

Methodology - OOTS

i The following page is an extract from the [SDG sandbox](#).

When setting up work package 4, it was decided to erect a common space for efficiency reasons. This space holds the latest version of data models as well as a methodology to develop common data models developed in the context of the preparatory work for the Single Digital Gateway Regulation. This common space, commonly referred to as 'SDG-sandbox', is a GitHub repo.

[Introduction](#) [Methodology](#) [Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#) [Quality](#)
[Review cycles and consensus](#) [Stakeholders](#) [Terminology](#)

Methodology

This folder contains the various steps that together make up the methodology used to develop the data models of the evidences.

Disclaimer

This process has been designed based on the work of modelling ten evidences qualified as relevant in the Study on Data Mapping for the cross-border application of the Once-Only technical system SDG. The process is general enough to be used for future procedures, that is, in addition to the ones identified by the SDGR. More importantly, as this process has been designed based on a sample of ten evidences, it should not be considered complete nor exhaustive.

Essentially, this is a living document which is set to evolve as more evidences are being modelled and feedback fed in.

Naturally, as the focus is on modeling evidences, aspects dealing with maintenance are not covered in this document.

Process

In order to tackle the problem of identifying common semantics for different types of evidence exchanged in cross-border administrative procedures, a specific process has been envisaged. The process consists of key phases as shown below.

There are six phases, which range from the identification of existing efforts, evidences and data models to the creation of distributions and the publication of documentation and focus essentially on agreeing on semantics. True to the European Commission' spirit, this methodology provides tools and guidelines on how to reach the wider consensus possible.

This process forms part of a broader context. It is therefore not the start point nor the end point of the initiative to develop common data models. It is clearly important to be conversant with what triggers the creation of a common data model and in what context this data model is going to be used.

It is also worth outlining that such methodology only applies on the basis of preliminary work carried out upstream, i.e. the identification of the evidence to be modelled through the definition of use cases. Use cases determine for what purpose you need specific evidence in the context of a specific procedure and need to be defined and assessed to arrive at a relevant list of candidate evidences.

The evidence selection process will imperatively require to fulfill certain criteria in order to be potentially retained. These different selection criteria pave the way for discussions aiming at defining which evidence is to be modelled.

The involvement of domain experts (preferably from each member state) in this kind of discussion plays a key role in the collaboration between member states. Their knowledge of the different specific features of national use cases and evidence enables them to discuss and select the most relevant evidence to be modelled.

Only once the list of evidence to be modelled is decided, the methodology can be applied.

Access each phase

- [Phase 1: Identify and analyse existing standardization efforts, evidences and data models](#)
- [Phase 2: Draft data model](#)
- [Phase 3: Select controlled vocabularies](#)
- [Phase 4: Review data model and incorporate comments](#)
- [Phase 5: Finalise data model](#)
- [Phase 6: Create distributions and publish documentation](#)

For each step, the key activities of every stakeholder group is described. If you would like to know the more general roles and responsibilities of a stakeholder group, please have a look at the section defining roles and responsibilities. In a similar manner, key terms are defined in the glossary. Next to the key activities, when relevant, additional information is provided in the form of rules and guidelines, tools or even examples. This is intended to make this methodology as actionable as it can be, giving the reader the means to develop common data models. Finally, for each step, three types of activities have been identified.

1. Business analysis, i.e. identifying business needs and determining solutions.
2. Technical analysis, i.e. identifying technical requirements and determining solutions.
3. Review, i.e. formal assessment potentially leading to changes.

Business analysis activities are more present in the beginning of the methodology and as moving forward through the methodology they make room for technical activities. Just like any project, business needs are defined before technical needs. Review activities are scattered through the methodology.

Tool-wise, GitHub was 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.)

It should also be considered that engagement is a key element to the process. Having a high participation and a lot of input leads to better output and a greater consensus. Working Group members, i.e. Member States representatives, should be well represented during activities along the way.

Consensus is the driving force, therefore, ideally, as many Member States as possible should be involved in any of the activities.

[Introduction](#) [Methodology](#) [Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#) [Quality](#)

[Review cycles and consensus](#) [Stakeholders](#) [Terminology](#)

[Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#)

[Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#)

Phase 1 Phase 2 [Phase 3](#) Phase 4 Phase 5 Phase 6

Phase 1 Phase 2 Phase 3 [Phase 4](#) Phase 5 Phase 6

Phase 1 Phase 2 Phase 3 Phase 4 [Phase 5](#) Phase 6

Phase 1 Phase 2 Phase 3 Phase 4 Phase 5 [Phase 6](#)

Introduction Methodology [Phase 1](#) Phase 2 Phase 3 Phase 4 Phase 5 Phase 6 Quality
Review cycles and consensus Stakeholders Terminology



Phase 1: Identify and analyse existing standardization efforts, evidences and data models



Quick links:

- [Step 1 Identify and share existing models and standardisation efforts](#)
- [Step 2 Identify and share entities, attributes and descriptions used in national implementations](#)
- [Step 3 Identify and analyse models used or standardisation efforts \(elsewhere\)](#)

Navigate to the different phases

[:arrow_left: Previous phase](#) | [Next phase :arrow_right:](#)

Step 1 Identify and share existing models and standardisation efforts

Business analysis - *identification of business needs and related solutions.*

Key activities

- The **Working Group members** and **domain experts** identify and share existing models, standardisation efforts or policies.
- The **responsible DG** in line with the evidence being modelled share existing models, standardisation efforts or policies.
- The **Editors** collect information from the Working Group members and the responsible DG.

Description

Working Group members will share information they possess related to the common data model being built. Similarly, DGs having competencies in relation with the scope of the evidence being modelled, will share relevant information and existing (legal) pieces of work.

The objective is to gather information in order to have a global overview of data models, and/or standardisation rules implemented and used across Europe and leverage this insight to develop a common data model.

This step is specifically looking at information being available at global, i.e. European level, rather than at national level, which is the scope of step 2.

Rules and Guidelines

One important aspect of this step is the 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 information shared. Also, reusing content based on intrinsic licenses may oblige to use a specific license for the model being developed.

Tool(s)

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

Example(s)

For example, for social security, [EESSI \(Electronic Exchange of Social Security Information\)](#) is an IT system already in place. For education related matters, [Europass](#), from DG EMPL, is in place.

Step 2 Identify and share entities, attributes and descriptions used in national implementations

Technical analysis - identification of technical requirements and related solutions.

Key activities

- The **Working Group members** share existing national data models or examples of evidences.
- The **Working Group members** contact relevant **domain experts** in order to identify and report features describing data models used in national implementations.
- The **Editors** collect information from the Working Group members.

Description

Step 2 is about the national implementation of data models or legislative pieces. Contrary to step 1, step 2 is looking at gathering elements from national contexts.

It might be that (semantic) data models do not exist or were not shared in step 1. Step 2 will remediate that by looking for elements in national implementations.

Working Group members will share information on:

- Examples of evidences
- Entities they judge paramount for the common data model being built
- Attributes they judge mandatory and optional;
- Descriptions of elements in their national implementations.

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)

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.

Italy 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.

Example of an implementation (Person Condition Register and Registration Register) shared by Germany: see [issue #89](#). Example of a data model shared by Spain: [issue #37](#).

Step 3 Identify and analyse models used or standardisation efforts (elsewhere)

Business analysis - identification of business needs and related solutions.

Key activities

| The **Editors** analyse European and global initiatives to standardise exchange of information.

Description

In parallel with step 1 and 2, the Editors document - the information received and - 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.

Step 1 and 2 are the source of information for step 3. While Working Group members and responsible DGs are gathering information, the editors will focus on documenting and analysing the information received. Editors should also do a research effort to not exclude any relevant data model and standardization effort used elsewhere.

This step supplements the statement made in step 2, as for existing harmonization of information contained in the evidences at European level. The editors may derive the necessary elements from these initiatives.

Rules and Guidelines

Reusing content based on intrinsic licenses may oblige the to use a specific license for the model being developed.

Tool(s)

Below are some links of input sources.

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

Example(s)

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

[Introduction](#) [Methodology](#) [Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#) [Quality](#)

[Review cycles and consensus](#) [Stakeholders](#) [Terminology](#)



Phase 2: Draft data model



Quick links:

- [Step 4](#) Analyse the models and derive necessary elements
- [Step 5](#) Select set of core entities, attributes and relationships
- [Step 6](#) Description from ISA² Core Vocabularies
- [Step 7](#) Description from other authoritative sources
- [Step 8](#) Create UML class diagram
- [Step 9](#) Create tables for all entities

Navigate to the different phases

[:arrow_left: Previous phase](#) | [Next phase :arrow_right:](#)

Step 4 Analyse the models and derive necessary elements

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** analyse the existing data models and information shared to check what is common and can be reused.

Description

The Editors analyse the data models, concrete examples and other useful documentation received from the Working Group and the DGs in the previous steps. They specifically look for similarities (and dissimilarities) between the different data models and documentation in order to identify a common set of entities, attributes and relationships, that are relevant for the respective evidence that is being modelled. Considering the procedure for which the evidence is modelled, and subsequently the use case(s) can serve to analyse models and documentation in order to derive necessary elements.

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. The Editors should look for the minimum common denominator when consolidating and analysing (fragment of) data models and information received.
- The **SKOS Mapping Properties** can be used to compare entities or attributes across different models.
- 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. For instance, if no other Member State mentioned the need for an attribute it will therefore be discarded.

Tool(s)

- **Linked Open Vocabularies** which is a source for predicates, i.e. existing attributes/relationships that might be candidates for reuse.
- 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. **insert picture** If provided, the table can also include definitions and URLs to ease comparison.

Step 5 Select set of core entities, attributes and relationships

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** select the entities, attributes and relationships that are needed to model the respective evidence.

- The **Editors** propose which attributes and relationships are mandatory / optional.

Description

With the output of the previous steps, the Editors select the entities, attributes and relationships that are common to most data models and that are necessary to model the evidence. They also determine which attributes should be mandatory and optional.

They do this by agreeing upon thresholds with the Working Group. These thresholds might be quantifiable, e.g. “if at least five Member States have 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.

Tool(s)

- A spreadsheet tool can be used to select the set of core entities, attributes 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.

Step 6 Description from ISA² Core Vocabularies

Technical analysis - identification of technical requirements and related solutions.

Key activities

The **Editors** assess whether the ISA² Core Vocabularies can be reused

Description

The Editors verify whether an ISA² Core Vocabulary can be reused. Reuse is a key objective 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 7.

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.
- Gender is a challenging topic due to the different recognition of non-binary gender, [issue #143](#).

Step 7 Description from other respected sources

Technical analysis - identification of technical requirements and related solutions.

Key activities

The **Editors** 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 elsewhere. ‘Not properly defined’ refers to a circular definition of a term, i.e. already containing the term that is to be defined.

1. Other respected sources can be considered when the terms are defined in a well-known domain-specific ontology.
In general, entities, attributes, relationships and definitions should be linked to existing terminologies.
2. In the event of information not being available in existing vocabularies, the editors propose definitions for new entities / attributes using respected and authoritative dictionaries (which are deemed of excellence).

A ‘respected dictionary’ refers to a dictionary widely regarded as the accepted authority on the English language.

Rules and Guidelines

Generic rules and guidelines

- Entities can be documented by using tools such as the [Interoperability Platform and Data Vocabularies Tools](#).

Specific rules and guidelines for the table per entity

- When defining a term, it should not be included in the tentative definition.

Tool(s)

- [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 8 Create UML class diagram

Technical analysis - identification of technical requirements and related solutions.

Key activities

- The **Editors** design an UML class diagram

Description

The Editors will leverage from the information collected in the previous phase 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 types are displayed.

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 9 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;
- Attributes should be presented in the following manner:attributeName: expected type. “Expected type” is further defined in step 11;
- 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 of the same size, if possible;
- Diagrams should show the cardinality of attributes and relationships as well;
- Entities names should start with an uppercase;
- Attributes names should start with a lower case.

Tool(s)

Some examples of proprietary and open source tools are the following:

Proprietary tools:

- [Enterprise Architect](#)
- [Microsoft Visio](#)
- [MagicDraw \(No Magic\)](#)
- [Visual Paradigm](#)

Open source tools:

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

Example(s)

- [Birth Certificate evidence](#)

Step 9 Create tables for all entities

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** create tables for all entities.

Description

Relying on the input gathered, the editors draft tables for all the entities of the data model. Per entity, the table consists of the following elements;

- Proposed attribute(s) / relationship(s)
- Proposed expected type
- Proposed definition
- Proposed 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 9

- Multilingualism, localisation and internationalisation aspects should be considered. A language neutral identifier for every concept and additional Member State language columns in the tables helps Member State participation.
- The scope of the data model should be described by a fact or an event that is proven by the evidence represented by the data model.
- The tables should have a language-neutral identifier that, along the creation and review of the data model, is agnostic to name changes.

Specific rules and guidelines for the table per entity:

- Sources of the entities/attributes should be added, e.g. existing regulation, reused model, etc.
- Entities, attributes and relationships should be accompanied by a definition as well as their cardinality.
- [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. The aforementioned approach could be reused for evidences other than public documents.

Tool(s) *The collaborative tool, e.g. Github.* **Example(s)**

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

Introduction Methodology Phase 1 Phase 2 [Phase 3](#) Phase 4 Phase 5 Phase 6 Quality
 Review cycles and consensus Stakeholders Terminology



Phase 3: Select controlled vocabularies



Quick links:

- [Step 10 Identify and propose controlled vocabularies across the model](#)
- [Step 11 Choose recommended controlled vocabularies](#)
- [Step 12 Create controlled vocabularies](#)
- [Step 13 Harmonise controlled vocabularies across the data model](#)
- [Step 14 Document core set of attributes and recommended vocabularies](#)

Navigate to the different phases

[:arrow_left: Previous phase](#) | [Next phase :arrow_right:](#)

Step 10 Identify and propose controlled vocabularies across the model

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Working Group members** and the **domain experts** propose controlled vocabularies for the different attributes defined in the previous phases.
- The **Editors** synthesise 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 enough, 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

- Controlled vocabularies at the EU level are multilingual which helps in cross- border data exchange scenarios.
- (domain-specific) Controlled vocabularies which are internationally accepted should be considered.
- Controlled vocabularies should have governance processes in place, be hosted in a sustainable manner and be provided free of charge.

Tool(s)

- [EU Vocabularies](#)
- [Core Public Service Application Profile](#)

Example(s)

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

Step 11 Choose recommended controlled vocabularies

Technical analysis - identification of technical requirements and related solutions.

Key activities

- The **Editors** put forward the different propositions for each attribute working towards a decision.
- The **Working Group members** and the **domain experts** discuss - through the collaborative tool - and select the controlled vocabularies.

Description

Based on the table of controlled vocabularies, the Working Group members discuss 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 Editors and the moderator to agree on the most controversial proposed solutions.

Rules and Guidelines

It is important to agree on common official controlled vocabularies that can harmonise across different countries the way in which specific values of properties can be specified, allowing for a uniform indexing and retrieving of data based on common terms.

Example(s)

As suggested by the Working Group, the editors have used the [language code list](#) as controlled vocabulary for the language attribute of all tertiary education related evidences ([see issue #120](#)).

Step 12 Create controlled vocabularies

Technical analysis - identification of technical requirements and related solutions.

Key activities

- The **Editors** create a proposition of new controlled vocabularies.
- The **Working Group members** review the proposition and provide comments.
- The **Publication Office** The Publications Office creates controlled vocabularies.

Description

In the event of no controlled vocabularies being available, the Editors (or Working Group members) have the opportunity to propose the creation of controlled vocabularies. Required controlled vocabularies, that do not exist yet, need to be created by the Publications Office, as part of the EU Vocabularies. Of course, existing controlled vocabularies can be updated, if necessary.

Tool(s)

- [The Publication Office](#)

Step 13 Harmonise controlled vocabularies across the data model

Technical analysis - identification of technical requirements and related solutions.

Key activities

The **Editors** 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 - and across all SDG data models -, 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 14 Document core set of attributes and recommended vocabularies

Technical analysis - identification of technical requirements and related solutions.

Key activities

The **Editors** document the consensus and construct the working draft.

Description

On the basis of discussions in phase 3 and phase 4, the editors will document the decisions and prepare to update the draft data model.

Introduction Methodology Phase 1 Phase 2 Phase 3 [Phase 4](#) Phase 5 Phase 6 Quality
Review cycles and consensus Stakeholders Terminology



Phase 4: Review data model and incorporate comments



Quick links:

- [Step 15 Publish draft data model](#)
- [Step 16 Review draft data model](#)
- [Step 17 Propose enhancements](#)
- [Step 18 Propose additional attributes](#)
- [Step 19 Perform semantic mapping of attributes](#)
- [Step 20 Harmonise entities, attributes and descriptions across the data model](#)
- [Step 21 Update draft data model](#)

Navigate to the different phases

[:arrow_left: Previous phase](#) | [Next phase :arrow_right:](#)

Step 15 Publish draft data model

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** finalise the data model based on information collected in step 8, step 9, step 10, step 11, step 12, step 13 and step 14.
- The **Editors** publish the output.

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 draft version of the data model based on the changes that have been agreed upon and derived from the previous seven steps. Additionally, the model is prepared for review.

Finally, it is important for Working Group members and the Editors to agree on an Open Licence to be used. Reusing content based on intrinsic licences may oblige editors to use a specific licence. Also, acknowledgement sections should be added specifying that data models developed heavily rely on the contributions of Working Group members, subsequently Member States.

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 another 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 17.
- Choose an open license, e.g. CC0, [EUPL](#).
- Publish the data model, its elements and related documentation via persistent (and ideally, dereferenceable) URIs.
- Provide machine access to the data model.

Tool(s)

The collaborative tool, e.g. GitHub.

Example(s)

Based on the steps described before, diagrams and [tables](#), in their first version, were published.

Step 16 Review draft data model

Review - formal assessment potentially leading to changes.

Key activities

- The **Working Group members** directly review the proposed model and/or contact the **domain experts** for reviewing it
- The **Editors** moderate and classify the issues.

Description

The Working Group members and the Editors agree on a tool to collaborate and capture the feedback. Using this tool, reviewers can create issues and the Editors follow up on them thanks to an issue tracker.

Then, the Editors publish the draft using the collaborative tool. The published draft of the data model is reviewed by the Working Group members and domain experts when relevant.

The Editors respond within an agreed timeframe to each issue made by the Working Group members, informing the reviewers that they have noticed and will process the issue. The Editors consolidate solutions to the issue and seek for additional contribution from the reviewers. This is done in collaboration with the moderator and rapporteur.

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

- [Step 17 Propose enhancements](#)
- [Step 18 Propose additional attributes](#)

Issues are categorised according to their type; (i) editorial (ii) minor or (iii) major.

- **Editorial issue:** issue stemming from errors in the data model, which are not affecting the semantic agreement in any way. These issues might be addressed directly and do not lead to another review cycle.
- **Minor issue:** issue leading to direct changes in the deliverables. These issues might be addressed directly and do not lead to another review cycle.
- **Major issue:** issue qualified as show stopper and/or transversal issue. Either stakeholders decide the issue to be addressed directly, without leading to another review cycle, or once the issue addressed, the data model needs to undergo another review round.

The moderator makes sure that the agreement process is transparent and acknowledged by all reviewers.

Rules and Guidelines

- Use case descriptions should be provided along with the data model.
- Model components should be translated.
- Editors organise issues as in a forum, by discussions, subjects and hierarchise the threads.
- 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 model 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)

The collaborative tool, e.g. Confluence, 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 the 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, the **domain experts** will try to answer the following questions:
 - Can you process the evidence in your country if only the mandatory attributes are provided? If not, what other optional or missing attributes do you need?
 - 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 week, to the initial issue created by acknowledging the issue or directly giving an initial answer.
5. The **Editors** propose resolutions or ask more details concerning the issue(s) raised to trigger discussion 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 a webinar to be organised following the review period.

Step 17 Propose enhancements

Review - formal assessment potentially leading to changes.

Key activities

- The **Working Group members** propose enhancements after reviewing the data model, if needed.
- The **Editors** consolidate the propositions and present them with resolutions to the Working Group members. If needed, the Editors seek for additional contributions 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 model published. Enhancements can take the form of requests regarding the proposed draft data model. It can be changes to the definitions, relationships, data types, cardinalities, etc.

In this context, it must be understood that enhancement also means restrictions, as one of the key principles of developing data models is data minimisation.

As outlined in Step 16. Review draft data model, the Editors invite opinions and feedback to the issues and moderate the ensuing discussion.

After consideration of the proposition, the Editors assess the type of issue, whether it is minor or major, and record the resolution. After that, a response is sent 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 members and the justification for the resolution, particularly in case the proposition is rejected.

Rules and Guidelines

The Working Group members 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. As indicated in the previous step, resolution will either lead to phase 5 or phase 4.

Tool(s)

There are no specific tools for this step. The GitHub issue feature can be used (or pull request feature for the more advanced users) to propose enhancements.

Example(s)

As described in [issue#125](#), a proposition was made to enhance an attribute as it was too narrowed down and did not encompass all the possibility for that attribute.

Step 18 Propose additional attributes

Review - formal assessment potentially leading to changes.

Key activities

- The **Working Group members** propose additional attributes after reviewing the data model, if need be.
- The **Editors** consolidate the propositions and present them with resolutions 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 attributes (and entities) that could or should be included in the draft data model published. It might be that in certain cases Working Group members request the deletion of an attribute, a controlled vocabulary, and/or entity.

As outlined in Step 16. 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 assess the type of issue, whether it is minor or major, and record the resolution. After that, a response is sent to the reviewers. The response usually includes the resolution agreed by the Working Group members and the justification for the resolution, particularly in case the proposed attribute(s) is (are) rejected

Rules and Guidelines

The Working Group members 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.

By default, attributes and entities added to the data model are optional.

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 adding the CO2 emission per KM as well as the environmental class attributes to the vehicle class. In [issue#73](#) additional dates were added to the model.

Step 19 Perform semantic mapping of attributes

Technical analysis - identification of technical requirements and related solutions.

Key activities

- Upon receiving additional attributes from the the **Working Group members**, the **Editors** perform a semantic clustering of attributes. Afterward, the 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

- For instance, [#issue 143](#) reported that in the [sex/gender code list from the Publication Office](#), the property “not applicable” related to the legal recognition of non-binary gender.

Step 20 Harmonise entities, attributes and descriptions across the data model

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- the **Editors** harmonise the entities, attributes and descriptions across the data model.

Description

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

Rules and Guidelines

In order to guarantee semantic interoperability amongst different common data models – that might be developed at the same time –, the same modelling patterns, especially for concepts independent from a specific domain, can be applied across data models (e.g. location, person, organisation) unless specific characteristics for them are required.

Example(s)

Following a discussion on the SDG sandbox, the editors proposed to align the Location entity for all tertiary education related evidences (see [issue #133](#)).

Step 21 Update draft data model

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

*the **Editors** create an updated coherent draft common data model based on information collected in the previous steps.*

Description

The draft data model expressed as an UML diagram with textual description (i.e. tables) of the entities, attributes, relationships, definitions, cardinalities and controlled vocabularies, i.e. codelists, is finalised. The Editors construct the new and final 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 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 another model at any time. Endorsement of the model will be sought in the `step 23`.

Introduction Methodology Phase 1 Phase 2 Phase 3 Phase 4 [Phase 5](#) Phase 6 Quality

Review cycles and consensus Stakeholders Terminology



Phase 5: Finalise data model



Quick links:

- [Step 22 Test the final data model with instance data](#)
- [Step 23 Review the final data model](#)
- [Step 24 Update the final data model](#)

Navigate to the different phases

[:arrow_left: Previous phase](#) | [Next phase :arrow_right:](#)

Step 22 Test the final data model with instance data

Review - *formal assessment potentially leading to changes.*

Key activities

- A selected number of **Working Group members** and domain experts test the model against instance data.
- The **Editors** assist the Working Group members in the testing by collecting and categorising the feedback.

Description

So far, the process of defining the elements of the data model was a theoretical exercise. The objective of this step is to test the final model against instance data, i.e. actual data, in order to discover potential flaws or blind spots in the model. In this step, working group members have to provide (dummy) instance data and report on the challenges they face when:

- mapping this instance data to the model (perspective of the data provider). Working group members must answer the question: “*Can we provide this information?*”.
- processing instance data that respects the data model (perspective of the data consumer). Working group members must now answer the question: “*Can we process this information?*”, where the information represents the minimum data required by the model and, in this case, considering that the data was hypothetically received from another party.

Mapping instance data is, in the jargon, looking from the data provider perspective. For instance, a person needs evidence of a diploma from studying in a Member State (A) for a procedure in another Member State (B). The mapping takes the perspective of Member State (A). From the other perspective, processing the instance data would take the role of the data consumer. In the example above, Member State (B) is the data consumer.

A likely process for this step could be as follow:

1. **Initiate** – All working group members have the possibility to volunteer for the testing of the data model with instance data. In the beginning of this exercise, editors will organise a meeting with the volunteers to walk them through the process and outline the expectations.
2. **Map** – Volunteers will put on the hat of the data provider and create instance data for the data model, with as many attributes as available in their national system, and map them to the attributes in the template provided.
3. **Process** – Volunteers will put on the hat of the data consumer and receive minimal evidence (mandatory fields only) data from another MS, i.e. another volunteer - as collected in the preceding step. Volunteers will then process the instance data received.
4. **Report** – Volunteers will report on (semantic) challenges arising from both the mapping and processing of instance data. This step should reveal potential flaws in the model thanks to a life-like situation of processing an evidence.

5. **Improve** – After the testing comes the reporting. Volunteers will therefore share their findings with the broad audience and discuss how to improve the models (e.g. by adding usage notes).

The feedback received during this step needs to be documented, categorized and analysed.

Rules and Guidelines

Questions to bear in mind when testing the model against instance data:

- How relevant do you think the data in the attribute is for cross-border exchange?
- For the mandatory attributes: how can you process them, and are there any specific requirements for the format of the data?
- For the optional attributes: what are the challenges for processing of the data if the attribute is missing?

Tools

For this exercise, a spreadsheet can come in handy.

Attribute	Expected type	Definition	Cardinality	Code list	Instance data	Mapping relation	Mapping Comment	Processing comment
Identifier	Identifier	An unambiguous reference to the Tertiary Education Evidence.	[1..1]	N/A				
issuing date	Date	The date on which the Tertiary Education Evidence was issued.	[1..1]	N/A				
language	Code	The language in which the Tertiary Education Evidence is issued.	[1..*]	Language				
qualification name	Text	Full name of the qualification, at least in the original language(s) as it is styled in the original qualification, e.g. Master of Science, Kandidat nauk, Maîtrise, Diplom, etc.	[1..*]	N/A				
issuing place	Location	The Location where the Tertiary Education Evidence was issued.	[1..1]	N/A				
belongs to	Student	The Student that is the holder of the Tertiary Education Evidence.	[1..1]	N/A				
obtained at	Education Institution	The Education Institution that educated the Student.	[0..*]	N/A				
issuing authority	Organisation	The Organisation that issued the Tertiary Education Evidence.	[1..*]	N/A				

Several columns to describe the model will be needed:

- Attribute
- Expected type
- Definition
- Cardinality
- Code list

Along with these elements, some input fields need to be provided:

- Instance data - Actual data to be provided. For instance, the given name for Johann Sebastian Bach is “Johann Sebastian”
- Mapping relation - e.g. exact match, no match, near match, etc. [For further information on the definitions of these mappings](#)
- Mapping comment - Comments in case there is a remark, suggestion, issue with the mapping, i.e. data provider perspective
- Processing comment - Comments in case there is a remark, suggestion, issue with the processing, i.e. data consumer perspective

Step 23 Review the final data model

Review - formal assessment potentially leading to changes.

Key activities

- The **Working Group members** and the **domain experts** review the final data model.
- The **Editors** assist the Working Group members, collect and categorise the feedback.

Description

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

In parallel, the Editors collect and, again, categorise the feedback. For instance:

- Editorial issue
- Minor issue
- Major issue

This step is also important to set the final agreement on cardinalities. To help with that, the Editors have the possibility to propose editable tables. The sole purpose of the tables is for the Working Group members to indicate whether they are in capacity to provide the attributes listed in the data model. But also whether a specific attribute is needed to process the evidence.

Ideally, the tables should be composed of the following columns:

- Entity
- Attribute
- Description
- Cardinality
- Country abbreviation
- multiple columns allowing Working Group members to specify whether an Attribute can be provided (Y) or not (N))
- multiple columns allowing Working Group members to specify whether an Attribute is needed (Y) or not (N))

By no means the tables will replace the collaborative tool selected. The latter will still be home to the data model 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 an attribute can be provided or not, the Working Group members have to be redirected to the collaborative tool selected.

Ultimately, the Working Group members have to come to a semantic agreement with regards to the data model reviewed. Unless there are major semantic changes, this step should be considered as a formal approval from the Working Group members for the data model.

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 model?
- 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?
- What information does my country need to process the evidence?

Example(s) 'Editable table' as described further above:

		Descr	Cardi																																
Attribute		ption	nalit	A	B	B	H	C	C	D	E	F	F	D	E	H	I	I	L	L	L	L	M	N	N	P	P	R	S	S	S	E	S		
		n	y	T	E	G	R	Y	Z	K	E	I	R	E	L	U	S	E	T	V	I	T	U	T	L	O	L	T	O	K	I	S	E		
Birth	Evidence	BirthEvid																																	
		ence.iden [Link] [1..1]								Y		Y															Y		Y				Y	Y	
		tifier																																	
		BirthEvid																																	
		ence.issui [Link] [1..1]								Y		Y																Y		Y				Y	Y
		ngDate																																	
		BirthEvid																																	
		ence.certi [Link] [1..1]								Y		Y																Y		Y				Y	Y
		fies																																	
		BirthEvid																																	
		ence.issui [Link] [1..1]								Y		Y																Y		Y				Y	Y
		ngAuthor																																	
		ity																																	

Step 24 Update the final model

Review - formal assessment potentially leading to changes.

Key activities

- The **Editors** process any last feedback and finish the final model.

Description

As the Working Group members have given feedback in the previous two steps, the Editors process these comments and make changes to the data model as agreed with the Working Group members. From this point, the Editors can only make changes for which the Working Group members have 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 members.

Rules and Guidelines

- No change - not agreed upon by the Working Group - is made.
- The change log is updated to reflect the final changes in order to achieve full transparency towards the Working Group.
- Every element, e.g. attributes, needs to have a persistent identifier alongside labels that could be in different languages.

Introduction Methodology Phase 1 Phase 2 Phase 3 Phase 4 Phase 5 Phase 6 Quality
 Review cycles and consensus Stakeholders Terminology



Phase 6: Create distributions and publish documentation



Quick links:

- [Step 25 Decide on the conformance requirements and develop a conformance statement](#)
- [Step 26 Create distributions](#)
- [Step 27 Publish all documentation](#)

Navigate to the different phases

[:arrow_left: Previous phase](#)

Step 25 Decide on the conformance requirements and develop a conformance statement

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** 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 Editors then include a conformance statement in the common data model.

It is possible that the data 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 26 Create distributions

Technical analysis - *identification of technical requirements and related solutions.*

Key activities

- The **Editors** 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 (RDF/turtle), RDF/XML, JSON-LD, SHACL, etc. Special care needs to be taken when using multiple formats, as conversion 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 provide a clear overview. This can for example be a UML-diagram, saved as a PNG-file. Next to this, human-readable documentation is also 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 and 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.

Rules and Guidelines

- Create both machine-readable as well as human-readable distributions of the data model.
- Automate, if possible, 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.
- UML diagrams can be published in machine-readable formats, e.g. XML.

Tool(s)

- [VocBench3](#)
- Sparx Enterprise Architect
- [Protégé](#)

Example(s)

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

Step 27 Publish all documentation

Technical analysis - identification of technical requirements and related solutions.

Key activities

- The **Editors** publish all documentation on the collaborative tool.

Description

The Editors publish the final version of the data model, in both machine-readable and human-readable formats, on the selected collaborative tool. The Editors must publish the data model as open (meta)data and specify which license is applicable.

Tool(s) The collaborative tool, e.g. Confluence, Github. Ideally, a collaborative tool allowing public access is more appropriate for transparency reasons.

[Review cycles and consensus](#) [Stakeholders](#) [Terminology](#)

Quality

Quality comes in three different shapes.

data models

This is ensured by using the **proposed methodology**, which is based on the existing SEMIC methodology. In addition, we build as much as possible on **existing resources**, like the ISA² Core Vocabularies, the Public/eJustice documents, EUCARIS, EU Vocabularies of the Publications Office etc., taking into account the **feedback and suggestions of the member states**, building consensus and delivering detailed documentation.

instance data

[the actual evidences to be exchanged] in terms of **correctness of the XML data** with respect to the data models: this can be supported by tools like **the Interoperability testbed** and can be included as a post-development step after phase 7 (finalisation) of the methodology.

source of data

This ensured by the requirement that all data comes from **authoritative sources**. Member states are responsible to identify and connect the authorities to the system.

[Introduction](#) [Methodology](#) [Phase 1](#) [Phase 2](#) [Phase 3](#) [Phase 4](#) [Phase 5](#) [Phase 6](#) [Quality](#)
[Review cycles and consensus](#) [Stakeholders](#) [Terminology](#)

Review cycles and consensus

The process through which semantic agreements can be reached among working group members in a consensus-building activity.

Consensus

Consensus is a generally accepted opinion or general agreement among a group of people.

Consensus is the heart of the process to develop common data models. It aims at developing a collective output, which is the reflection of the greatest possible number of views.

Indeed, consensus involves looking for solutions that are acceptable to all. When everyone agrees with a decision, they are more likely to implement it and, in our case, ultimately use the common data models being built. Consensus is built through iterations, called review cycles.

In the process defined, consensus takes the form of proposals shared, valued and debated to work towards [semantic agreement](#). Semantic agreements aim to meet everyone's most important needs and find a balance between what different Working Group members want, while bearing in mind data minimization and data sensitivity.

Transparency and record keeping are important aspects of achieving consensus. Therefore, all proposals must be debated and documented.

Once a proposal has been dealt with, stakeholders are informed of the group's decision and reasoning. However, there may be times when consensus cannot be reached on an issue or on a comment received. In such cases, one possible course of action is to seek external guidance.

Review cycle

Review cycle is when a (working) draft model is shared with the Working Group so that the members can provide comments and proposals for change. It is during this activity that the consensus is built.

All stakeholders should bear in mind to always ensure that the broadest possible consensus is achieved when a review cycle is carried out. Once reviewed, proposals are categorized and addressed, leading to a new version of the (working) draft model.

Introduction Methodology Phase 1 Phase 2 Phase 3 Phase 4 Phase 5 Phase 6 Quality
 Review cycles and consensus [Stakeholders](#) Terminology

Stakeholders

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

Roles and responsibilities

This section describes the stakeholders identified in the process of developing data models along with their roles 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 cases where no consensus could be reached.

In the context of the SDG Work Package 4, the European Commission is taking this role.

Working Group members

The Working Group members contribute to the different deliverables and help others to meet the incremental goals and deadlines mutually agreed upon upfront. Working Group members will be responsible for achieving consensus.

Ideally, knowledge of the SDG is required and semantic awareness is recommended.

In addition to the core activities - defining data models - it is important for the Working Group to know how the output of this methodology will fit the technical aspect of the SDG OOP. For example, "how are the data models going to be used in the exchange of information?". This requires IT knowledge which could be included under

the responsibility of the Editors or by including an IT representative of the SDG OOP in all relevant activities of the methodology.

In the context of the SDG Work Package 4, the Working Group is composed of representatives of Member States. Representatives attend the webinars and coordinate the work at the national level. It was recommended to have not only people with “semantic awareness” but also data modellers and data stewards.

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. Domain experts should be reachable and available throughout the development of the data model.

In the context of the SDG Work Package 4, one expert per domain should ideally be reachable by the representatives of Member States composing the Working Group. 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.

Editors

The Editors lead the drafting of the deliverables and specification (i.e. data model) by integrating and consolidating the input received from the Working Group. Specifically, the role of the Editors 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.

In the context of SDG Work Package 4, the editors are external to the European Commission and the Working Group. They are responsible for doing the groundwork, collecting and aggregating the input.

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. The moderator communicates with other Work Packages to ensure alignment.

In the context of SDG Work Package 4, the moderator is an official of the Commission, who is in contact with other work packages as well as the directing bodies.

Rapporteur

The rapporteur collects input from the Working Group, ensures that the Working Group is on time in respect of the deadline of each deliverable in collaboration with the moderator. Also, both the moderator and rapporteur communicate with other Work Packages to ensure alignment. The rapporteur is drawn from the Working Group.

In the context of SDG Work Package 4, the rapporteur is a member of the Working Group. At the outset, Working Members were given the possibility to take up the role of rapporteur.

Introduction Methodology Phase 1 Phase 2 Phase 3 Phase 4 Phase 5 Phase 6 Quality
Review cycles and consensus Stakeholders [Terminology](#)

Terminologies

This pages contains the definitions (and illustrations) of the different concepts and terms used throughout the repository.

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.

Controlled vocabulary

A controlled vocabulary is an authoritative list of terms to be used in indexing. Controlled vocabularies do not necessarily have any structure or relationship between terms within the list.

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.

Evidence

An evidence means any document or data, including text or sound, visual or audiovisual recording, irrespective of the medium used, required by a competent authority to prove facts or compliance with procedural requirements

Procedure

Set of administrative formalities or steps to be followed in order to carry out a request.

Example

Life events	Procedures	Expected output subject to an assessment of the application by the competent authority in accordance with national law, where relevant
Birth	Requesting proof of registration of birth	Proof of registration of birth or birth certificate
Residence	Requesting proof of residence	Confirmation of registration at the current address

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

A specification of a data model and entities for which stakeholders reached consensus.

Vocabulary

A set of concepts and relationships (also referred to as “terms”) used to describe and represent an area of concern.