

D05.01 20230202 Meeting Minutes: Webinar on Style Guide for semantic data specifications

Project:	SEMIC	Meeting Date/Time:	19/04/2023 10:00 - 12:00
Meeting Coordinator:	Eugeniu Costetchi, Bert Van Nuffelen	Issue Date:	24/04/2023

Meeting Agenda

1. Welcome
2. SEMIC overview & webinar introduction
3. Tour of the revised Style Guide
4. Changelog
5. Plenary discussion
6. What's next?

Meeting Slides

[LINK](#)

Participants


Name	Initials	Organisation
William Verbeeck	WV	SEMIC Team
Bert Van Nuffelen	BVN	SEMIC Team
Jitse De Cock	JDC	SEMIC Team
Anastasia Sofou	AS	SEMIC Team
Pavlina Fragkou	PF	SEMIC Team
Emidio Stani	ES	SEMIC Team
Makx Dekkers	MD	SEMIC Team

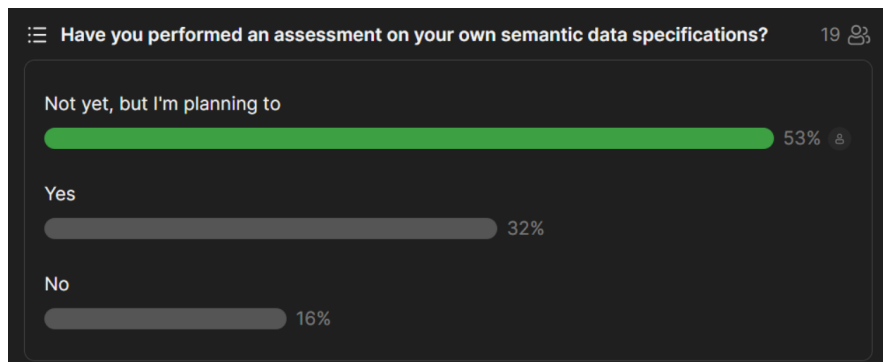
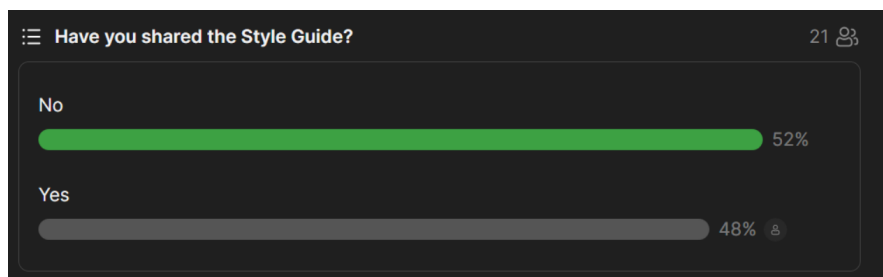
Eugeniu Costetchi	EC	SEMIC Team
CSongor Nyulas	CN	SEMIC Team
Hans Overbeek	HO	Kennis- en exploitatiecentrum Officiële OverheidsPublicaties (KOOP)
Agata Majchrowska	AM	European Commission
Alejandro Villar	AV	Open Geospatial Consortium
Alexandros Vassiliades	AV	Aristotle University of Thessaloniki
Andreea Pasare	AP	QIAGEN
Björn Hagström	BH	Hagström Consulting AB
Costas Simantos	CS	Sigma Cubed
Olli Hurskainen	OH	Digital and Population Data Services Agency Finland
Riita Alkula	RIA	Digital and Population Data Services Agency Finland
Ivan Penava	IP	Central State Office for Development of Digital Society Hungary
Jasna Knezevic	JK	Bundesverwaltungsamt
Kees Trautwein	KT	Logius
Ludger Rinsche	LR]Init[
Lyubomir Blagoev	LB	USW Ltd
Marie Muller	MM	Unknown
Maya Borges	MB	Digitaliseringsstyrelsen
Miha Jesenko	MJ	Ministry of Public Administration Slovenia
Norman Caleja	NC	MITA
Jim Yang	JY	Norwegian Digitalisation Agency
Oskari Nenonen	ON	Finnish Digital Agency
Peter Bruhn Andersen	PBA	The Danish Agency for Digitisation
Rob Atkinson	RA	OGC

Robert Czarny	RC	DG EAC
Stratos Kontopoulos	SK	Centre for Research & Technology Hellas (CERTH)
Svein Olsen	SO	Unknown
Thomas Francart	TF	Sparna
Vassilis Tzouvatas	VT	National Technical University of Athens
Honza Förster	HF	Cognizone
Mihai Paunescu	MP	Publications Office of the European Union

Summary	
Welcome	Slides 1 - 4
Semic overview & webinar introduction	Slides 5 - 24
Tour of the revised Style Guide	Slide 25
Changelog	Slides 26 - 37
Plenary discussion	Slides 42 - 51
What's next?	Slides 52 - 55

Full meeting minutes

Welcome	<p>PF welcomed the participants, presented the agenda and explained the objective of the webinar. BVN explained the context and summarised the previous webinar.</p>						
SEMIC overview & webinar introduction	<p>BVN elaborated on the road to the Style Guide and the motivation of creating it.</p> <p>First poll:</p>  <table border="1"> <caption>Poll Results: Have you read the Style Guide?</caption> <thead> <tr> <th>Response</th> <th>Percentage</th> </tr> </thead> <tbody> <tr> <td>Yes</td> <td>79%</td> </tr> <tr> <td>No</td> <td>21%</td> </tr> </tbody> </table>	Response	Percentage	Yes	79%	No	21%
Response	Percentage						
Yes	79%						
No	21%						



Motivation of the Style Guide:

BVN explained the intended audience of the Style Guide.

Next, BVN elaborated on the two semantic asset types and the technicalities of the Style Guide.

Question from RA on how to retrofit UML models for all the common vocabularies we want to incorporate - e.g PROV, GeoSPARQL, RDF-Datacube, SKOS etc - and how to have normative JSON-LD contexts for each model.

RIA asked why one should start from UML and not from OWL (and SHACL) and then generate a visualisation from those in UML format?

RA also mentioned that OWL has the advantage of a canonical URI which can be dereferenced to find alternative forms, such as UML. It's harder to do the other way around unless you define a canonical way to bind namespaces to UML and ways to publish combinable UML models (which flavour of XMI?).

LR replied that generally speaking, it is easier to start with the representation that is the easiest to understand.

RA agreed but also mentioned that from experience he has found the precise understanding is better in OWL than UML. Is UML better for a general understanding of scope?

TF added to the conversation that practically speaking UML is easier to explain the model to other stakeholders than OWL. What is questionable in the approach in the derivation of both OWL & SHACL from the UML diagram.

RA answered that if UML data models are stored in OWL (not the RDF version of UML, MOF, but actual equivalent data models) then we can have our cake and eat it too. But then it needs a limited profile of UML and specific interpretation rules.

VT stated that the question is not why there is UML in the workflow but why UML is the single source of truth instead of OWL/RDF.

TF replied that the OWL is SSoT for the ontology, SHACL is SSoT for the application profile. HTML documentation is derived from both. Complement them with UML diagrams. Not have UML as SSoT for both.

HF replied that he thinks the logic of not using OWL as SSOT is because of the rule SC-R2: The OWL should be very light-weight on logic and constraints so it cannot be SSoT for APs. On top of that, he mentioned that he thinks the aim is that the UML can be used for both: extracting the broader semantics into OWL and specific constraints into SHACL.

LR added that tooling is missing the RDF/OWL and SHACL generation and that they have built something similar: <https://github.com/GovDataOfficial/plantuml-to-ontology>
It is based on PlantUML, so no nice what you see is what you get UML editor.

TF answered to LR that they are doing SHACL to PlantUML diagrams in SHACL Play (<https://shacl-play.sparna.fr/play/draw>) and that he was thinking that they could try doing it the other way around.

RIA provided a link to the Finnish approach: <https://github.com/VRK-YTI>.

RA added that there is another factor here around the semantic grounding for concepts - in government applications this is perhaps bound up in object registration processes - what is the model used to identify individual resources? These identifiers are what is directly referenced in data integration, so the concepts need to be exposed.

EC explained the reasoning behind the choice for UML:
On one side we need domain experts and business people to understand and read it. At the same time we need the developers to understand the technical part of it, as well as the machine must understand the specification. Unfortunately the languages spoken by the business/domain experts, developers and machines are different. In the Style Guide the separation of concerns, explains the concerns for the different stakeholders. Please keep in mind we are not talking about UML diagrams, UML is a language for conceptual models.

AM added to the conversation that the conceptual model is a pretty blurry concept by the means by which it is described (to recall the discussion of the more or less clear distinction between the conceptual model and the visual representation or visualisation of information).

Secondly, from the perspective of the methodology of digital literacy, one may probably expect a little more extended set of examples that use general diagrams or pretty simple techniques of presenting the information in a way that one does on presentations or slides (as they were already provided during the SEMIC group meetings).

The final remark is: why can't one use another repository than GitHub to present the data? Perhaps the European Data Hub set of tools already offers a similar solution that is widely accessible).

Another remark from RA was whether the UML and the OWL+SHACL diagrams are isomorphic. If they are, then they can be translated. It's the actual diagram layout UML adds, and this can be retrofitted to any starting point.

A final question from TF on this section, related to visual diagrams: the Style Guide does not recommend anything related to "instance diagrams" yet from experience they are important to convey the semantics of the model to business and developer stakeholders. What is your opinion on this?

	Instance diagrams are out of scope for this iteration.
Tour of the revised Style Guide	<p>EC gave a walkthrough of the Style Guide.</p> <p>TF commented that what he found missing in the Style Guide is a recommendation to use “example/instance diagrams” which are much easier to understand by business and developer stakeholders. Out of scope of this iteration.</p> <p>EC gave an explanation of the reuse section in the Style Guide and in general the Style Guide as a whole.</p> <p>On top of that, people can edit the Style Guide via the Github ecosystem, using a pull request. This happens automatically when people click on the “edit” button in the Style Guide itself.</p> <p>EC continued on the changes based on the public review.</p> <p>BVN explained the reason for using UML, same reasoning as EC shared during the webinar: On one side we need domain experts and business people to understand and read it. At the same time we need the developers to understand the technical part of it, as well as the machine must understand the specification. Unfortunately the languages spoken by the business/domain experts, developers and machines are different. In the Style Guide the separation of concerns, explains the concerns for the different stakeholders. Please keep in mind we are not talking about UML diagrams, UML is a language for conceptual models.</p>
Changelog	EC explained the different elements in the changelog .
Plenary discussion	<p>RC mentioned that usually they start from something that exists. He agreed to have a sort of visualisation of the model, not only for the reviewers but also for the business people. He added to consider a sort of path that could help colleagues generate UML diagrams from the data assets they already have.</p> <p>TF stated that the process of going from the conceptual diagram to OWL and SHACL is what they do most of the time. He asked about instances and example diagrams and if they were omitted on purpose?</p>

BVN replied that this is correct, including instance/example diagrams would lead us too far. He added that we see it more as a next step/future work.

EC added that for drawing instances we have been using object oriented diagrams, where you can instantiate the classes. However we have not come up with a very specific set of rules of how to implement this. EC agreed that this is future work.

SO added that they document in AsciiDoc which can then be converted to html or word document, a conceptual model tool should be able to export AsciiDoc.

AM stated that in many cases the object-oriented modelling is prior to the "translation" of the desired model into the technical part of developing the software.

EC added that based on the European Interoperability Framework (EIF), we want to focus primarily on the semantic interoperability in the Style Guide, we are not yet focussing on the technical interoperability but we will have to in the future.

RIA gave the comment that there should be a statement that UML is not mandatory and that the usage of URIs instead of IRIs should be explained too.

BVN replied that many of the rules are UML agnostic.

SO added that he would like to swap the order: Ontology (RDFS-Plus, lightweight) | Reasoning Ontology (OWL2, Complete) | Data Shape (SHACL, Permissive) | Documentation (AsciiDoc, Picture, Precise).

EC proposed English to be the lingua franca.

RIA stated that country specific data specifications, which are meant to be used primarily at a national level, might decide to relax on this convention and permit the use of Unicode characters in the Element names, which will result in the generations of IRIs, instead of URIs [iri]. However, this is unnecessary, and strongly discouraged, in the SEMIC context.

(Element names and URIs - paragraph)

MP added that using an extended set of UNICODE might also lead to confusion due to similar graphical presentation of some characters.

CN replied that internationalisation, if required, should be provided through the use of the UML Element Tags to specify labels that will be used for generation of specification documents in languages other than English.

BVN added that for URIs, users and reusers should try to read them through dereferencing.

AM answered that people will not understand guidelines as you wish them to understand them because they interpret the conceptual model in a different way.

According to RA there are multiple sets of resources and tools needed to make this work:

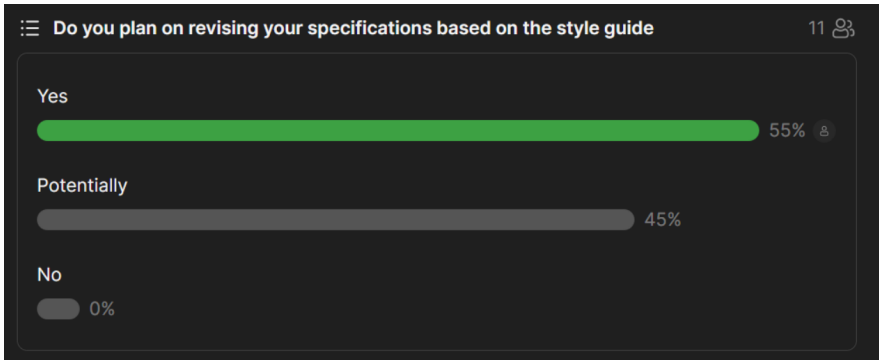
- 1 - Existing ontologies available as UML.
- 2- UML profile enforcement with URI binding and serialisation to OWL + SHACL.
- 3 - A UML access mechanism (repository - e.g. PyPi for Python).

TF replied to the comments of AM that this gives extra motivation to provide "example diagrams" to convey the meaning of the model.

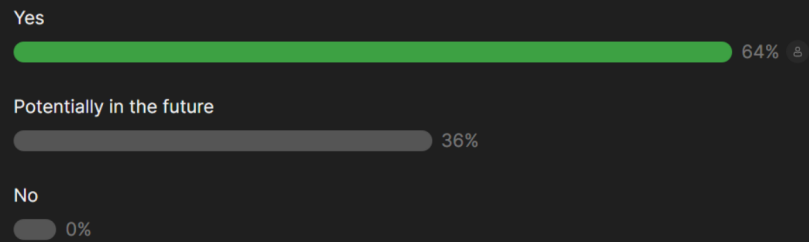
VT commented that the term lightweight ontology is not very good because it implies the others are a heavy weight ontology. Maybe naming it core ontology and adding other layers might be better. EC replied that is a good point, and that we can add another guideline explaining what the heavyweight part of the ontology is.

AM stated that those who understand the GOV-service-related substance matter should be able to deliver the conceptual model that fits for purpose, and then, the "UML formal conceptual model" can be "translated" and applied. Still, one should perhaps take into consideration that the approach is not in practice as formal as one would like to put it for the purpose of the Style Guide.

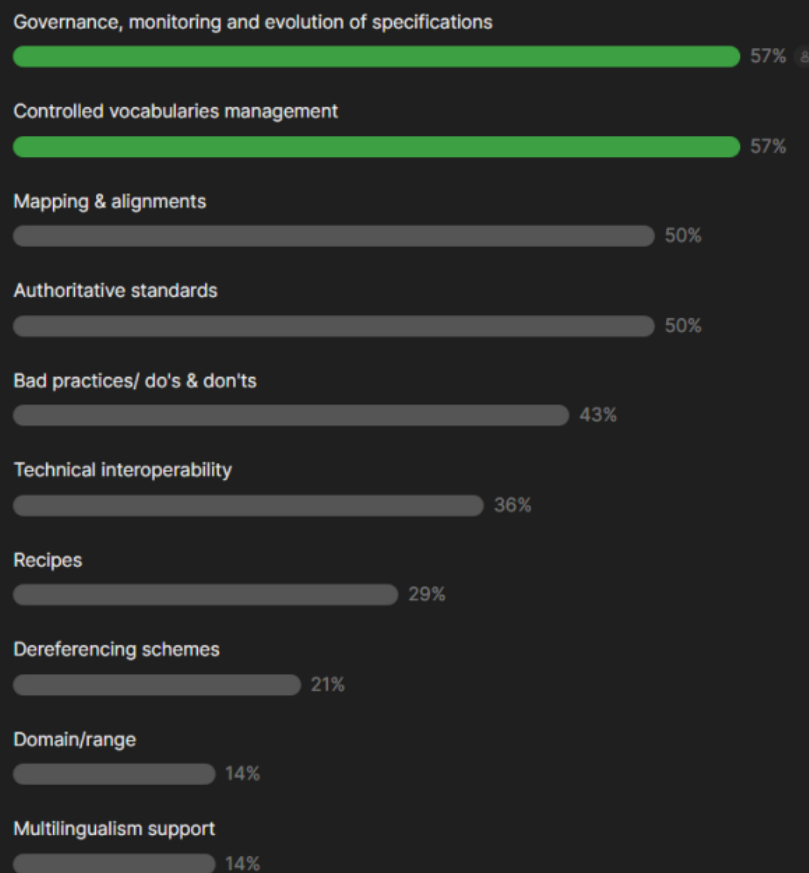
RA stated that the point he wanted to make is that the reality is that data integration means working with real models that come from both existing ontologies and existing UML that won't conform to this

	<p>profile. We cannot throw out the community concerns that lead to these existing specifications and the systems that implement them. So round tripping is really just a way to bind nice UML diagrams to the canonical ontology.</p>								
<p>What's next?</p>	<p>BVN explained the (self-)assessment of specifications.</p> <p>RC commented that he misses implementation guidelines that would show how you should start considering at least two use cases. The first where you have data at hand, the second when you build a model from scratch.</p> <p>JDC replied that this is definitely something we would like to introduce in the future.</p> <p>JDC added that one should also see this assessment as a stress-test for the rules themselves. It could be possible that rules might need modification given your specific context/situation, so please let us know via GitHub if you run into these occurrences.</p> <p>Question from TF: In practice, Application Profile needs to be complemented with two additional artefacts: the specification of data sources to be converted, that can contain only a part of the full data, and the specifications of datasets that are published, when the datasets are only chunks of the large AP. The articulations of these three SHACL variants is interesting to think about.</p> <p>LR stated: Missing Answer = Tools and Processes</p> <p>Results of the second poll:</p>  <table border="1"><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>Yes</td><td>55%</td></tr><tr><td>Potentially</td><td>45%</td></tr><tr><td>No</td><td>0%</td></tr></tbody></table>	Response	Percentage	Yes	55%	Potentially	45%	No	0%
Response	Percentage								
Yes	55%								
Potentially	45%								
No	0%								

☰ Are you planning on adapting the style guide based to your own (national/regional/organisational) needs? 14 👤



☰ Which topics would you like to read blog articles about? 14 👤



Wrap-up

BVN wrapped up the webinar and thanked everyone for their collaboration.

HF answered that it would be nice to see expansion on technical implementation (automation toolchains, validations etc.).

	<p>Final remark of LR: I have not tested it, but Eclipse Papyrus might be a good Open Source tool for creating the UML model:</p>
--	---

<https://www.eclipse.org/papyrus/>