European Commission

26 January 2023

Style Guide for semantic data specifications

interoperable europe

Ground rules for the webinar

Audio Connection

Click “Connect audio” to hear the presenters but please mute your mic when you are not speaking.

Q&A Sessions

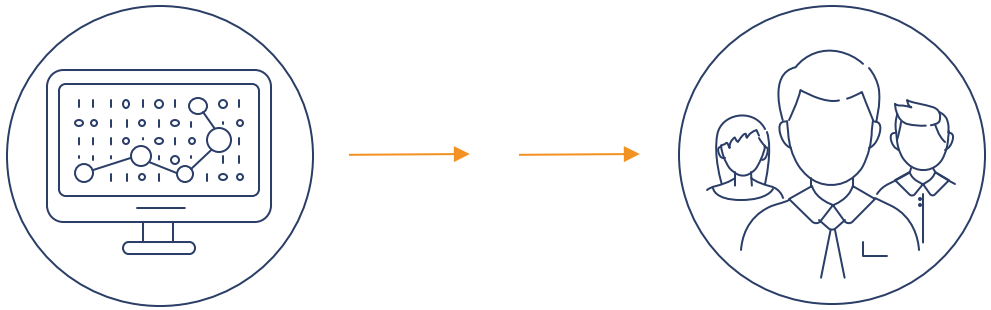
Use the function of raising your hand. We will enable your audio for you to address the speakers directly. You can also use the WebEx chat (optionally).

Technical difficulties?

If you have technical issues with WebEx, please submit your questions in writing to the HOST (privately) by using the WebEx chat.

Meeting flow

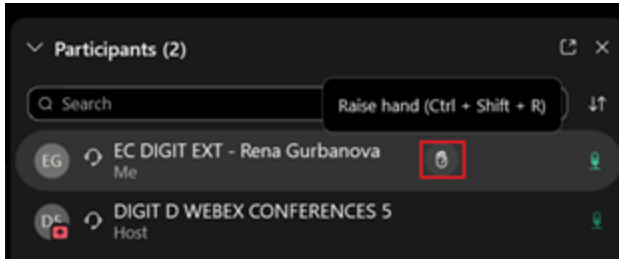
Presentations of the meeting will be followed by Q&A sessions



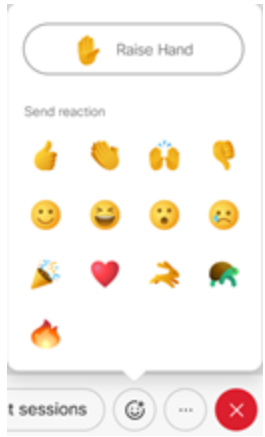
Presentation **Q&A session**

How to raise your hand?

1. Click ‘Participants’ and then click ‘Raise hand’ next to your name.



2. You can also raise and lower your hand from the Reactions menu.



European Commission

Agenda

1. **Welcome**
2. **SEMIC overview & webinar introduction**
3. **Introduction Style Guide**
   - **Audience**
   - **Terminology**
   - **Reuse principles**
   - **Overview of rules & guidelines**
4. **Feedback**
5. **Conclusion**

European Commission

SEMIC and introduction to the webinar

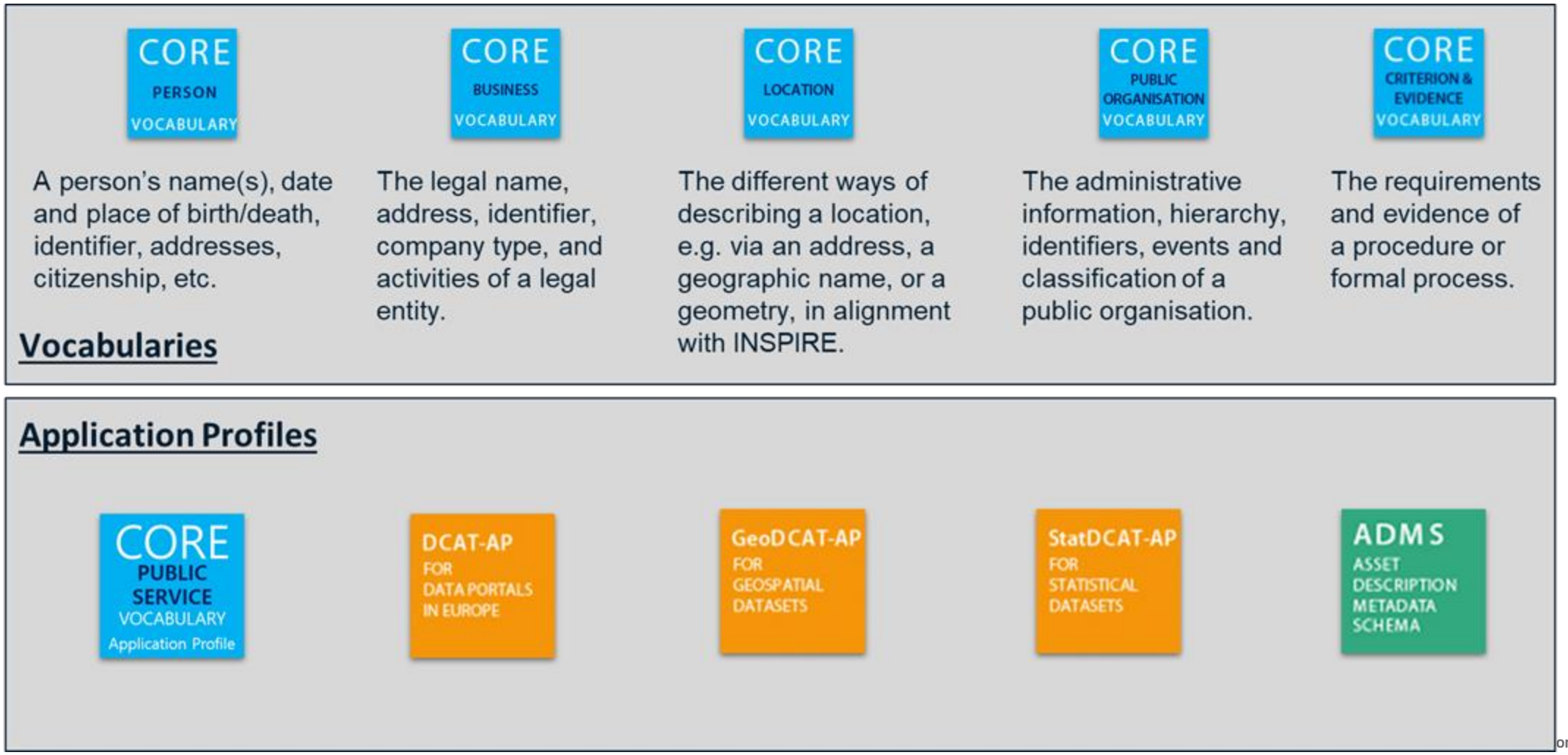
European Commission

Introduction to SEMIC

The objectives of the SEMIC action is to promote Semantic Interoperability amongst the EU Member States by:

- Promotion, sharing and reuse of semantic assets, experience and tools and facilitating agreements in key areas.
- Identifying opportunities for alignment on semantic definitions, metadata and reference data sources with special focus on identification and definitions of Core Concepts / Vocabularies.
- Raising awareness on the importance of data and metadata management.

Current SEMIC assets



Objective of this webinar

Addressing the question:

how do we build SEMIC assets?

https://semiceu.github.io/style-guide/public-review/index.html 

Road to this Style Guide

10 Rules For Persistent URIs

Core Vocabularies Handbook

SEMIC initiate activities to have more coherent and quality assets

Webinars on SEMIC Style Guide and toolkit

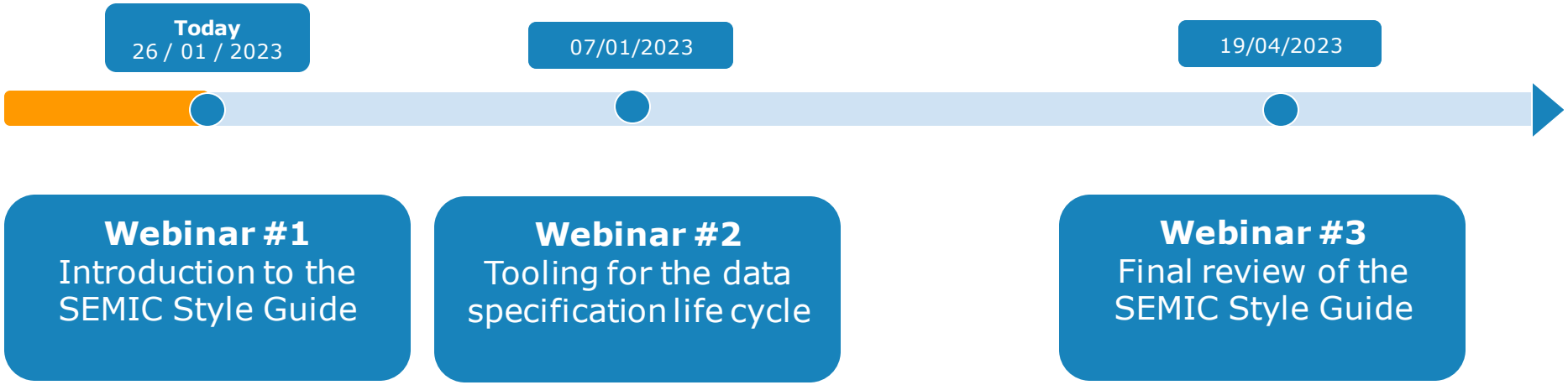
+ Growing experience through practice and piloting with(in) the SEMIC community

Motivation / Context

Share the existing experience and current practices in a document to

- Improve the **quality** of the SEMIC assets
- Improve the **coherency** of the SEMIC assets
- **Streamline and support extension** of Core Vocabularies and Application Profiles by MS
- And, address the challenges of a growing need for data specifications through the **data spaces** initiative

Connected SEMIC activities



Collecting Feedback

Collecting feedback

The full guide: https://semiceu.github.io/style-guide/public-review/index.html

Early feedback already collected from bilaterals:

- Impacted this webinar
- Some of the remarks are already addressed
- Will be taken into account with feedback received today

As usual: feedback via issues on GitHub page https://github.com/SEMICeu/style-guide/issues

Feedback

https://app.mural.co/t/beadvtc7549/m/beadvtc7549/1674502990849/fddfe3cea13387a17a21b75b2283e9b69925490d?sender=ue9933922f73b776a33f43670

Introduction to the style guide

European Commission

Introduction Style Guide

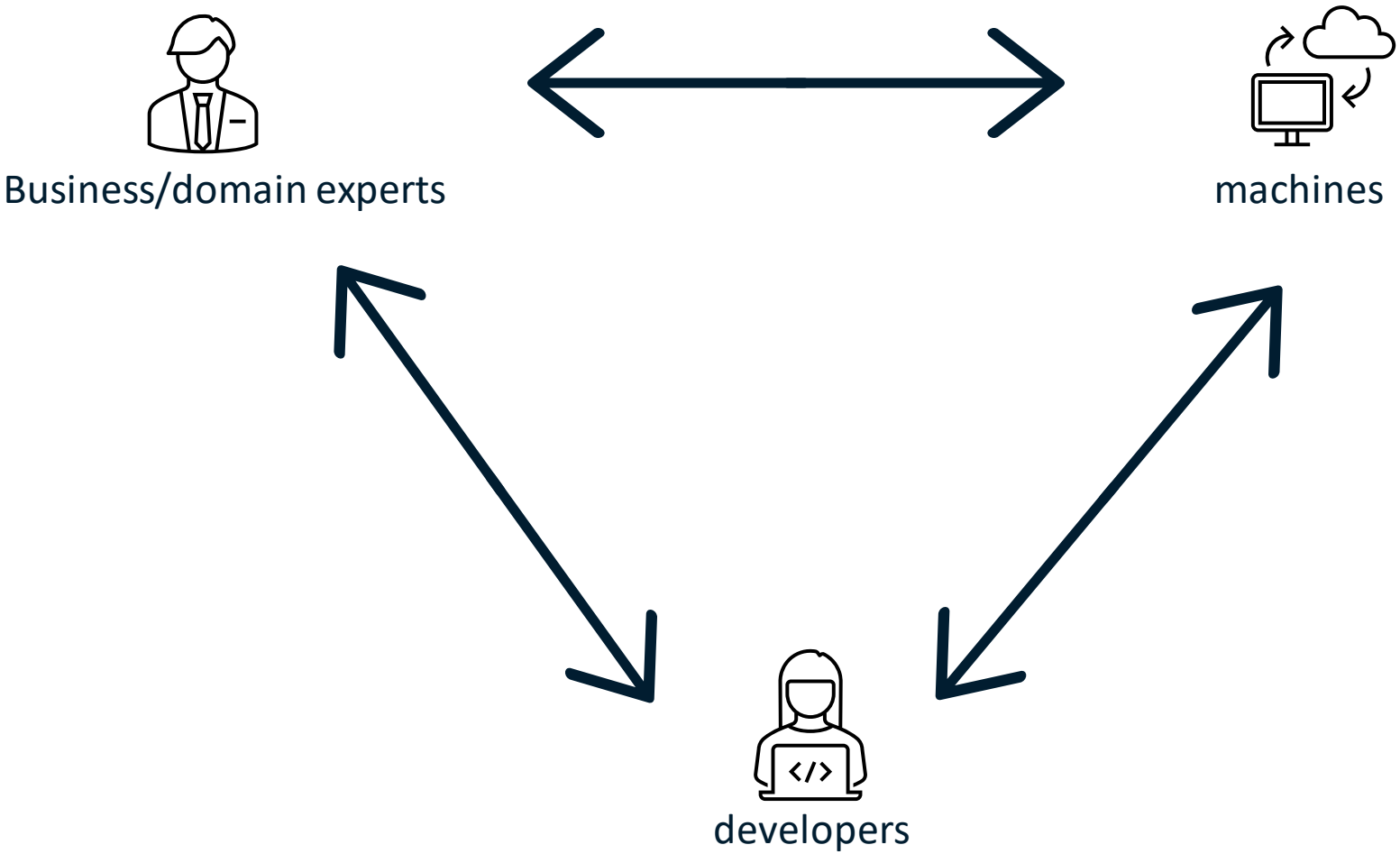
1. **Audience**
2. **Terminology**
3. **Reuse principles**
4. **Overview rules & guidelines**

European Commission

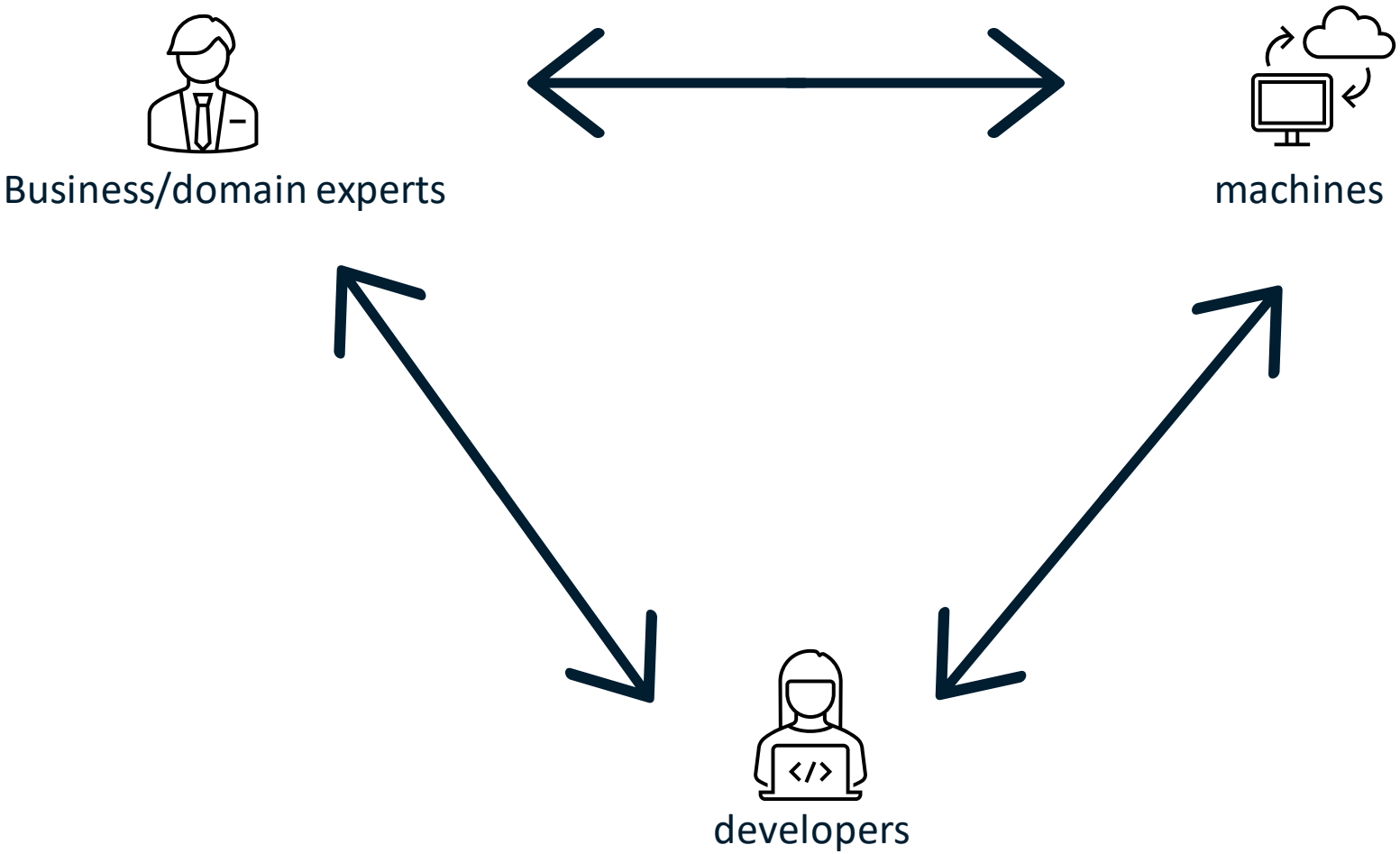
Audience

European Commission

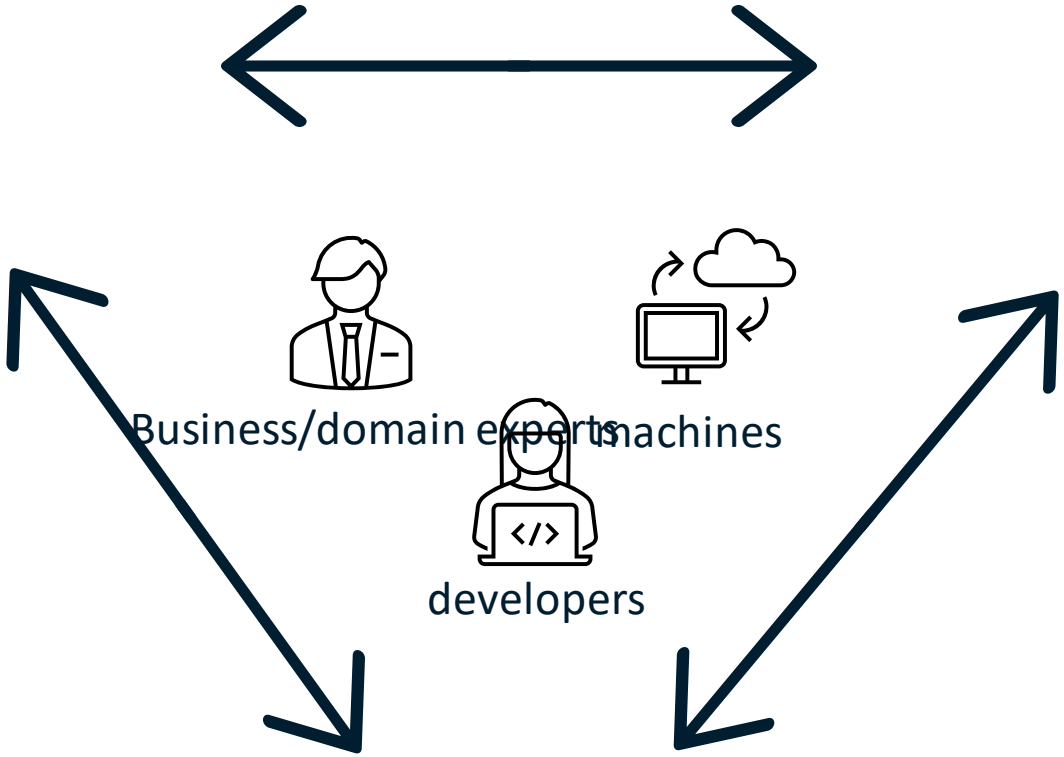
Consumer context



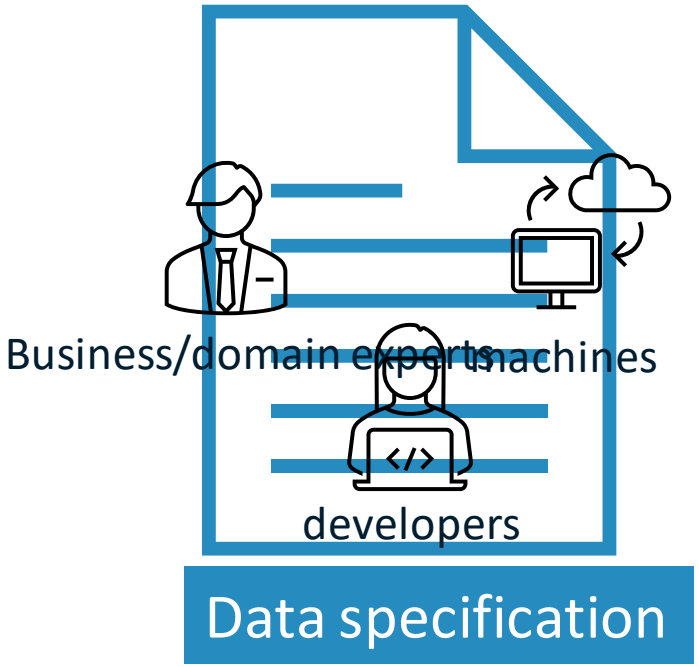
Consumer context



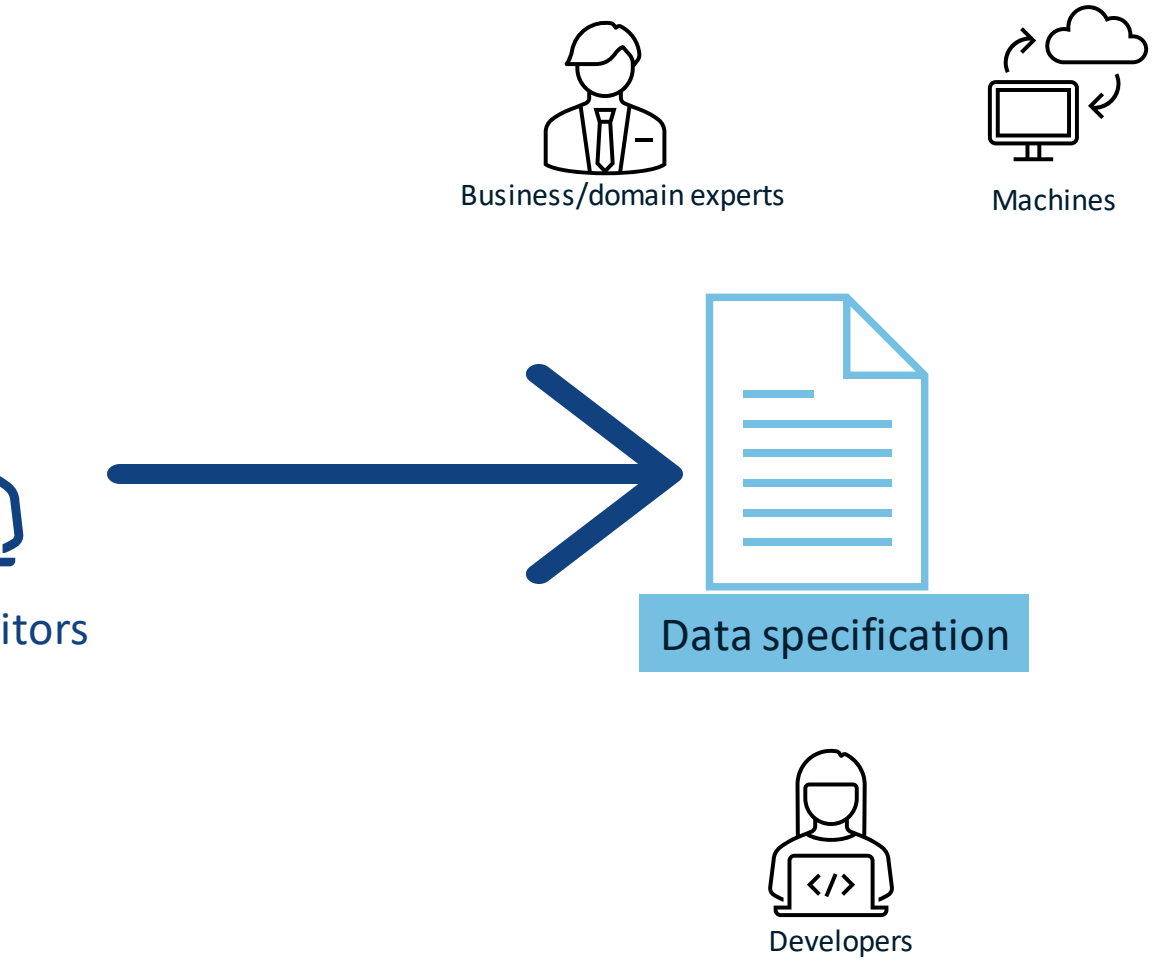
Consumer context



Consumer context



Audience Style Guide



Audience (summary)

Mainly meant for the **editors** of the data specifications (e.g. semantic engineers, data architects, knowledge modelling specialists).

The aim is to ensure the creation of **coherent** data specifications that can be read and used by *domain experts*, *developers* and *machines*.

Scoped to what is needed within SEMIC.

Terminology

European Commission

Disclaimer / challenge

- There is no neutral terminology available.
- Our terminology meaning is overloaded by historic uses and episodic assumptions.
- The same terms are used in different contexts with different expectations
  - For example: the term *vocabulary* or *schema* or *class*

A. Introducing new terms might lead to more confusion.
B. Our terminology is still based on widely accepted terms, which we define/precise in the document.

Semantic asset types

Semantic data specification

- Core Vocabulary
- Application Profile

Collection of terms
+
Structures/rules how to combine these terms
+
(Re)use expectations

Artefacts

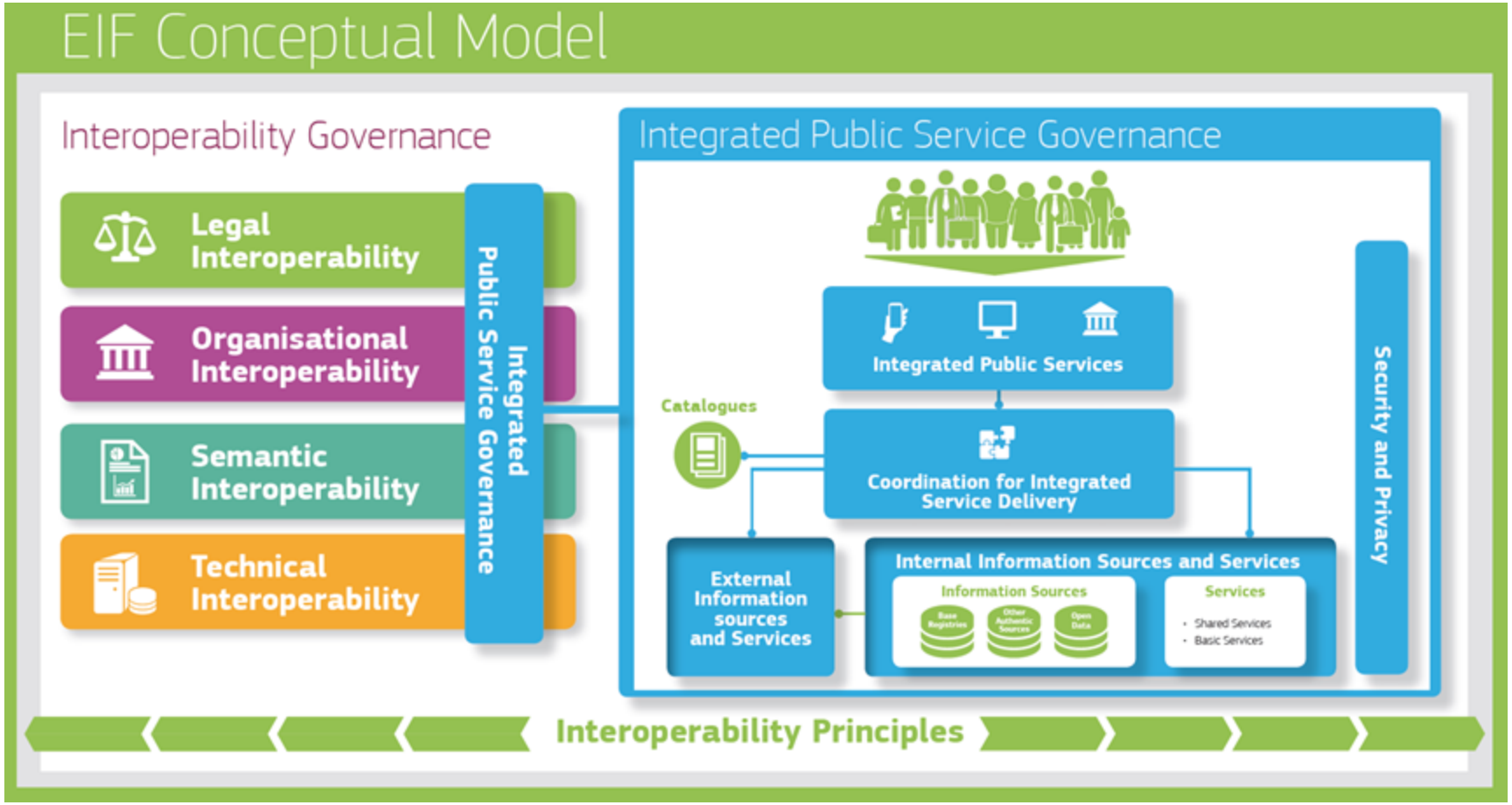
- Persistent URIs
- RDF representation
- OWL representation
- SHACL representation
- HTML representation
- Pictures/Diagrams
- UML representation
- JSON-LD representation
- XML representation
- …

Format
+
Use with a purpose
+
Addressing one concern

Architecture

European Commission

Approach



Semantic interoperability → Technical interoperability

Data specifications composition

Data specification

=

Visual representation

Machine interpretable representation

Human readable text

Data specifications composition

Data specification

=

Visual representation

Machine interpretable representation

Human readable text

Supporting adoption representation

Data specifications composition

Data specification

=

UML

Data Shape (SHACL)

nte en

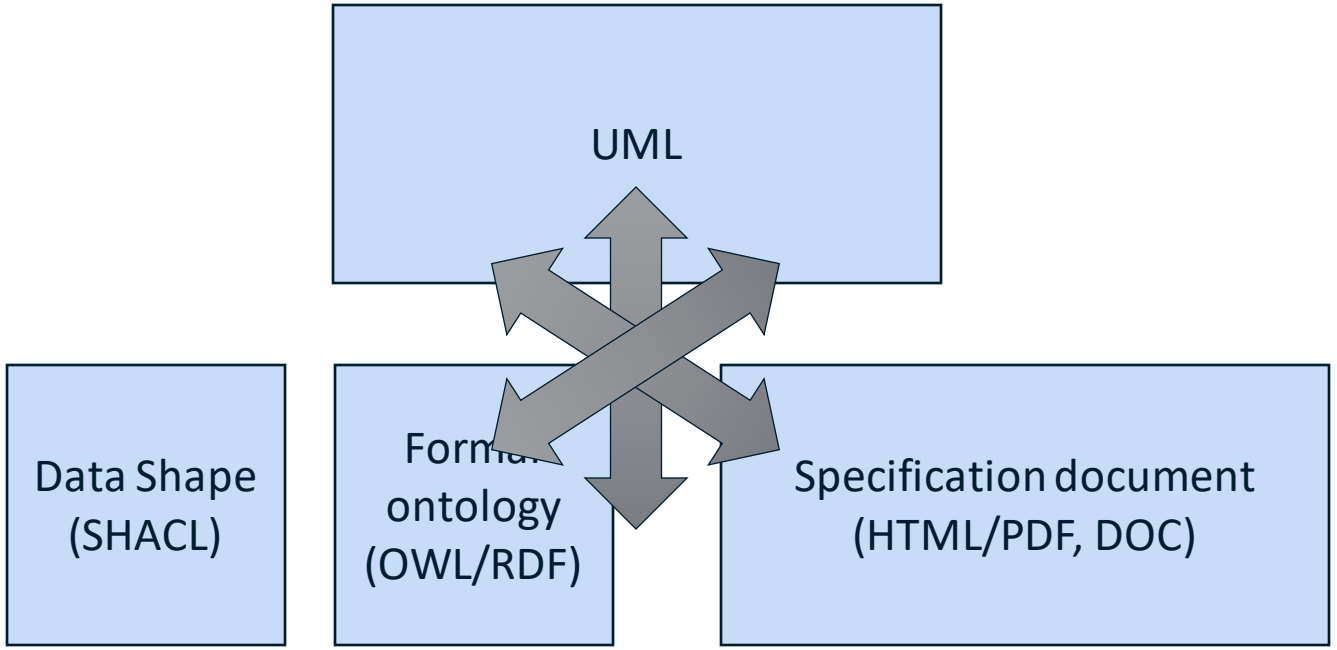
Formal ontology (OWL/RDF)

Specification document (HTML/PDF, DOC)

Editorial support: manual

Data specification

=

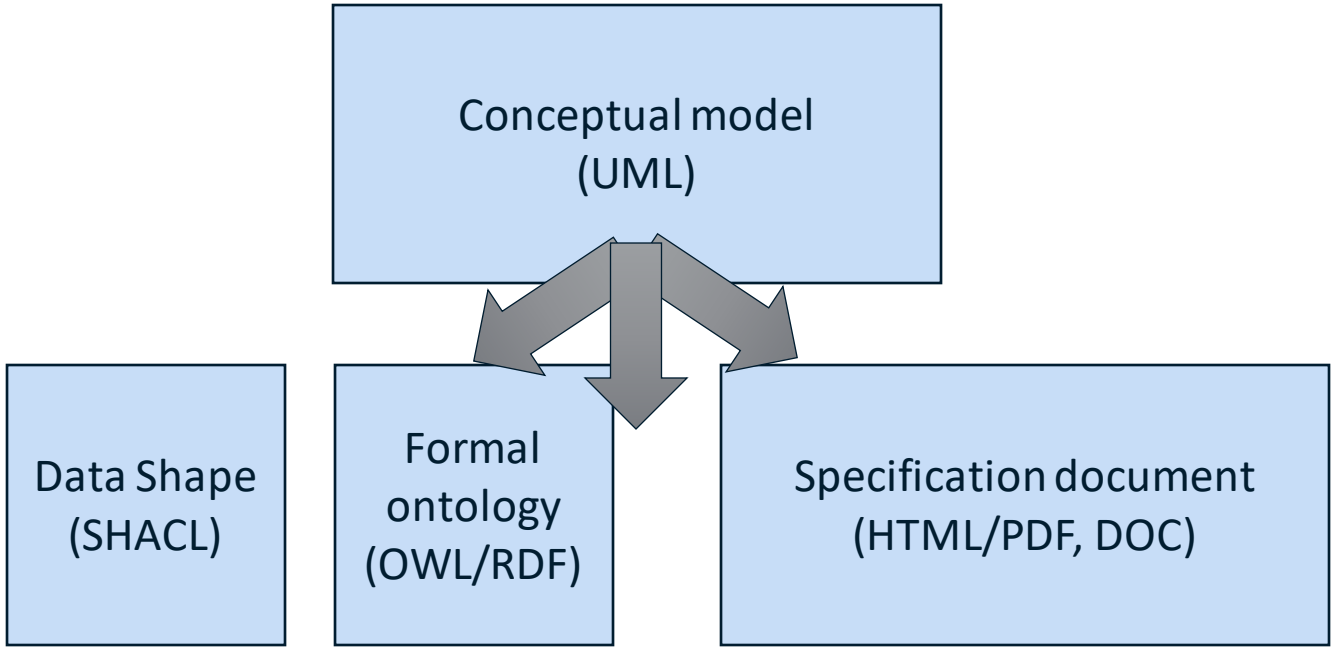


- Keeping the different representations coherent is a tedious and error-prone process

Editorial support: automation

Data specification

=

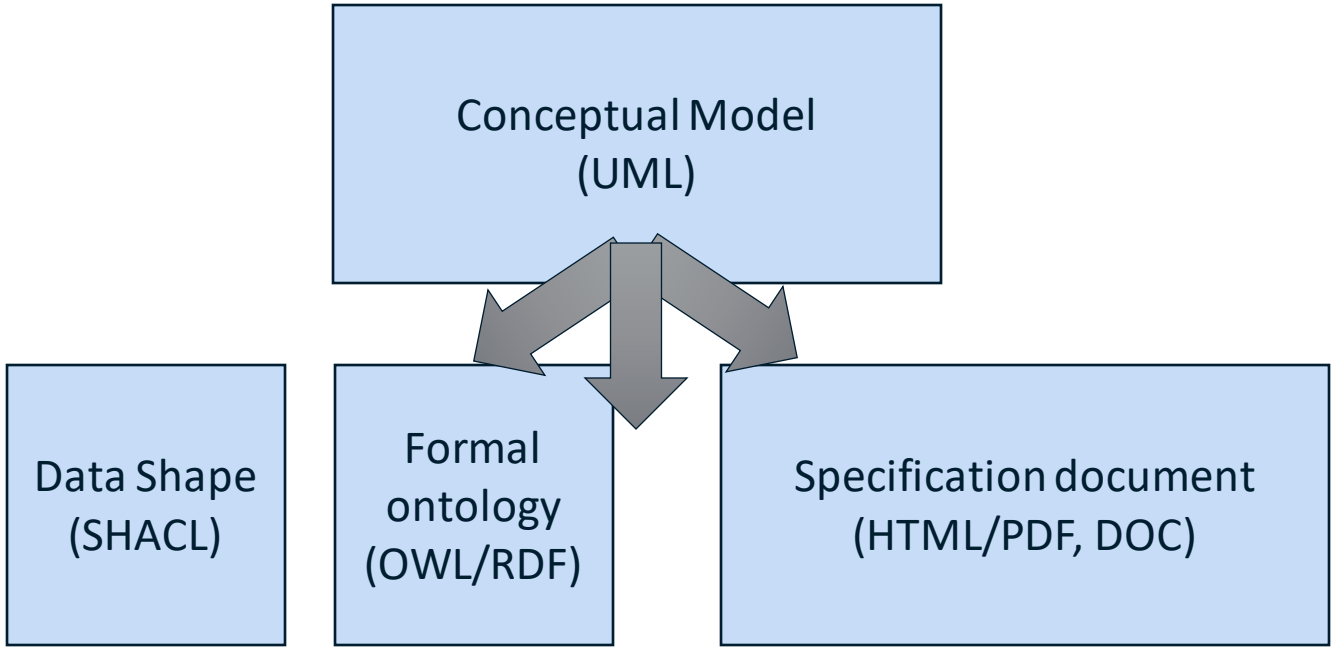


- Conceptual Model is introduced to support automation.
- Decision to make UML (not only a visual representation but) a language for the **conceptual model**.

Motivation for Conceptual Model

Data specification

=



- Practice of using UML as main interaction medium with the domain experts.
- Hard to create correct, coherent UML diagrams from the formal specification (e.g. colour of boxes, location on the diagram, kind of arrows, etc.)
- Realised in two independent *automation tools*

Reuse principles

European Commission

Reuse

Takes advantage from the other's investment & efforts

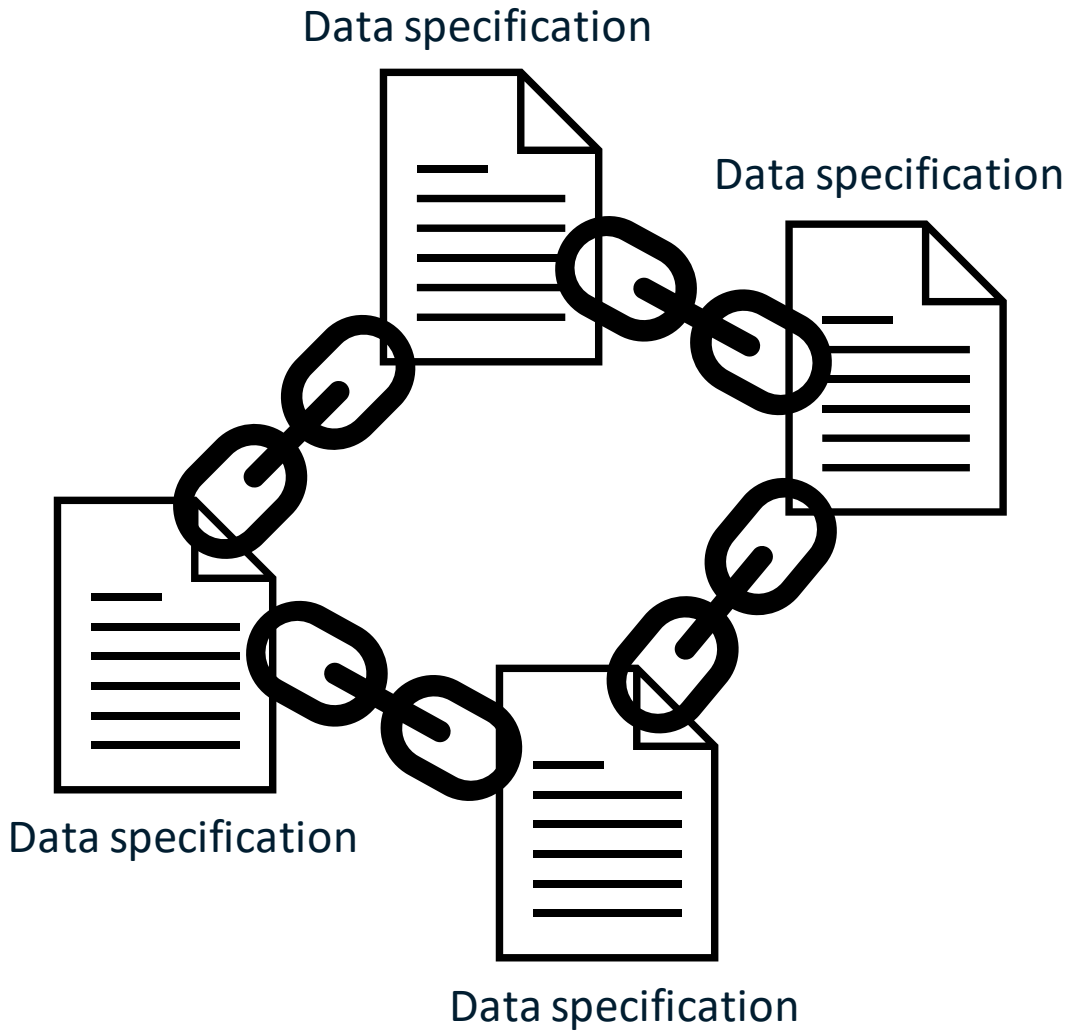
But …

- It means becoming connected with the other
- It means becoming bounded by the other
- It means being limited by the other
- It means to collaborate and invest in the other

Motivation

- How to manage reuse?
- How to apply reuse so that I am connected but not bounded?
- How to facilitate that my work is reusable by others? (i.e. become part of a larger community).
- What does reuse mean and how it is signaled?

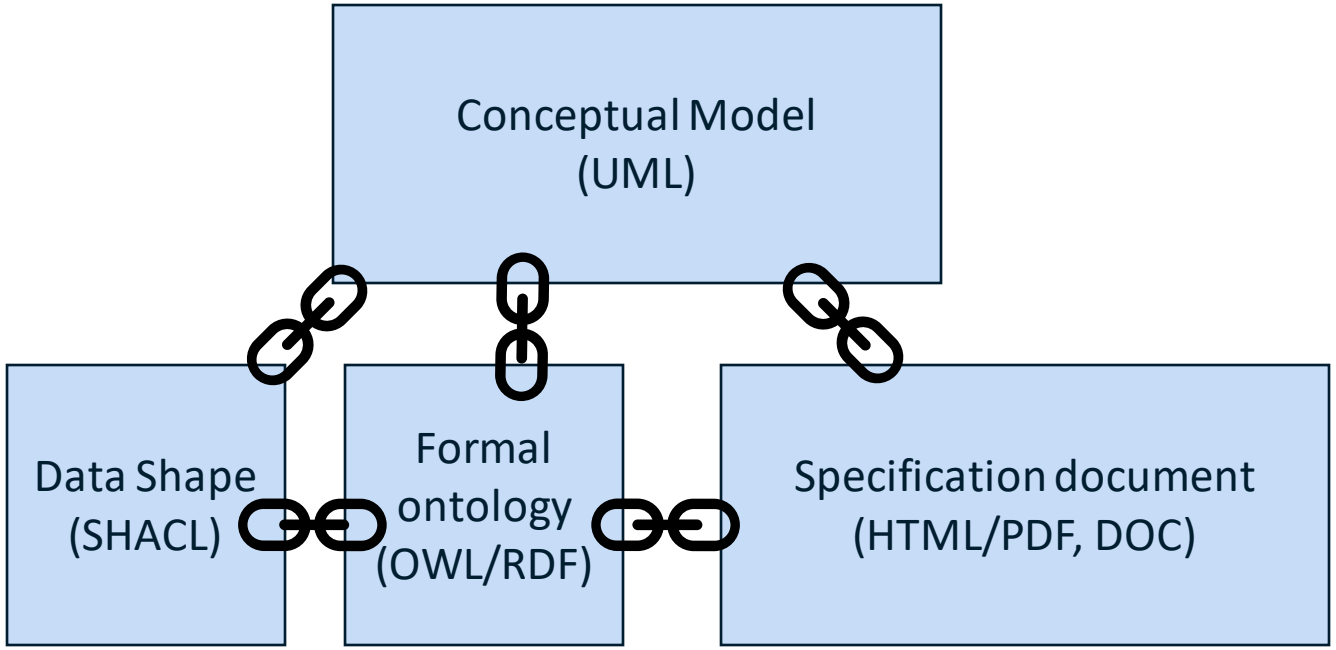
Reuse bounded data specifications together



Reuse binds representations together

Data specification

=



Reuse principles (and the actual Style Guide rules) are designed in such a way that

- Knowledge representation activities in the **conceptual model** have clear effects on the other representations (and implicitly in the derived artefacts).
- The resulting specification is reusable in a coherent way.

Reuse cases

What

- Class
- Property

How

- As-is
- With terminological changes
- Semantic adaptations

Impact

- How to denote reuse
- Label | definition
- (Classes only) Original properties | new properties
- (Properties only) Domain | Range
- Other constraints & additional info

How to denote reuse

Via URI

- The URI is the reference/pointer to the semantics
- Dereferenceable
- Implicit in the conceptual model

Via subclass | subproperty relation

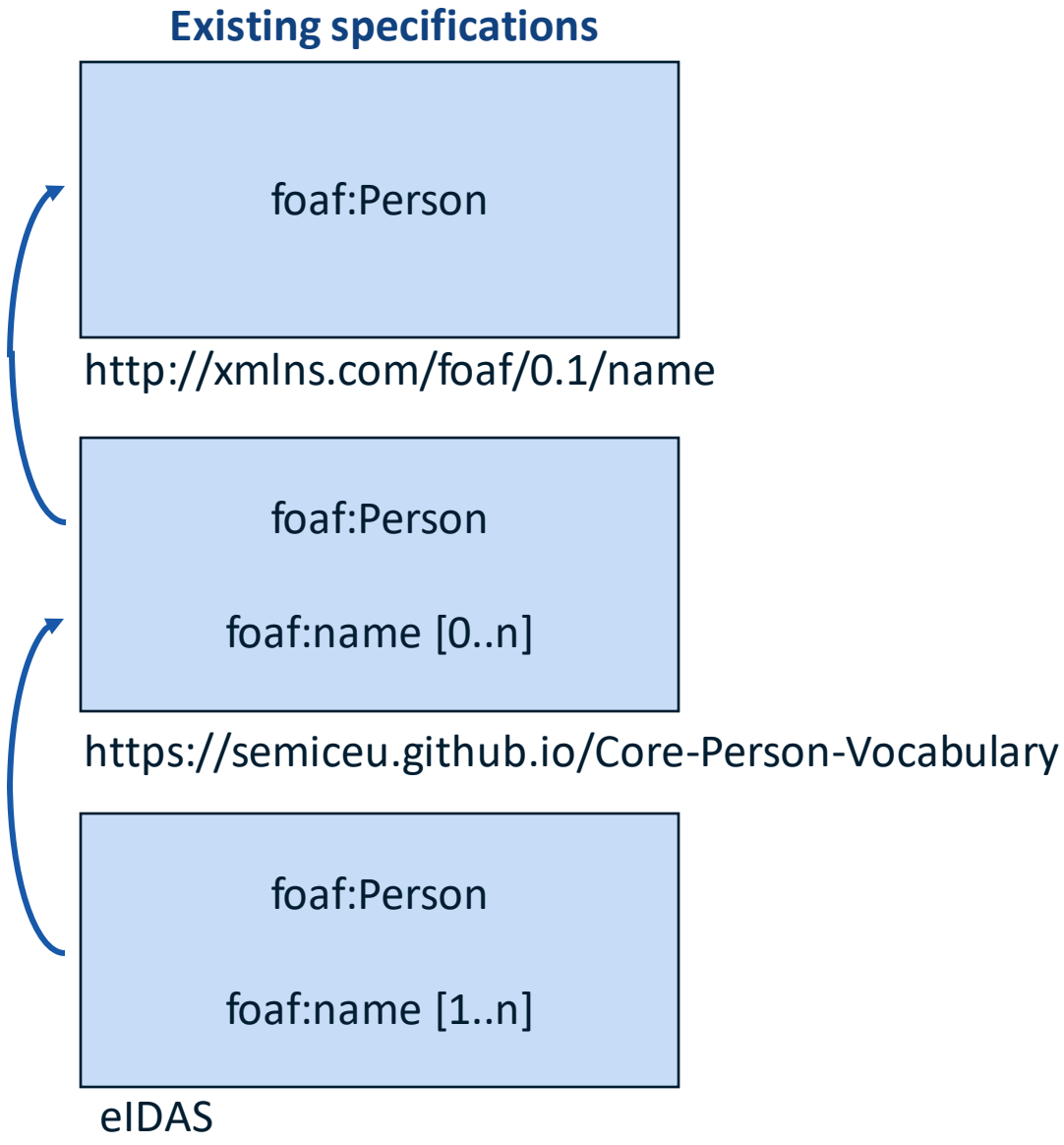
- Is a logic structure
- Explicit in the conceptual model

With notes

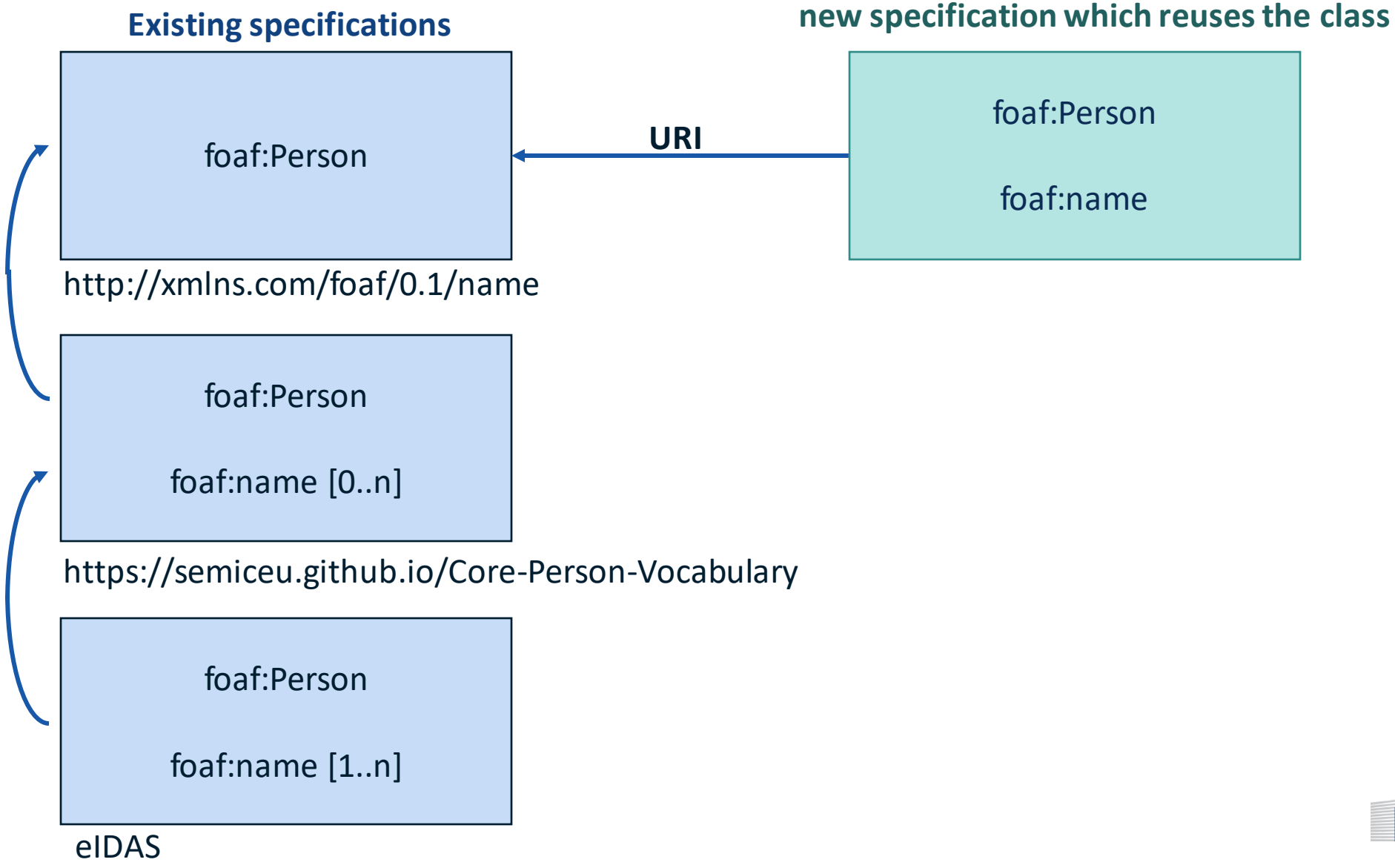
- for human understandability

Goal is to make the reuse chain as clear as possible

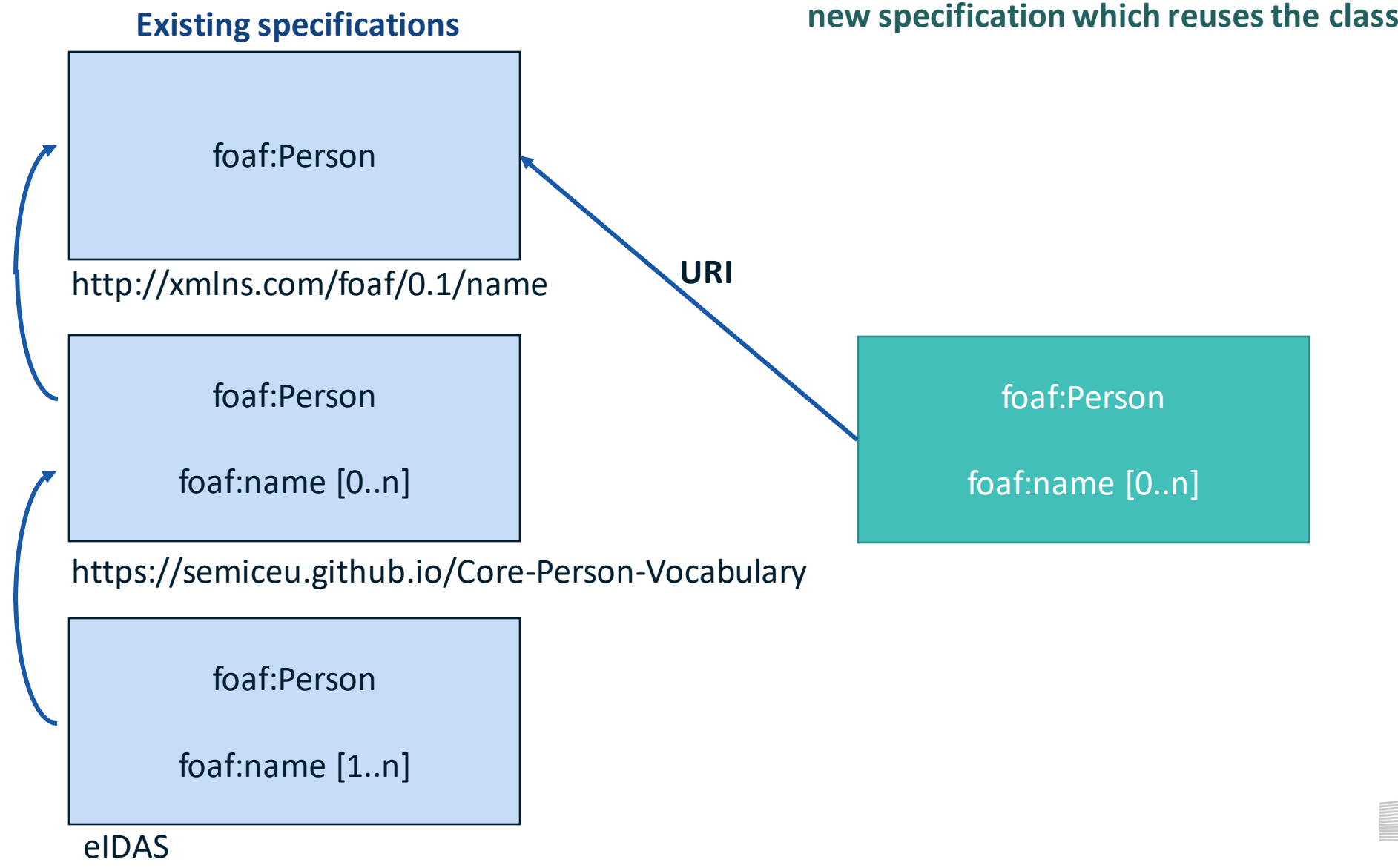
How to denote reuse



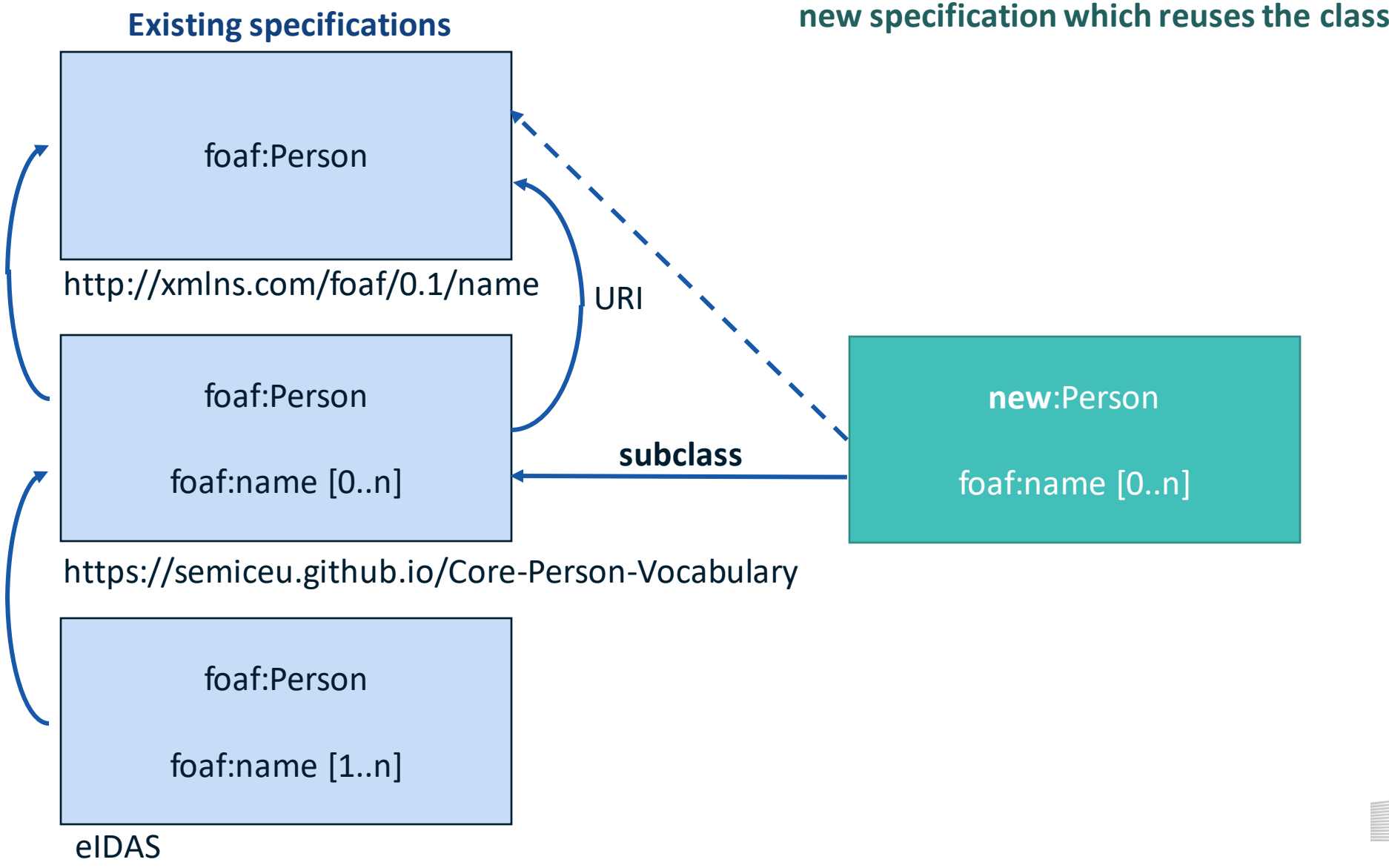
How to denote reuse



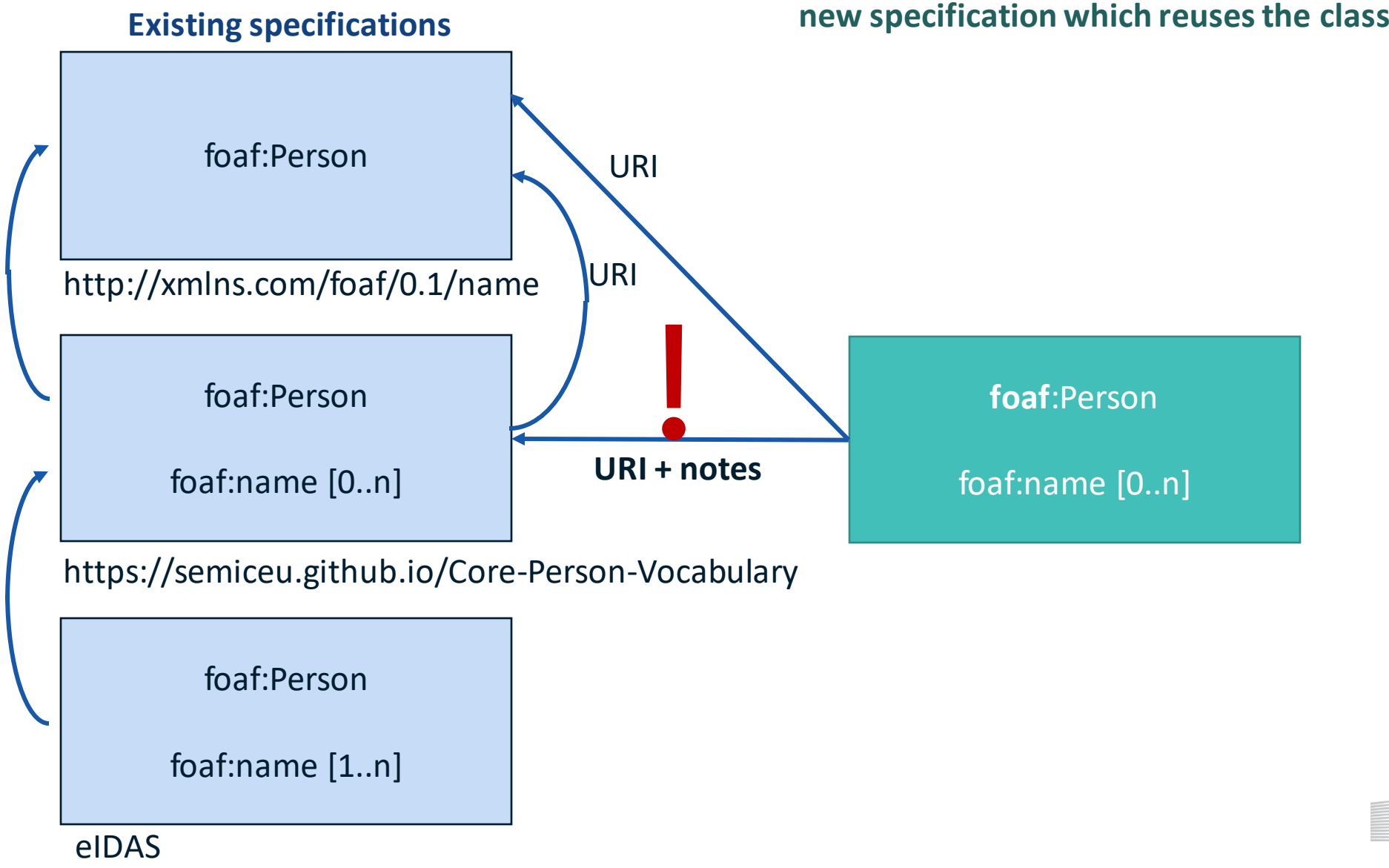
How to denote reuse



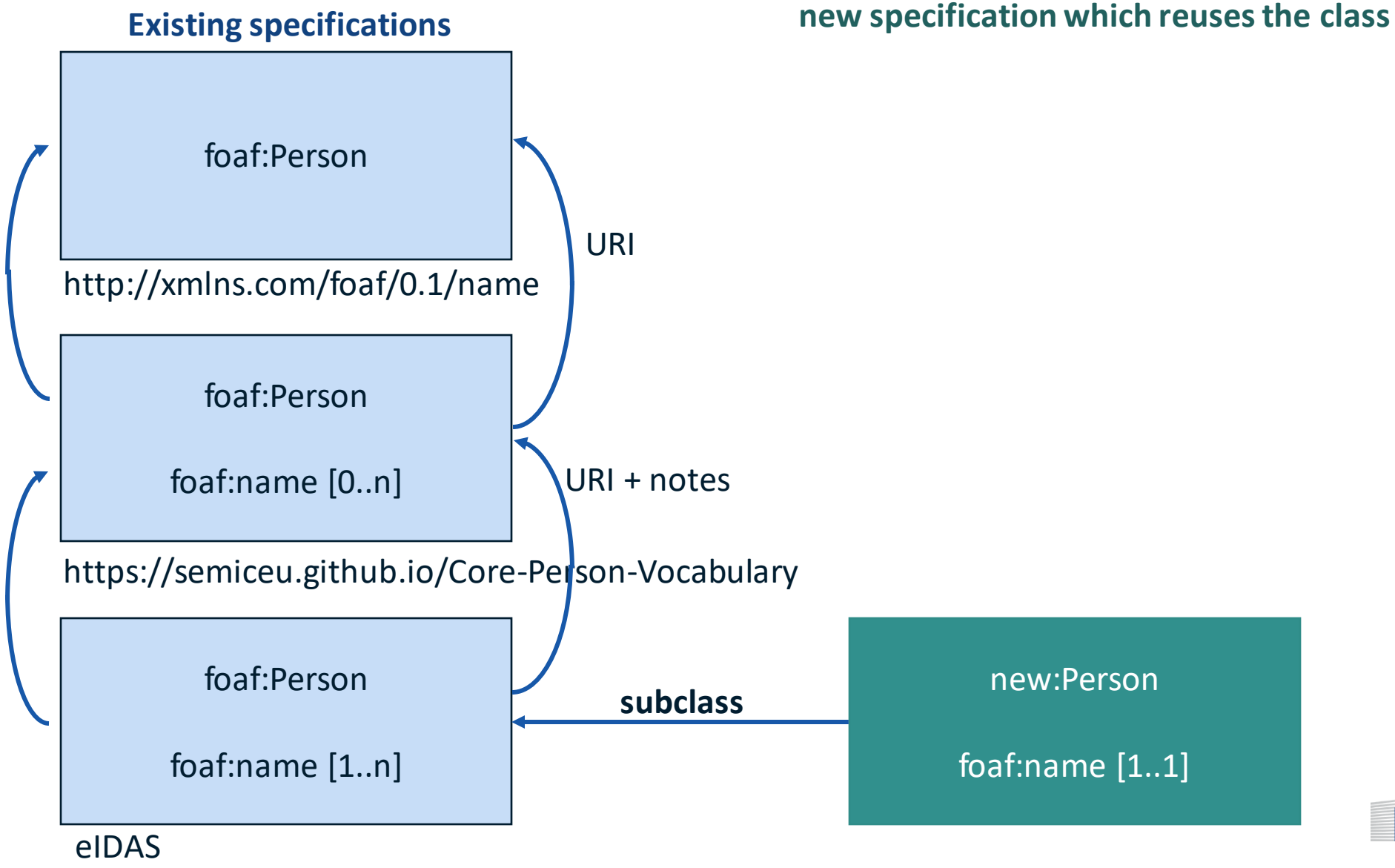
How to denote reuse



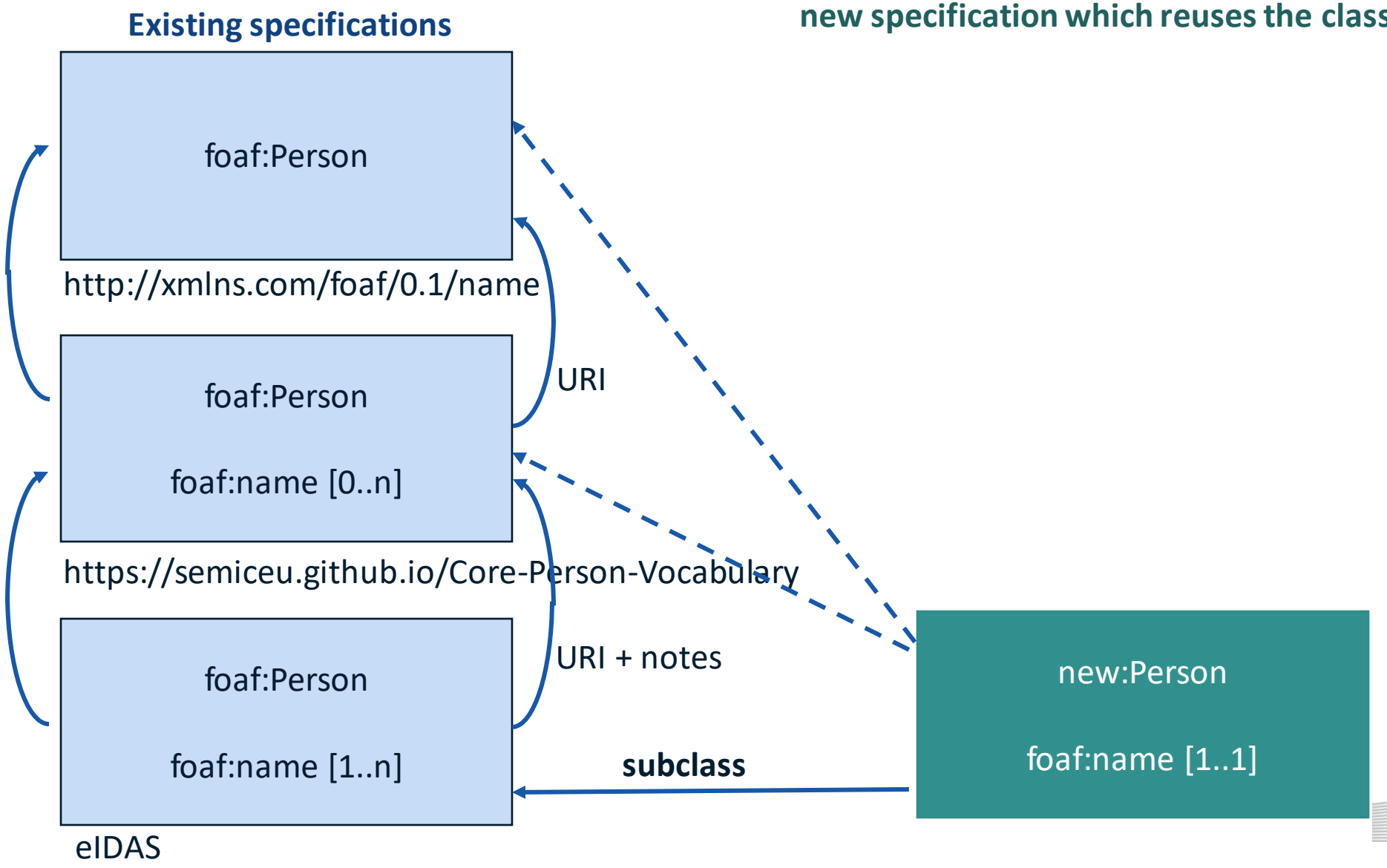
How to denote reuse



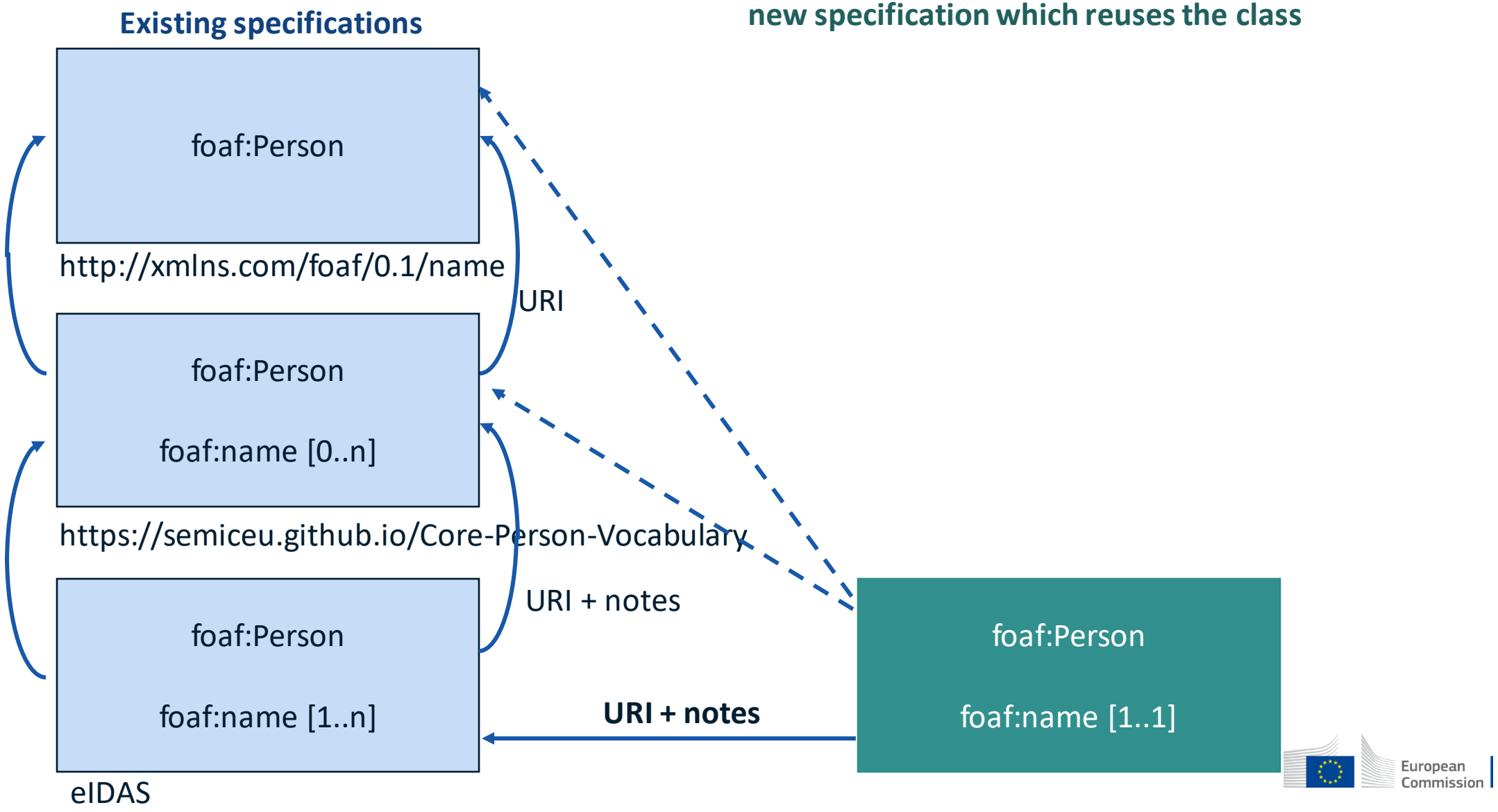
How to denote reuse



How to denote reuse



How to denote reuse



Reuse motivations

As-is

In the specification the term is used in the way as it is being defined in the source specification

With terminological changes

In the specification the term is used in a slightly other wording but with the semantics as defined in the source specification

- Typical case: domain specific words, preferred synonyms

With semantical adaptations

- The term strongly overlaps with the semantics in the source specification, but its usage is independent of that source term, or is narrower.

Reusing Classes

Class	Denotation	Label	Definition	Usage Note
As-is	URI	⊘	⊘	
Term. changes	URI	✓	⊘	explain nuance
Sem. changes	Subclass	✓	✓	

Label *Term. changes*

- New label becomes visible (in UML)
- Original label is maintained
- Original label => skos:prefLabel
- New label => skos:altLabel

Reusing Properties

Property	Denotation	Label	Definition	Usage Note
As-is	URI	⊘	⊘	
Term. changes	URI	✓	⊘	explain nuance
Sem. changes	Subproperty	✓	✓	

Label *Term. changes*

- New label becomes visible (in UML)
- Original label is maintained
- Original label => skos:prefLabel
- New label => skos:altLabel

Scope of a term per semantic data specification type

Vocabulary

- Class and properties are (highly) independent
- May provide a few broad constraints

Application Profile

- Classes and properties are dependent
- Class = class + properties (outgoing arrows)

Scope of term per asset type

Existing specifications

foaf:Person

http://xmlns.com/foaf/0.1/name

Vocabulary = classes and properties are independent

foaf:Person

foaf:name [0..n]

https://semiceu.github.io/Core-Person-Vocabulary

Core Vocabulary = classes and properties are independent

foaf:Person

foaf:name [1..n]

eIDAS

Application Profile = classes and properties are dependent

Reusing Classes

Original properties	New properties
Visibility of 'mandatory properties' is mandatory	Add, if no semantic relation to an original property exists
Visibility of 'optional properties' is optional	Otherwise apply reuse principles of properties
Semantics are not changed, only nuanced	

Basic principle

= Use (inherit) all constraints that are within the scope of the class (and its superclasses)

No property duplication allowed within a class

- No usage of the same URI twice with two different semantics associated
- Avoid conflicts!

Reusing properties

Basic principle

= Use(inherit) all constraints that are within the scope of the property (and its superproperties)

Property has domain or range defined:

- (explicit) in machine readable representation (OWL/RDF)
- (implicit) in the human readable definition

--> Honor them

Reusing Properties

Property	Has Domain or Range	Domain or Range	Other constraints
Via URI	no	Can be added to the desired class (as domain) and with the desired range	Free to add
Via URI	yes	Can only be added if the desired class matches the domain and if the desired range matches the range	Constrained by reuse
Via subproperty	no	Can be added to the desired class (as domain) and with the desired range	Free to add
Via subproperty	yes	Can only be added if the desired class matches the domain of the superproperty and if the desired range matches the range of the superproperty	Constrained by reuse

Reusing Properties: Challenge

Reusing properties may lead to multiple domains and ranges being provided.

- OWL default semantics is Intersection and not Union (as intuition may indicate)
- This is the motivation for recommending expressing domain and range as SHACL statements rather as rdfs:domain/rdfs:range statements.

Quick overview rules

European Commission

Organisation of the rules

- General conventions (5)
- Conceptual model conventions (15)
- Semantic conventions (5)
- Data shape conventions (5)
- Methodology conventions (3)
- Publication conventions (6)

Organisation/formulation of the rules

Conciseness of the document

- Avoid repetition: that is why more rules are in the conceptual model conventions

Balance between precise scope, strong guidance and leaving room for acceptable variations

- Coherent experience / way of working
- By promoting reuse, choices of others get imported:
  - E.g. opaque URIs versus human readable URIs
  - American English vs British English (dcat:Catalog vs Catalogue)

Provision of a minimal set of aspects to be considered

(for reference)

rule	statement
GC-R1	Reuse existing concepts as much as possible, respecting the original semantics and lexicalisation.
GC-R2	Quality of maintenance and governance should be reviewed before reuse. Preference should be given to vocabularies that are well maintained and governed.
GC-R3	Each concept shall be (a) referenceable by a URI, (b) formally defined and (c) described by a precise, unambiguous human-readable label and definition.
GC-R4	The terminology style shall be consistent across the vocabulary.
GC-R5	The concept definitions shall be elaborated consistently across the vocabulary.

(for reference)

rule	statement
CMC-R1	The UML conceptual model should be used as the single source of truth.
CMC-R2	Using UML as a graphical language to encode a data specification according to the reuse guidelines requires defining a set of pre-established conventions on the UML notation to reach a fixed interpretation.
CMC-R3	All UML element names should be fit for URI generation with clear namespace organisation.
CMC-R4	All UML element names are case sensitive and shall follow the CamelCase convention.
CMC-R5	Element names shall be organised by namespaces. Namespaces may be indicated through prefixes delimited by colon (:) character, forming qualified names.
CMC-R6	Tags can be conveniently used for annotating the elements.

(for reference)

rule	statement
CMC-R7	The UML diagrams should depict how the developed model relates to external (reused) models
CMC-R8	Make sure that the attributes and associations of a superclass apply to all its subclasses.
CMC-R9	Classes that are not intended for instantiation can be marked as abstract.
CMC-R10	UML Attributes shall be used to define properties taking simple data type values. An attribute declaration should specify its datatype and multiplicity whenever possible
CMC-R11	The multiplicity of connectors and class attributes should be specified, indicating the minimum and maximum cardinality. The cardinality shall be as permissive as possible in Core Vocabularies and as restrictive as possible in Application Profiles.
CMC-R12	UML Connectors shall be used to define relations and properties taking non-atomic type values. A connector declaration should specify multiplicity whenever possible

(for reference)

rule	statement
CMC-R13	The visibility of all UML elements should be "public".
CMC-R14	The controlled lists of values shall be modelled as enumerations and specified whenever possible.
CMC-R15	Use packages to logically organise the model.
CMC-R16	UML class diagrams shall be organised for readability.

(for reference)

rule	statement
DSC-R1	The data shapes shall be expressed in SHACL.
DSC-R2	In a Core Vocabulary, then the data shape constraints shall be as loose as possible, i.e. permissive, while in an Application Profile, the data shape constraints shall be as rigid as necessary, i.e. restrictive.
DSC-R3	The data instantiating a core model may be fragmented across information systems. For representation purposes, open-world assumptions shall be adopted, and for validation purposes, a closed-world assumption shall be adopted.
DSC-R4	Each ontology class shall be mirrored by a `sh:NodeShape`, while the property constraints shall be embedded (contextualised) within node shapes rather than being defined as freestanding.
DSC-R5	It is recommended to use constraint severity for distinguishing critical violations from non-critical recommendation warnings and from nice-to-have information.

(for reference)

rule	statement
MC-R1	It is strongly recommended that a clear, explicit methodology is adopted for the content development and lifecycle management of a semantic data specification. Conversely, no data specification shall be developed and managed without following a clear methodology.
MC-R2	A semantic data specification shall have clear goals and a well-established scope definition.
MC-R3	If a domain contains a lot of concepts that have to be modelled, then it is recommended to split it into subdomains and manage them as modules.

(for reference)

rule	statement
PC-R1	Publishing the Core Vocabularies and Application Profiles shall be compliant with 5-star LOD criteria.
PC-R2	• The URIs that are used to identify the terms modelled in the data specifications should be *dereferenceable Persistent URIs* [1] and comply with a URI policy (or URI strategy). • URIs that are well-defined according to such criteria are often denoted as PURIs [1]. • SEMIC adheres to the EU persistent URI policy. Each SEMIC URI is dereferenceable in both machine representation (RDF artefacts) and human-readable (HTML artefact).
PC-R3	A consistent version management shall be implemented for the vocabulary evolution respecting the “semantic versioning” principles.
PC-R4	The vocabularies and Application Profiles shall be versioned as a whole and never to atomic elements (concepts, relations or constraints).
PC-R5	Any URI identifiable resource devised in a data specification shall be dereferenceable.
PC-R6	Each artefact shall have a corresponding human-readable form representing the model documentation.

Short break

Feedback

European Commission

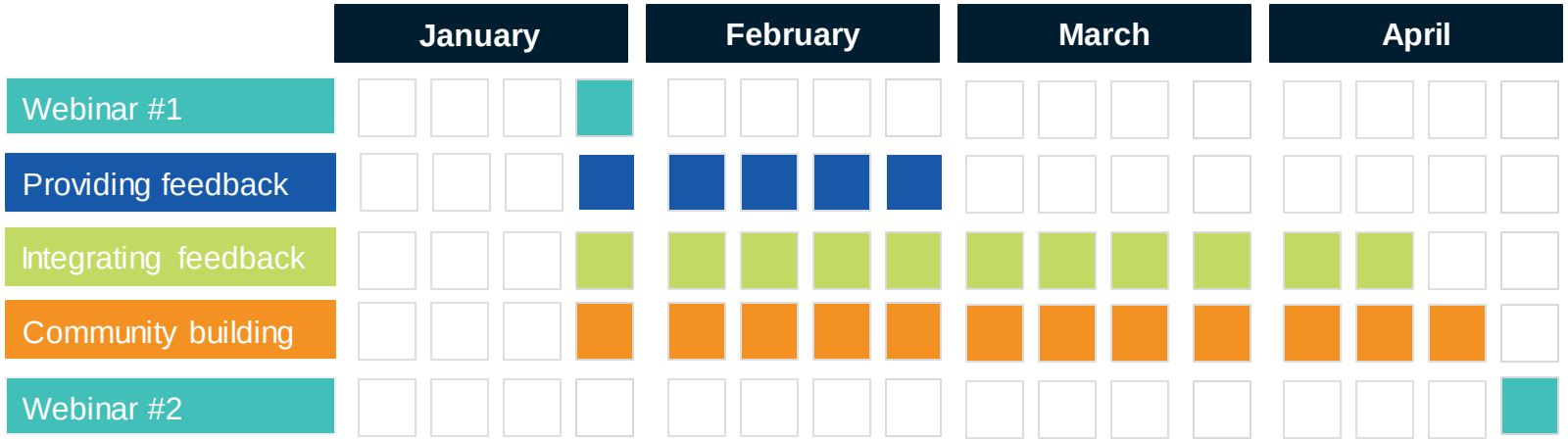
4 groups

- Generic questions/feedback
  - Missing topics / perspective
  - Organisation / style / approach
- Feedback on terminology + architecture
- Feedback on reuse
- Detailed rule feedback:
  - What should not be done / bad practice

Wrap up

European Commission

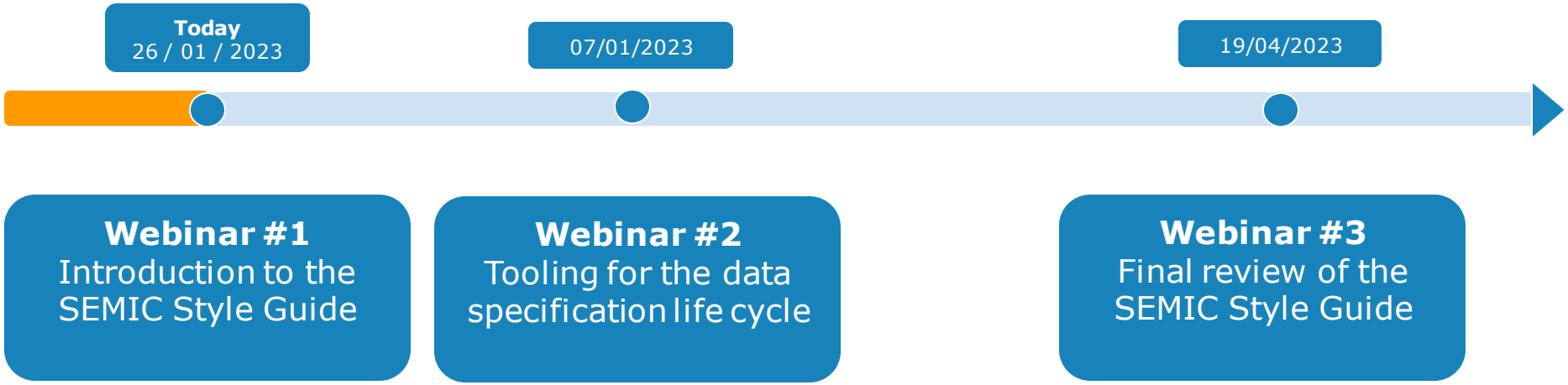
Next steps



The style guide: https://semiceu.github.io/style-guide/public-review/index.html

Feedback via GitHub: https://github.com/SEMICeu/style-guide/issues

Part of the activity around building semantic data specifications



SEMIC Support Centre

European Commission


Thank you

European Commission

interoperable europe

innovation ∞ govtech ∞ community

Stay in touch

(@InteroperableEU) / Twitter

Interoperable Europe - YouTube

Interoperable Europe | LinkedIn

DIGIT-INTEROPERABILITY@ec.europa.eu

https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe