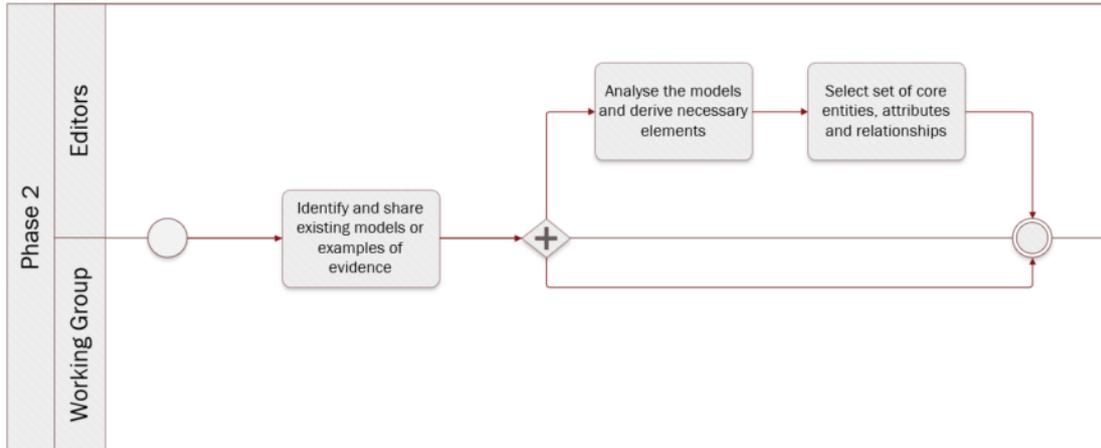
A collaborative tool, e.g. Confluence, GitHub.

Example(s)

For instance, taking into account all the factors mentioned in the preceding steps, the study on data mapping for the cross-border application of the Once-Only technical system SDG has identified the following categories of evidence as relevant for the modelling exercise.

| Category of evidence | Evidence | Relevant procedure(s) |
|---|---|---|
| Vital records and personal public documents | Birth certificate | Requesting academic recognition of diplomas, certificates or other proof of studies or course |
| Vital records and personal public documents | Certificate of change of name, such as marriage certificate | <ol style="list-style-type: none"> 1. Requesting academic recognition of diplomas, certificates or other proof of studies or courses 2. Recognition of professional qualifications of doctors (doctor of medicine) 3. Recognition of professional qualifications of nurses |

PHASE 2: ANALYSE EXISTING EVIDENCES AND MODELS



Process Phase 2

Step 3 Identify and share existing models or examples of evidence

Key activities

- The Working Group members identify and share existing models or examples of evidence.
- The Editor(s) collect information from the Working Group members.

Description

Working Group members will share information they possess related to the common data model being built.

The objective is to gather information in order to have a global overview of data models implemented and used across Europe and leverage this insight to develop a common data model. This task is assigned to the Working Group members who will report back to the editors using the channels and collaborative tools defined.

One important aspect of this step is **source of data quality**. This is ensured by the requirement that all data comes from authoritative sources. Working Group members are responsible to identify and connect the authorities to the system.

Rules and Guidelines

Before sending any data, the Working Group members should bear in mind the following;

- The data model has been validated and implemented by a competent authority; and
- The data model has been issued in a final version.

Tool(s)

The collaborative tool, e.g. GitHub.

Example(s)

Example of a data model shared by Spain: [issue #37](#).

Step 4 Analyse the models and derive necessary elements

Key activities

- The Editor(s) analyse the existing models and check what is common and can be reused.

Description

The Editors analyse the models, concrete examples and other useful documentation that they received from the Working Group in the previous step. They specifically look for similarities (and dissimilarities) between the different models in order to identify a common set of entities, attributes and relationships, that are relevant for the respective evidence that is being analysed.



Rules and Guidelines

- The [SKOS Mapping Properties](#) can be used to compare entities or attributes across different models.

Tool(s)

- A spreadsheet tool can be used to present and compare the different data models.

Example(s)

The table below illustrates how SKOS mapping properties can be used to compare models.

| SDG data model | Spain data model | SKOS mapping value |
|----------------|------------------|--------------------|
| Person | Person | exact match |
| Birth | | no match |



If provided, the table can also include definitions and URIs to ease comparison.

Step 5 Select set of core entities, attributes and relationships

Key activities

- The Editor(s) select the entities, attributes and relationships that are needed to model the respective evidence.

Description

The Editor(s) analyse the output from the previous step and derive the entities, attributes and relationships that are common to most data models and that are necessary to model the evidence. They do this by agreeing upon thresholds with the Working Group. These thresholds might be quantifiable, e.g. "if at least three

Member States request an attribute, the attribute is included” or “if one Member State is not able to provide an attribute, the attribute is made optional”.

Rules and Guidelines

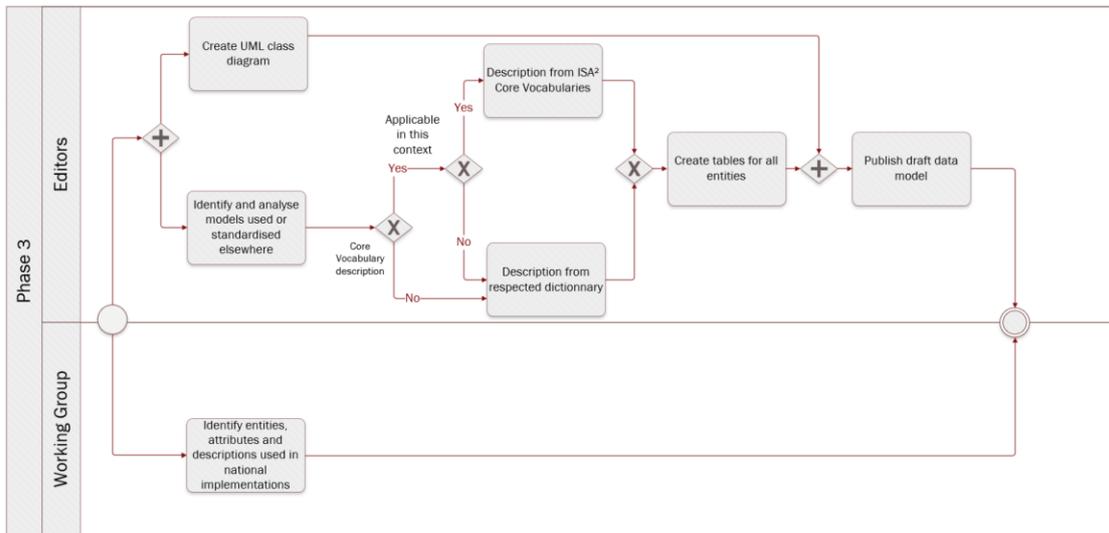
- Be as specific as possible, without restricting local flexibility too much.
- When selecting the core entities, attributes and relationships the editors can define thresholds allowing to decide which of the latter will be mandatory, optional or discarded.



Tool(s)

- A spreadsheet tool can be used to select the set of core entities, attribute and relationships of the common data model.
- The collaborative tool can be used to hold the discussion on the inclusion of entities, attributes and relationships.

PHASE 3: DRAFT DATA MODEL



Process Phase 3

Step 6 Identify entities, attributes and descriptions used in national implementations

Key activities

- The Working Group members identify and report features describing data models used in national implementations.
- The Editor(s) collect information from the Working Group members.

Description

Unlike step 3, step 6 is about the implementation (rather than the models). It might be that (semantic) data models do not exist or were not shared in step 3. Step 6 will remediate that by looking for elements in national implementations. Working Group members will share information on:

- Entities they judge paramount to the common data model being built
 - Attributes they judge mandatory and optional to the common data model being built;
 - Descriptions of elements in their national implementations with the ultimate purpose of harmonizing elements across member states.

Tool(s)

The collaborative tool, e.g. GitHub.

Example(s)

Example of an implementation (Person Condition Register and Registration Register) shared by Germany: see [issue #89](#).

Step 7 Identify and analyse models used or standardised elsewhere

Key activities

- The Editor(s) analyse European and global initiatives to standardize exchange of information.

Description

In parallel with step 6, the Editors document any European and/or global initiatives that aim at standardizing data exchanges across Member States. The output of this step will serve as a basis to draft the common data model.

The scope of step 6 and step 7 are complementary: while Working Group members are gathering information on what is nationally implemented, the editor team will focus on existing initiatives that are used to standardise the exchange of data. Step 6 will provide baseline information while step 7 will provide information to supplement the first draft of the common data model.

Rules and Guidelines

The data model to be produced is not modelling paper documents but rather evidence itself, i.e., information required by competent authorities to prove a fact. So, the grain of the data should be limited to the fact to prove, when modelling evidence types.

Tool(s)

- [Study on Data Mapping for the cross-border application of the Once-Only technical system SDG](#)
- [Linked Open Vocabularies](#)
- [Core Vocabularies](#)
- [Euro Vocabularies](#)
- [Public Documents forms | DG Justice](#)

Example(s)

The Core Person Vocabulary can be used when modelling data related to people.

Step 8 Create UML class diagram

Key activities

The Editor(s) design a UML class diagram.

Description

The editor team will leverage from the information collected in step 3, step 6 and step 7 to develop a UML class diagram. The latter aims at visually describing how entities of the data model will interact with each other. The different entities, the relationship between entities, and their attributes as well as the expected type are displayed.

Attributes should be presented in the following manner: *attributeName: expected type*. "Expected type" is further defined in step 11.

The exclusive focus on entities, attributes and relationships will allow the Working Group members to concentrate on the semantic aspects of the model. Supplementary modelling elements are added in step 11 when entities are documented in tables.

Rules and Guidelines

- Follow the [UML design rules](#):
 - Each element and their relationships should be identified in advance;
 - Attributes of each class should be clearly identified;
 - Avoid as much as possible lines crossing each other;
 - Ensure orthogonality of relationships;
 - Parents elements are higher than the child elements, so the subclass arrows always point upwards;
 - Align elements either by one of their sides or by their centers;
 - Make elements the same size if possible.

Tool(s)

Proprietary tools:

- [Enterprise Architect](#)
- [Microsoft Visio](#)



Open source tools:

- [Modelio](#)
- [UMLet](#)

Example(s)

[Birth certificate evidence](#)

Step 9 Description from ISA² Core Vocabularies

Key activities

- The Editor(s) assess whether the ISA² Core Vocabularies can be reused

Description

The editors verify whether an ISA² Core Vocabulary can be reused. Reusability is a key principle when drafting data models. In case there is no ISA² Core Vocabulary reusable, or it is not coherent to the context of the data model, the editors will consider other possibilities as presented in step 10.

Core Vocabularies are simplified, re-usable and extensible data models that capture the fundamental characteristics of an entity in a context-neutral fashion. Public administrations can use and extend the Core Vocabularies in the following contexts:

- *Development of new systems*
- *Information exchange between systems*
- *Data integration*

- *Open data publishing*

Tool(s)

- [Core Person Vocabulary](#)
- [Core Business Vocabulary](#)
- [Core Location Vocabulary](#)
- [Core Criterion and Core Evidence Vocabulary](#)
- [Core Public Organisation Vocabulary](#)
- [Core Public Service Vocabulary Application Profile](#)

Example(s)

The Core Person Vocabulary describes a class/entity Person that has an attribute/property "gender" that expects a Code as data type, coming from four possible controlled vocs: ISO, Eurostat, HL7 or SDMX.

Step 10 Description from respected dictionary

Key activities

- The Editor(s) fetch information elsewhere than the ISA² Core Vocabularies

Description

Should an entity or attribute not be (properly) defined in the ISA² Core Vocabularies, the editors find adequate documentation in other vocabularies.

In the event of information not being available in existing vocabularies, the editors propose definitions for new entities / attributes using respected and authoritative dictionaries.

Tool(s)

- [Euro Vocabularies](#)
- [Linked Open Vocabularies](#)
- [Oxford dictionary](#)
- [Merriam-Webster](#)

Example(s)

For instance, for the [Completion of secondary education evidence](#) the **course name** definition comes from [Merriam-Webster](#) ; i.e. "Name given to a number of lectures or other matters dealing with a subject."

Step 11 Create tables for all entities

Key activities

- The Editor(s) create tables for all entities.

Description

Relying on the input gathered in step 9 and step 10, the editors draft tables for all the entities of the data model (as presented in step 8). Per entity, the table consists of the following elements;

- Attribute(s) / relationship(s) 
- Expected type
- Definition
- Cardinality

Tables are a way to provide further information and context to the data model, unlike the UML class diagram which can be seen as a visual representation of the data model. Both form the data model referred to in the further steps.

Rules and Guidelines

Generic rules and guidelines for **step 11**

- The scope of the data model should be described by the fact or event that is proven by the evidence represented by the data model.

Specific rules and guidelines for the table per entity:

- Entities names should start with an uppercase.
- Attributes names should start with a lower case.
- Entities as well as attributes and relationships should be accompanied by a definition.
- The regulation 2016/1191 on  Public Documents sets a set of fields for the production of multilingual standard forms. Each field has a code and a text label that has been officially translated into the Member States' official languages. It is essential to provide (when possible) the correspondence between the attributes of the proposed data model to the fields of the multilingual standard forms of the regulation on Public Documents for evidences related to such domain.

Tool(s)

There are no specific tools for this step.

Example(s)

Here is are examples of tables created per entity, namely:

- [Birth Evidence](#)
- [Birth](#)
- [Person](#)
- [Public Organisation](#)
- [Location](#)

Step 12 Publish draft data model

Key activities

- The Editor(s) publish the output of step 8 and step 11, i.e. the draft data model

Description

The editors will publish the draft version of the data model, if considered stable enough, based on the output of step 8 (i.e. UML class diagram) and step 11 (entity tables). The draft data model will be published on the collaborative tool selected.

Rules and Guidelines

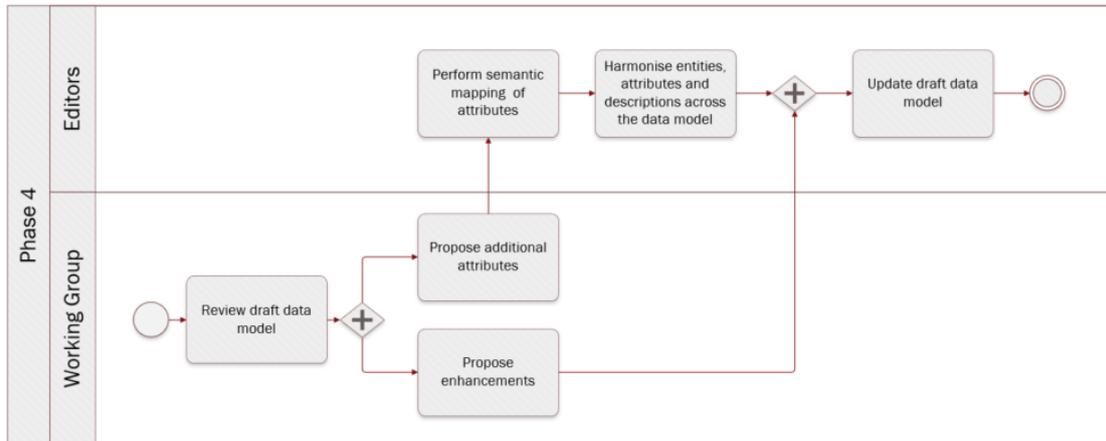
Publication as a Working Draft does not imply endorsement by the Working Group members or its representatives. This is a draft model and may be updated, replaced or made obsolete by other model at any time. It is inappropriate to cite this model as other than work in progress. Comments on the model are invited. Further details on Step 13

Tool(s)

A collaborative tool, e.g. GitHub.



PHASE 4: REVIEW DATA MODEL AND INCORPORATE COMMENTS



Process Phase 4

Step 13. Review draft data model

Key activities

- The Working Group directly review the proposed model and/or contact domain experts for reviewing it.
- The Editor(s) moderate and classify the issues (e.g. tag major issues which should be commented in priority by the reviewers).

Description

Each published draft of the Data Model is reviewed by the Working Group members and domain experts when relevant.

Beforehand, the Working Group members and the editors agree on a tool to collaborate and capture the feedback. The reviewers can then create issues using the designated tool. When comments are captured outside of the collaborative tool, the editors make an issue out of them in the issue tracker (for each comment).

The editors respond within an agreed timeframe to each issue, informing the reviewer that they have noticed and will process the issue. The editors consolidate solutions to the issue and seek for additional contribution from the reviewers (i.e. remainder of the Working Group members) when needed in collaboration with the moderator and rapporteur.

The issues can be in many different forms. For instance, a issue can deal with a modification to an existing entity or attribute, the addition or removal of an entity and/or attribute. For further details about these the types of issues, please check;

- Step 14. [Proposition enhancements](#)
- Step 15. [Propose additional attributes](#)

In addition to that, an issue can be qualified as `major` issue in the case it requires specific attention from the Working Group and the reviewers for commenting the issue and the potential resolutions. Further categorization (i.e. [labels](#)) is proposed when registering an issue.

The moderators make sure that the agreement process is transparent and acknowledged by all reviewers. Concerning *how consensus will be reached* you can find more information [here](#).

Rules and guidelines

- Reviewers are encouraged to directly create issues on the collaborative tool.
- Reviewers are encouraged to propose a solution in case they raise an issue.
- Reviewers are encouraged to use labelling and tagging for increasing searchability and responsiveness of contributors.
- Reviewers should consider how to present and discuss issues (e.g. technical versus business aspects).
- Reviewers are encouraged to provide context to their issues (e.g. data models used).
- Reviewers are encouraged to structure their issues and especially their denomination to increase comprehension. For instance:

Name of the data model or sub-part (e.g. relevant entity) and a short statement of the issue

+ **VehicleRegistrationCertificate evidence should contain registration status**

- Additional commenting guidelines are described in the [Wiki](#). These guidelines are specific for the SDG OOP but generic across the Work Packages (and therefore not limited to this methodology).

Tool(s)

We propose to use GitHub as the preferred tool for reviewing the data models. GitHub is a collaborative tool with built-in versioning control as well as other features that make it easy to propose suggestions and raise issues. ([Here](#) you can find the documentation on how to create issues on GitHub.) Nevertheless, other tools such as Jira or Confluence, can also be chosen. There are some important aspects to consider when choosing a tool, on which we elaborate further on [GitHub](#).

Example(s)

The following example describes the review of a draft data model followed by the creation of an issue and its processing by the editors and Working Group members. The process is the following:

1. The editors publish on GitHub the diagram and tables describing the [Vehicle registration certificate](#).
2. While reviewing the model, a semantic expert or a domain expert will try to answer the following questions:
 - Are the elements and their relationships correctly used and labelled?
 - Do you agree with the definition of the elements?
 - Are all elements necessary for this evidence described in the model?



- Are there conflicts between the elements of the model and the elements used in your country?
 - Is the element mandatory or optional in your country (cardinality)?
 - Do you have specific codes or expected types (e.g. format of date, address etc.) for attributes?
3. The reviewers document their issues on GitHub. For instance, concerning the [Vehicle registration certificate](#), the following issue was created [#45](#).

You may notice that the issue describes in practice several comments related to the vehicle registration certificate as well as an image of the data model used within the country.

To simplify the contribution of other reviewers to this issue, the editors will analyse the proposition, categorise it with labels, verify whether the issue should be restructured and describe the pros and cons of the issue documented.

In our example, each bullet point from the general comment should represent a separate issue. However, the editors should avoid as much as possible to complexify the structure of GitHub issues by creating complex hierarchies between the issues. For instance, the visual data model proposed by the issue owner does not need to be separated from the initial issue #45 since it represents a direct source of information which may be relevant for more than one issue.

4. The editors or the moderators answer, usually within one working day, to the initial issue created by acknowledging the issue or directly giving an initial answer.
5. The editors give more details about the pros and cons of (the various resolutions for) the issue(s) raised to trigger the discussions and comments from other Working Group members.
6. The discussion continues as reviewers comment on the issue.
7. When no agreement has been reached, the editors prepare the discussions and alternatives to be tackled during the next webinar following the public review period.

Step 14 . Propose enhancements

Key activities

- Working Group propose enhancements after [reviewing the data model](#).
- The Editor(s) consolidate the propositions and explain the pros and cons of the different propositions to the Working Group members. If needed, the editors seek for additional contribution from the reviewers in collaboration with the [moderator](#) and [rapporteur](#).

Description

Working Group members create semantic issues which deal with enhancements to the draft data models published. Enhancements can take the form of new features or requests regarding the proposed draft data models. It can be adjustments to the definitions, relationships, datatypes, cardinalities, etc.  As outlined in [Step 13. Review draft data model](#), the editors invite opinions and feedback to the issues and moderate the ensuing discussion.

After consideration of the propositions, the editors record the resolutions and send a response to the reviewers. To a semantic issue, the response usually includes a summary of the context of the proposition, the resolution agreed by the Working Group and the justification for the resolution, particularly in case the proposition is rejected.

Rules and Guidelines

- The Working Group must resolve each proposition in one of three ways:
 - Accepted: This usually means that changes will be made to the data model that will be reflected in the next release.
 - Rejected: No changes will be made to the draft data model.
 - Partially accepted: Part of the change is accepted, but other parts are rejected.

Tool(s)

There are no specific tools for this step. Similar to the previous step, we propose to use the GitHub issue feature (or pull request feature for the more advanced users) to propose enhancements.

Example(s)

TBD

Step 15 . Propose additional attributes

Key activities

- Working Group propose additional attributes after [reviewing the data model](#).
- The Editor(s) consolidate the propositions and explain the pros and cons of the different propositions to the Working Group members. If needed, the editors seek for additional contribution from the reviewers in collaboration with the [moderator](#) and [rapporteur](#).

Description

Working Group members create issues which deal with attributes (and entities) that could or should be included to the draft data models published. It might be that in certain cases Working Group members request the deletion of an attribute and/or entity.

As outlined in [Step 13. Review draft data model](#), the editors invite opinions and feedback to the issue and moderate the ensuing discussion.

After consideration of the proposition, the editors record the resolution and send a response to the reviewers. The response usually includes the resolution agreed by the Working Group and the justification for the resolution, particularly in case the proposed attribute(s) is (are) rejected.

Rules and Guidelines

The Working Group must resolve each proposition in one of three ways:

- Accepted: This usually means that changes will be made that will be reflected in the next draft data model.
- Rejected: No changes will be made to the draft data model.

- Partially accepted: Part of the change is accepted, but other parts are rejected.

Tool(s)

There are no specific tools for this step. Similar to the previous step, we propose to use the GitHub issue feature (or pull request feature for the more advanced users) to propose additional attributes/entities.

Example(s)

For instance, issue [#26](#) suggested to add the CO2 emission per KM as well as the environmental class attributes to the [vehicle class](#).

Step 16 . Perform semantic mapping of attributes

Key activities

- Upon receiving additional attributes from the Working Group members, the Editor(s) perform a semantic clustering of attributes. Afterward, editors will map the 'semantic clusters' to existing attributes, if any. Should there not be an attribute to map a 'semantic cluster' to, the editors will propose a new attribute (or entity).
- The Working Group members discuss the 'semantic clusters' - and potentially the new attribute(s) - and work towards consensus.

Description

Wherever attributes do not convey exactly the same information, 'semantic clusters' of similar attributes should be constructed to find a common, higher-level, and more general attribute to which the more specific attributes can be mapped.

Rules and Guidelines

- The relationships among different attributes (or entities) can be given a value according to the [SKOS \(Simple Knowledge Organization System Mapping system\)](#). The different values of which are
 - exact match
 - close match
 - related match
 - broader match
 - narrower match
 - (no match, i.e. absence of match)

Tool(s)

- This step can be performed using a spreadsheet tool, such as Microsoft Excel, in which related attributes are juxtapositioned in two columns and given a semantic mapping value in a third column.

Example(s)

speed hasCloseMatch velocity

Step 17 . Harmonise entities, attributes and descriptions across the data model

Key activities

- The Editor(s) harmonise the entities, attributes and descriptions across the data model.

Description

The editors consider all the entities, attributes and descriptions across the data model and check their consistency. Editors may propose changes to the attributes, for example to harmonise the names and definitions across entities or solve inconsistencies.

Rules and Guidelines

Example(s)

TBD

Step 18 . Update draft data model

Key activities

- The Editor(s) update the draft data model based on information collected in step 14, step 15, step 16 and step 17.

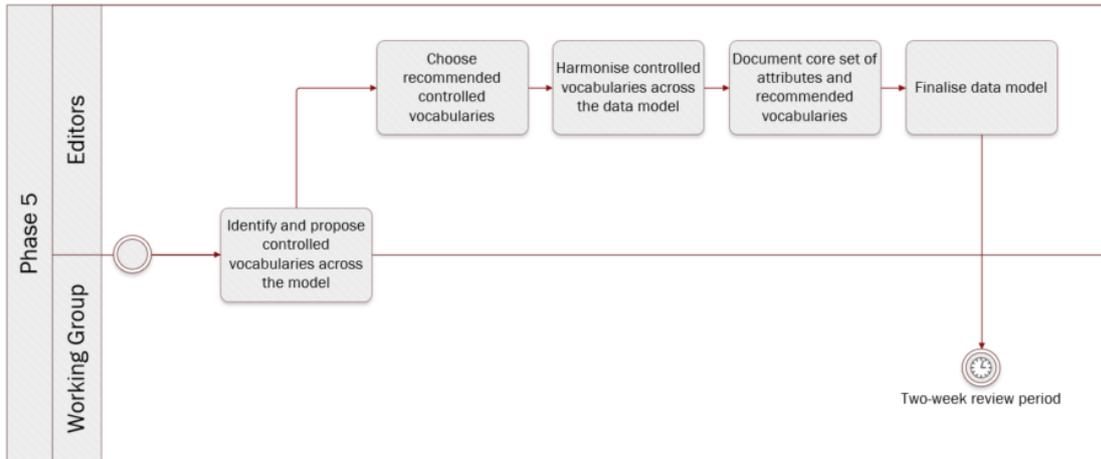
Description

The draft data model expressed as an UML diagram with textual description (i.e. tables) of the entities, attributes, relationships, definitions and cardinalities is updated. The editor constructs the new version of the data model based on the changes that have been agreed upon and derived from the previous four steps.

Rules and Guidelines

Publication as a Working Draft does not imply endorsement by the Working Group members or its representatives. This is a draft model and may be updated, replaced or made obsolete by other model at any time. It is inappropriate to cite this model as other than work in progress.

PHASE 5: SELECT CONTROLLED VOCABULARIES



Process Phase 5



Step 19 Identify and propose controlled vocabularies across the model

Key activities

- The Working Group and the domain experts propose controlled vocabularies for the different attributes defined in the previous phases.
- The Editor(s) synthesize the propositions and complement with additional standard controlled vocabularies where relevant.

Description

Once a core set of common attributes has been agreed upon and the [draft data model](#) is stable, the set of controlled vocabularies, for those attributes where a controlled vocabulary is needed, needs to be analysed. The editors create a table with the common attributes along one dimension and the local implementations along the other dimension, placing the controlled vocabularies suggested in the cells. Along with the controlled vocabularies, the Working Group is tasked to propose usage notes for all the attributes agreed upon.

Rules and Guidelines

Tool(s)

- [Core Person Vocabulary \(ISA\)](#)[Core Business Vocabulary \(ISA\)](#)[Core Location Vocabulary \(ISA\)](#)[Core Criterion and Core Evidence \(ISA\)](#)[Core Public Organisation \(ISA\)](#)
- [Core Public Service Vocabulary Application Profile](#)

-

Example(s)

For instance, for the [gender attribute](#) the [Human Sex](#) controlled vocabulary has been identified and proposed.

Step 20 Choose recommended controlled vocabularies

Key activities

- The Editor(s) put forward the different propositions for each attribute working towards a decision.
- The Working Group and domain experts discuss - through the collaborative tool - and select the controlled vocabularies to recommend.

Description

Based on the table of controlled vocabularies, the Working Group discusses which controlled vocabularies are the most appropriate to be recommended as well as the soundness of the proposed usage notes. This may be based on the status of particular vocabularies (e.g. if they are based on an international standard) or on their usage across multiple implementations.

In the case of divergent views, a live discussion may be organised by the Editor(s) and moderators to agree on the most controversial proposed solutions.

Rules and Guidelines

Step 21 Harmonise controlled vocabularies across the data model

Key activities

- The Editor(s) harmonise the controlled vocabularies and usage notes across the data model while ensuring the alignment between data models.

Description

The editors consider all controlled vocabularies and usage notes across the data model, check their consistency and identify any overlaps or gaps. Editors may propose changes to the recommendations, for example if different controlled vocabularies have been recommended for identical or similar attributes. Editors may also propose slight changes to the usage notes, for example to harmonise the writing style across the model or solve inconsistencies.

Step 22 Document core set of attributes and recommended vocabularies

Key activities

- The Editor(s) document the consensus and construct the working draft.

Description

On the basis of discussions in phase 4 and phase 5, the editors will document the decisions and prepare to update the draft data model.

Step 23 Finalise data model

Key activities

- The Editor(s) finalise the data model based on information collected in step 19, step 20, step 21 and step 22.

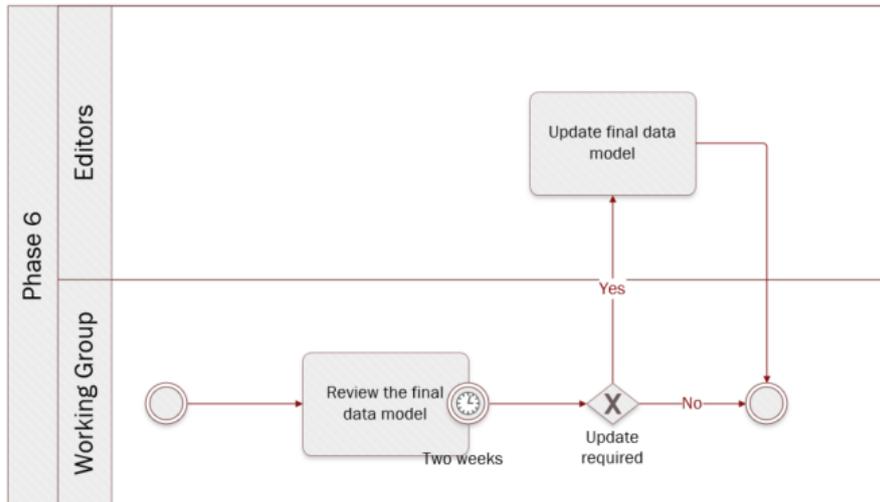
Description

The draft data model expressed as an UML diagram with textual description (i.e. tables) of the entities, attributes, relationships, definitions, cardinalities, controlled vocabularies and usage notes is finalised. The editors construct the final version of the data model based on the changes that have been agreed upon and derived from the previous four steps. Additionally, the model is prepared for review.

Rules and Guidelines

Publication as a *last call* Working Draft does not imply endorsement by the Working Group members or its representatives. This is a draft model and may be updated, replaced or made obsolete by other model at any time. It is inappropriate to cite this model as other than work in progress. Endorsement of the model will be sought in the next step 24

PHASE 6: FINALISE DATA MODEL



Process Phase 6

Step 24 Review the final data model

Key activities

- The Working Group members and the domain experts review the final data models
- The Editor(s) assist the Working Group members, collect and categorize the feedback.

Description

Working Group members discuss and validate the model with the business and functional experts and share their questions and / or remarks, if any, with the editors via the adequate channel.

In parallel, the editors collect and categorize the feedback. For instance,

- Editorial changes
 - Minor semantic changes
 - Major semantic changes - this entails a new version of the data model to be drafted and proposed for review.

Ultimately, the Working Group members have to come to a [semantic agreement](#) with regards to the data models reviewed.

To help reach that semantic agreement, the editors have the possibility to propose editable tables. The sole purpose of the tables is for Member States to indicate whether they are in capacity to provide the attributes listed in the data model. Ideally, the tables should be composed of the following columns;

- Entity
- Attribute
- Description
- Cardinality

- Country abbreviation (multiple columns allowing Member States to specify whether an attribute can be provided (Y) or not (N))

By no means the tables will replace the collaborative tool selected. The latter will still be home to the models and a place to discuss the latter. The tables are a way to collect input on whether an attribute can be provided or not in a structured manner. In case further information is necessary to provide an answer whether or not an attribute can be provided, Member States have to be redirected to the collaborative tool selected.

Rules and Guidelines

Aspects to bear in mind while reviewing:

- Data elements and entity names
- Model appearance
- Rules of normalization
- Definitions
- Model flexibility

Questions to bear in mind while reviewing:

- Do I agree with the proposed controlled vocabularies?
- Do I agree with the proposed changes to the data models?
- Are the entities and attributes definitions clear enough?
- Does the modelling approach make sense?
- Do I agree with the proposed cardinalities (i.e. mandatory versus optional)
- With data minimisation in mind, should some of the entities and or attributes be stripped off?
- Will my country be able to provide all the mandatory information?

Step 25 Update the final model

Key activities

- The Editor(s) process any last feedback and finish the final models.

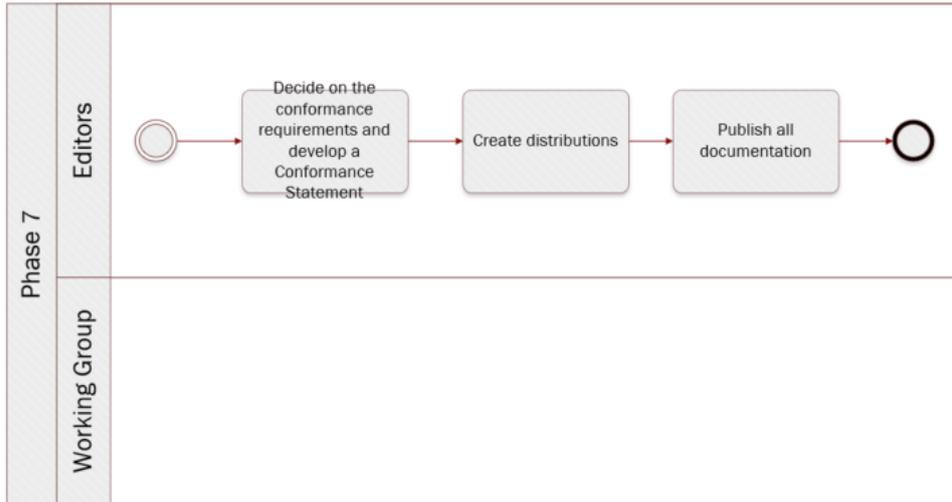
Description

In case the Working Group has given feedback in the previous step, the editors process these comments and make changes to the models as agreed by the Working Group. From this point, the editors can only make changes on which the Working Group has reached a consensus. Since there is no review period anymore, all changes that are carried out during this step should have been discussed with the Working Group.

Rules and Guidelines

- No changes are made during this step that were not agreed upon by the Working Group.
 - The change log is updated to reflect the final changes in order to achieve full transparency towards the Working Group.

PHASE 7: CREATE DISTRIBUTIONS AND PUBLISH DOCUMENTATION



Process Phase 7

Step 26. Decide on the conformance requirements and develop a conformance statement

Key activities

- The Editor(s) write a conformance statement.
- The Working Group members agree on the conformance statement.

Description

A conformance statement declares a minimum set of requirements that an implementation must adhere to, in order to be considered conformant with the respective data model. The Working Group members must agree on these conformance requirements. The editor then includes a conformance statement in the common data model.

It is possible that the model has natural divisions so that it might be appropriate to set different conformance levels. For example, a model used to describe vehicles may have a group of terms related specifically to motor vehicles that could be used in an implementation that has no needs to understand the terms that relate to bicycles. This will consequently lead to the establishment of different conformance levels.

Rules and Guidelines

- Publish the conformance statement together with the common data model.

Step 27. Create distributions

Key activities

- The Editor(s) create the required distributions for the data model.

Description

The data model can be expressed (or serialized) in various formats depending on the specific needs and context. Each distribution (format) will have its own uses and advantages, but also its own disadvantages and limitations.

Semantic data models can be expressed in different serialisation formats, such as TTL (turtle), RDF/XML, JSON-LD, SHACL, etc. Special care needs to be taken when using multiple formats, as conversions between different serialisation formats can potentially introduce inconsistencies.

Aside from these machine-readable formats, human-readable formats also need to be created. A visual representation of the entities, attributes and relationships of the data model is always recommended to give a clear overview. This can for example be a UML-diagram, saved as a PNG-file. Next to this, human-readable documentation is required with all the necessary information to construct the data models, i.e. the entities and attributes with their definitions, cardinalities, proposed codelists, etc. This can for example be distributed as an HTML-page or a PDF-document.

All these distributions can either be manually created, or automatically via one or multiple tools. If possible, preference should be given to the usage of an automated toolchain, reducing the risk of introducing inconsistencies during updates.

During this step, URIs are also created (or reused when possible) for the data model itself, its entities and their attributes. These identifiers need to be minted and maintained by a (European Commission) service.

Required codelists that do not exist yet, need to be created by e.g. the Publications Office, as part of the EU Vocabularies.

Rules and Guidelines

- Create both machine-readable as well as human-readable distributions of the data models.
- Automate the creation of the distributions as much as possible in order to avoid inconsistencies.
- Use URIs under data.europa.eu which allows for flexibility for where the URIs resolve to.



Tool(s)

- [VocBench3](#)
- Sparx Enterprise Architect

Example(s)

For instance, the Birth evidence was distributed in [XML](#).



Step 28. Publish all documentation

Key activities

- The Editor(s) publish all documentation on the collaborative tool.

Description

The Editor(s) publish the final version of the data model, in both machine-readable and human-readable format, on the collaborative tool selected. If possible, the

editors should publish the data model as open (meta)data and specify which license is applicable. 

Rules and Guidelines

- Choose an open license, e.g. CC0.
 - Publish the data model, its elements and related documentation via persistent (and ideally, dereferenceable) URIs.
 - Provide machine access to the data model.

Tool(s)

A collaborative tool, e.g. Confluence, GitHub.

ANNEX 1. GLOSSARY

Application profile

A data model defining which entities and attributes to use, what the cardinalities of the attributes are and recommendations for core vocabularies to be used, in order to support a particular application or use case(s).

Attribute

A characteristic of an entity in a particular dimension such as the weight of an object, the name of an organisation or the date and time that an observation was made, often representing things or events in the real world.

Data model

A data model is an abstract model that organises elements of data and standardizes how they relate to one another. It specifies the entities, their attributes and the relationships between entities.

Entity

A 'thing', such as a vessel, a geographic location, a sensor, a map or something more abstract like an incident, an event or an observation.

Relationship

A link between two concepts; examples are the link between an observation and the sensor that produced it, the link between a document and the organisation that published it, or the link between a map and the geographic region it depicts.

Semantic agreement

Consensus among the working group on the model and data entities that support common services. Apart from the typology of the data entities, the consensus also covers the characteristics of the data entities as expressed in metadata and the use of common controlled vocabularies.

