# Meeting Minutes: Webinar on review of Core Vocabularies and Style Guide Blog post (SEMIC - A04.01)

| Project: | SEMIC | Date and Time: | 09/04/2024 10:00 - 12:00 |
|---|---|---|---|
| Meeting Type: | Webinar | Location: | Virtual |
| Coordinators: | Pavlina Fragkou | Issue Date: | 30/04/2024 |

| Agenda of the webinar | | |
|---|---|---|
| 13:00 - 13:10 | Introduction | Slides 1 - 7 |
| 13:10 - 13:20 | CPOV | Slides 8 - 11 |
| 13:20 - 13:40 | CPSV-AP | Slides 12 - 16 |
| 13:40 - 13:50 | CV Related Assets | Slides 17 - 22 |
| 13:55 - 14:55 | Blog Post | Slides 23 - 33 |
| 14:55 - 15:00 | Wrap-up & next steps | Slides 34 - 38 |

| Meeting Slides |
|---|
| LINK |

| Participants | | |
|---|---|---|
| Name | Initials | Organisation |
| Alberto Abella | AA | Fiware |
| Alexandros Gerontas | AG | University of Macedonia, Greece |
| Anastasia Sofou | AS | SEMIC Team |
| Anastasija Nikiforova | AN | European Open Science Cloud |

| Participants | | |
|---|---|---|
| **Name** | **Initials** | **Organisation** |
| Andreea Pasare | AP | Meaningfy |
| Annika Strupp | AS | Init, Germany |
| Arthur Schiltz | AS | SEMIC Team |
| Benny Lund | BL | Swedish Companies Registration Office |
| Bert Van Nuffelen | BVN | SEMIC Team |
| Cornelia Strauß | CS | Init, Germany |
| Costas Simatos | CS | Sigma Cubed |
| Cristian Vasquez | CV | Publications Office |
| Csongor Nyulas | CN | Meaningfy |
| Emidio Stani | ES | SEMIC Team |
| Emiel Dhondt | ED | SEMIC Team |
| Eugeniu Costetchi | EC | Meaningy |
| Frans van der Zande | FvdZ | Logius, The Netherlands |
| Igor Trickovic Rifelj | ITR | Ministry of Public Administration, Slovenia |
| Jitse De Cock | JDC | SEMIC Team |
| Jorge Sousa | JS | AMA, Portugal |
| Josema Alonso | JA | SEMIC Team |
| Kees Trautwein | KT | Logius, The Netherlands |
| Kurt Stian | KS | Norwegian Digitisation Agency |
| Lukas Cladders | LC | Init, Germany |
| Marco Aarts | MA | Logius, The Netherlands |
| Mario Bravo | MB | NTT Data |
| Maya Borges | MB | Danish Agency for Digitisation |
| Morten Borrebaek | MB | Norwegian Mapping Authority |
| Pavlina Fragkou | PF | SEMIC Team |

| Participants | | |
|---|---|---|
| **Name** | **Initials** | **Organisation** |
| Peter Bruhn Andersen | PBA | Danish Agency for Digitisation |
| Riitta Alkula | RA | Finnish Digital and Population Services Agency |
| Sebastian Sklarß | SS | Init, Germany |

# Points discussed and decisions taken

| Topic discussed | Outcome |
|---|---|
| **CPOV** | |
| **Issue #31 (& CPSV-AP #123)** <br> Administrative Territorial Unit. | Approved. |
| **CPSV-AP** | |
| **Issue #125** <br> Relating cv:Output and cv:Evidence to a dcat:Dataset. | Approved. |
| **Issue #128** <br> The execution of a Public Service. | Approved. |

# Full Meeting Minutes

| **Welcome & Introduction** <br><br> Slides 1 - 7 <br><br> **Speaker:** Pavlina Fragkou | PF welcomes the participants and presents the objectives of the webinar. <br><br> **Closing of the public review** <br> The various open issues related during the public review of the Core Vocabularies and CPSV-AP will be discussed and attempted to be closed. <br><br> **Gathering input for improvements** <br> Input for those issues that could not be resolved in the latest release for public review will be gathered. <br><br> **Discussing the Style Guide blog post** <br> The Style Guide blog post on mapping the Core Vocabularies to XML will be discussed and community feedback can be provided. <br><br> The objectives of the SEMIC action are to promote Semantic |
|---|---|

| | Interoperability amongst the EU Member States by: |
|---|---|
| | <ul><li>Promoting the share and reuse of semantic assets, experience and tools and facilitating agreements in key areas.</li><li>Identifying opportunities for alignment on semantic definitions, metadata and reference data sources with special focus on identification and definitions of Core Concepts / Vocabularies.</li><li>Raising awareness on the importance of data and metadata management.</li></ul>The objective of the Core Vocabularies are to capture the fundamental characteristics of a data entity in a context-neutral and syntax-neutral fashion. The aim of these Core Vocabularies is to be simple, reusable and extensible. |
| **CPOV**<br><br>Slides 8 - 11<br><br>**Speaker:** Emiel Dhondt | The issues discussed are issues that were raised during the public review of the latest draft release of CPOV (2.1.1).<br><br>**Issue #30:**<br>The optional property cv:frequency was added to the examples that are included in CPOV as was requested by the community.<br><br>**Issue #31 (& CPSV-AP #123):**<br>Administrative Territorial Unit is a specific code list.<ul><li>It does not align with the definition:<ul><li>"A code from a list ..."</li><li>Which implies multiple code lists can be used.</li></ul></li><li>It does not align with the usage note.<ul><li>"In Europe, this is likely to be the Administrative Territorial Units Named Authority List"</li><li>Which implies multiple code lists can be used.</li></ul></li><li>It does not align with Style Guide rule CMC-R14<ul><li>Enumeration should be used.</li><li>Only when one specific code list is possible.</li></ul></li><li>Within Public Service dct:spatial has range dct:Location.<ul><li>However, the intention and the usage note is the same as in PublicOrganisation.</li></ul></li></ul>**SEMIC Proposition**<br>The proposition contains two parts:<ul><li>In CPOV and CPSV-AP change the range to dct:Location.</li><li>Maintain a Usage Note with the recommendation to use ATU.</li></ul>**Resolution**<br>The group agrees on the incorporation of the proposed resolution. |
| **CPSV-AP**<br><br>Slides 12 - 16 | **Issue #125:**<br>The need to relate cv:Output and cv:Evidence to dcat:Dataset using an optional relation was expressed by the community. Specifically it was |

| | |
|---|---|
| **Speaker:** Emiel Dhondt | requested in the context of the Once Only principle where Evidence and Outputs can be required to be related to a Dataset. |
| | Context of previous webinar:<br>● The mandatory subclassing of cv:Evidence as a dcat:Dataset was removed<br>● A relation similar to the requested relation between cpsv:PublicService and dcat:Dataset was changed from cv:isDescribedAt to dct:isPartOf. |
| | **SEMIC Proposition**<br>Have an optional,dct:isPartOf, towards dcat:Dataset from:<br>● cv:Evidence<br>● cv:Output<br>The relation is optional and therefore should not affect implementers who do not relate their Evidence and Outputs to a dataset. |
| | **Resolution**<br>The group agrees on the incorporation of the proposed resolution. |
| | **Discussion**<br>AG believes it is a good idea to connect both classes to a Dataset. However, he believes it would be good to exclude certain output types from being related to a Dataset. For example, a certificate as an Output. BVN replies that output can be anything and does not have to be a piece of data per se, it can also point to a physical object like an identity card. |
| | ED asks whether a usage note to indicate the distinction between a piece of data or a physical object is desired. AG replies that he cannot answer the question at the time as he is insufficiently prepared. |
| | **Issue [#128](#):**<br>The request from the community was threefold for this issue:<br>1. How to use cv:Evidence and cv:Output when describing a cpsv:PublicService?<br>2. Update Language definition to allow both usages of CPSV-AP for the description and execution of Public Services because it does not allow the usage of CPSV-AP in both high-level use cases described below. Therefore, it will be rephrased.<br>3. Update the relatedDocumentation usage note to allow the "execution" use case, making it more generic. |
| | **Examples**<br>CPSV-AP can be used for two high-level use cases:<br>A. Using CPSV-AP to describe a Public Service as needed for a Catalogue of Service to describe what is available and how it functions.<br>    a. In this case Evidence is not commonly used as it points to a particular object. |

b. It would be more suitable to use the generic concepts of cv:EvidenceType such as the generic concept of a driver's licence.

B. Using CPSV-AP during the execution of a public service, here CPSV-AP is used in support of the public service. In this case more detail is possible as it is an instance of CPSV-AP.

a. In this case cv:Evidence is used and can point to a particular object, for example John Doe's driver's licence.

**Context**
1. As described in issue #95: CPSV-AP can be used to describe a Public Service or describe the execution of a public service.
   a. For describing Public Services it is recommended to use cv:EvidenceType rather than cv:Evidence which is more suitable for the execution of Public Services.
2. Language: The language(s) in which the Evidence must be provided.
   a. This definition is catered to "describing" while the definition should allow "execution".
3. Related documentation: a particular template for an administrative document, an application or a guide on formatting the Evidence.
   a. This usage note is catered to "describing" while it should focus on "execution".

**SEMIC Proposition**
1. Explicitly add the "execution" use case to the introduction of the specification.
   a. Reference the use case in cv:Evidence & cv:Output's usage notes.
2. Update Language definition to allow both usages.
   a. The language(s) the piece of Evidence is in.
3. Update the related documentation usage note.
   a. Add the phrase: "For instance, an energy audit report provides more context to the evidence of a home energy efficiency score." Here the audit report is the related documentation while the energy score is the evidence.

**Resolution**
The group agrees on the incorporation of the proposed resolution.

**Discussion**
MA indicates that they mostly use it for use case A (the first example).

| CV Related Assets | **ReSpec** |
|---|---|
| [Slides 17 - 22](#) | From the public review of the Core Vocabularies onwards, all SEMIC specifications will be published using ReSpec styling. |
| **Speaker:** Emiel Dhondt | **Examples** |
| | In the ReSpec HTML page of each specification examples are provided to facilitate the use of the SEMIC specifications in Turtle and JSON-LD. |
| | The examples are linked to: |
| | <ul><li>a [converter](#) that allows to perform inline transformation of the RDF into JSON-LD, Turtle, RDF/XML and Trig formats.<ul><li>With inline transformation, users no longer need to upload any files and can edit directly inline inside the browser.</li></ul></li><li>a [validator](#) that reuses the SHACL shapes associated with the specification by performing a strict validation (expected values, cardinalities, etc).</li><li>The ITB validation service is currently being tested. The current platform will be phased out if possible. The goal will be to move the converter and the validator to the ITB to provide these services within the Interoperable Europe ecosystem.</li></ul> |
| | SS mentions that they would like to use the ITB Testbed to check instance data of a future cccev-ap.de as was done with dcat-ap.de. CS, the team leader of ITB, replies to SS that setting up a new validator can be done with very little effort. By sending an email to [DIGIT-ITB@ec.europa.eu](mailto:DIGIT-ITB@ec.europa.eu) the process can be started. |
| | **Consolidated Core Vocabularies** |
| | A HTML page will be published containing all m8g terms. |
| | <ul><li>It will contain all classes, properties and relations defined under the data.europa.eu/m8g/ name space</li><li>It will contain a consolidated diagram combining all Core Vocabularies.</li></ul> |
| | SEMIC's aim is to facilitate the reuse of m8g classes, properties and relations. By increasing findability. With this page, a user can find all the assets in one overview. This way reuse can be facilitated if the exact location of a certain asset is unknown. |
| | **An updated mapping** |
| | <ul><li>The mapping has been updated to the latest release of the Core Vocabularies.</li><li>The mapping has been updated to v26 of schema.org.</li><li>The mapping now also includes CPSV-AP.</li></ul> |
| | Feedback on whether community members reuse such a mapping is requested. |
| | RA mentions that schema.org is quite messy and quality is not always of the standard they expect. Therefore, they try to avoid using it. BVN |

| | |
|---|---|
| | replies that that is why the mapping is a separate document and not directly integrated.<br><br>MA adds that tying the CVs to schema.org should really help mapping and integrating all kinds of models, both European and national. He adds that colleagues of his use it in projects for products and services. Therefore, schema.org presents a nice way to map local models to international models.<br><br>PBA indicates that mappings to other vocabularies should always be provided if possible, including schema.org.<br><br>BVN adds that SEMIC would be interested in reuse of such mappings as otherwise it would only be for 'academic' purposes since they would not be used in practice. |
| **Blog post**<br><br>Slides 23 - 33<br><br>**Speaker:** Emidio Stani | This part of the webinar moves forward from semantic interoperability to technical operability. In the SEMIC Style Guide, we start from a conceptual model and derive different types of artefacts such as ontologies, SHACL shapes, JSON-LD contexts for implementing APIs, etc. Other types of artefacts could also be derived such as an XML schema, which is the topic for today.<br><br>The participants are pointed to the blog post on Joinup. The idea of the article is to go from a conceptual model to a physical one, in this case with an XML Schema.<br><br>Following the introduction the article goes into the challenges that are introduced by entering in an XML Schema. The XML Schema has certain restrictions that have to be taken into account when concepts are to be described in an XML Schema when coming from RDF.<br><br>Next are considerations when designing an XML Schema to make it flexible. As with conceptual models, the schema will be adapted over time, therefore flexibility is an important characteristic.<br><br>Lastly, there are different approaches to move between RDF and an XML Schema.<br><br>The focus during the webinar is on certain open questions for those that are interested in dealing with XML Schema, especially in the context of the Core Vocabularies.<br><br>**Need for metadata**<br>The first question for the community is on the minimum metadata needed for an XML schema aside from the versioning.<br><br>Like any other conceptual model you would like to define metadata. Indicating the version of the physical model is one of them, but there can be many more. |

Currently SEMIC does not provide an XML schema. A question is raised on what the minimum metadata needed for an XML schema aside from versioning are. Additional feedback on this question can be provided under the dedicated Style Guide issue.

**How to define data types**
When creating an XML Schema, aside from the concepts of the types that are reflected by the conceptual model, such as a person, data types have to be defined. Some data types are intrinsic to XML Schema, but others are from RDF such as a literal. If this is to be translated into an XML Schema it could be done by using a simple type with a restriction on a string.

Another way is to use the W3C recommendation SAWSDL. The recommendation comes in different steps. SEMIC recommends the XML attribute that refers to any type, simple or complex, to the URI that identifies the concept in the conceptual model. It is possible to use one or multiple URIs. In the case of a literal pointing to the URI of rdfs:Literal would be the solution.

The question is raised whether it is important to bring a mapping like the Literal expression above within the XML schema and whether there are Member States using SAWSDL. Additional feedback on this question can be provided under the dedicated Style Guide issue.

An example can be found on the Finnish data portal where a conceptual model can be exported in an XML Schema.

MA mentions that in the Netherlands CPSV(-AP) is not used directly, but mostly as a blueprint. Therefore, the need for an XSD does not really exist.

**Reusing existing XML schemas vs re-implementing**
The CVs and Applications Profiles rely on existing specifications, such as DCTERMS and FOAF. In the case of DCTERMS an XML Schema is available, however for FOAF this is not the case. Therefore, FOAF has to be recreated in XML Schema. This may create problems of authoritativeness and different implementers might use different approaches to recreating these FOAF concepts.

The question of which approach should be considered when integrating external namespaces that do not have an authoritative XML Schema is raised by SEMIC. Additional feedback on this question can be provided under the dedicated Style Guide issue.

SS mentions that in Germany for the SDG implementation, an XSD serialisation already exists, which they reuse. PF wonders if they used CVs in their implementation and are creating their own XSDs for this. SS replies they use XSD files provided by the Commission for the CVs.

**How the XML Schema could be structured**
The idea is to divide the schema into different blocks which would allow different levels of reuse. CV Simple Types, for example a language string, would be included in a CV Element. A CV Element would in turn be included in a CV Complex Type. Lastly, these different CV Complex Types could be included in various different CVs, such as Core Person XML Schema or Core Location XML Schema. This modularity allows different levels of reuse.

SEMIC raises the question to the community whether this is a good approach for structuring the CVs in an XML Schema. Additional feedback on this question can be provided under the dedicated Style Guide issue.

**Extension Mechanism**
In XML, once you define the concept of a person and add a given name and a family name, then the only way to extend this type would be to extend the type. However, another approach would be to include an intrinsic concept in each type, which would be an extension element. In reality the extension element embeds an xsd:any. This element allows adding any type of element within it.

Another approach could start at the same point, but the person type now includes a PersonAugmentationPoint. This element is defined as an abstract.

SEMIC raises the question whether providing an extension mechanism for each defined type is useful, and if so, what kind of approach should be taken. Additional feedback on this question can be provided under the dedicated Style Guide issue.

SS uses a restriction approach. They use a big construction set and then restrict it by deleting elements. BVN wonders whether they then assume that this big set covers all their needs in terms of basics. SS is convinced that in the public administration domain they do not need the open world paradigm and that everything is known in the legislation texts. Therefore anything can be built from a big set of building blocks.

**Multiple Inheritance and Multiple Instances**
Moving from a conceptual model into RDF makes it possible to have multiple inheritance in the case of a class being a subclass of more than one other class. Additionally, an instance can be a type of different kinds of classes. In XML only single inheritance exists and concepts can only be a single type.

One approach to implement this is by means of groups. At first a person could be defined by a group with a sequence of properties such as first and last name. Next, the Complex Type person is defined by such a group of elements. In essence the properties coming from a

group are encapsulated in the Complex Type.

SEMIC raises the question whether this is too complex and whether such a need exists. However, currently in SEMIC there is no case of multiple inheritance, therefore such a need does not exist in SEMIC. Additional feedback on this question can be provided under the dedicated Style Guide issue.

RA mentions that they do not support multiple inheritance. Their tool contains ontologies, which do not support multiple inheritance. She recognises however that such a use case might arise in the future. However, as of now it is not supported because of implementation difficulties. For example when a host would be a subclass of a legal entity it would inherit the attributes of the legal entity. However, if attributes of a person would need to be added to the host they would be added manually, without creating a subclass relation from host to person. In that sense a host can have attributes from both classes but with only single inheritance.

**Naming and Design Rules to be followed**
In general there are different types of approaches to deal with limitations posed by the use of XML Schema. Rules are necessary to define all sorts of elements.

Different standards exist such as NIEM and UBL. NIEM is designed to be cross-domain with a specialisation for each domain. UBL applies more to a business context, for example invoicing. Both standards define their own rules, however, they have some rules in common. For example all element declarations must be global. Other rules are different such as in the case of extending where NIEM prefers an augmentation point compared to UBL providing extension elements.

SEMIC raises the question on what kind of approach could be followed up, aside from the common rules that could already be adopted. Additional feedback on this question can be provided under the dedicated Style Guide issue.

RA mentions that if rules are well established they should be adhered to, but she cannot mention any specific rules. However, if there are common, international naming conventions it would be good to adhere to them.

**XML Validation Service**
Validation mechanisms exist to validate an instance of an XSD in XML against the XSD. The question is whether there is a need from the community for a validation service for XML Schemas for the CVs, such as is provided for CPSV-AP and DCAT-AP validators on ITB.

| | Currently SEMIC does not provide an XML schema validation service for the Core Vocabularies such as it is done for ITB CPSV-AP and DCAT-AP SCHACL validators. The question is raised whether such a service is needed. Additional feedback on this question can be provided under the dedicated [Style Guide issue](#).<br><br>CS confirms that XSD validation and Schematron are available in the ITB platform.<br><br>The Finnish tool (RA) generates XSD but validation is currently not supported. ES wonders if they have any measurements for the reuse of their artefacts. RA replies that that currently is not the case, but there are discussions on the inclusion of enumerations and/or rootnodes. The need for XML is currently being evaluated within the development of their new tool. RA adds that she believes users rely on API serialisation and are less interested in XML. ES adds that the challenges that were mentioned for XML Schema could also be applicable for other serialisations. |
|---|---|
| **Wrap-up & next steps**<br><br>[Slides 34 - 38](#)<br><br>**Speaker:** Bert Van Nuffelen, Pavlina Fragkou | BVN presents the next steps for the Core Vocabularies:<br><ul><li>The mural will remain open for 24h, after which the content is moved to the [Style Guide GitHub Issue page](#).</li><li>All feedback on the Core Vocabulary XML Schemas is to be provided by 15 May 2024.</li><li>Close all issues posted during the Core Vocabulary public review.</li><li>Release Core Vocabulary assets as SEMIC Recommendations.</li></ul><br>PF thanks the participants for their participation and encourages them to provide feedback on GitHub, the mural and participate in the [survey on usage of SEMIC assets](#). |