# Meeting Minutes: Webinar dedicated to the Core Vocabularies (SEMIC - A04.01)

Project:	SEMIC - Core Vocabularies	Date and Time:	27/10/2023 13:00 - 15:00 CET
Meeting Type:	Webinar	Location:	Virtual
Coordinators:	Emiel Dhondt	Issue Date:	02/11/2023

Agenda of the webinar			
13:00 - 13:05	Introduction	Slides 1-8	
13:05 - 13:15	Core Public Event	<u>Slides 9-12</u>	
13:15 - 13:25	Core Business	<u>Slides 13-16</u>	
13:25 - 13:45	Core Location	<u>Slides 17-20</u>	
13:45 - 14:05	Success story: Slovakia	Slides TBA	
14:05 - 14:10	CCCEV	<u>Slides 23-28</u>	
14:10 - 14:25	CPSV-AP	<u>Slides 29-33</u>	
14:25 - 14:40	Core Person	<u>Slides 34-36</u>	
14:40 - 14:45	Style Guide	<u>Slides 37-39</u>	
14:45 - 14:50	Wrap-up	<u>Slides 40-44</u>	

Meeting Slides	
LINK	

Participants		
Name	Initials	Organisation
Alessio Nardin	AN	European Commission
Ana Rosa	AR	Spanish Government
Anastasia Sofou	AS	SEMIC Team
Andreea Pasare	AP	SEMIC Team
Angsar Mondorf	AM	European Commission
Annika Strupp	AS	INIT
Antonios Stasis	AS	Ministry of Digital Governance, Greece
Arthur Schiltz	AS	SEMIC Team
Augusto Herrmann	AH	Herrmann
Bart Hanssens	вн	FPS Policy and Support, Belgium
Benny Liund	BL	Swedish Companies Registration Office
Bert Van Nuffelen	BVN	SEMIC Team
Carola Schneider	cs	PD consulting
Davide Bruno	DB	Tuscany government
Dorota Kazanecka	DK	Ministry of Digital Affairs, Poland
Emidio Stani	ES	SEMIC Team
Emiel Dhondt	ED	SEMIC Team
Erik Mossing	EM	Swedish Companies Registration Office
Eugeniu Costetchi	EC	SEMIC Team
Felicitas Loeffler	FL	Ministry of Finance, Thüringer
Frans van der Zande	FvdZ	Ministry of internal affairs, Netherlands
Franz-Josef Gietz	FJG	Eurojust
Giampaolo Sellitto	GS	National Anti-Corruption Authority, Italy
Ida kamienska	IK	Cyfra Poland
Igor Trickovic Rifelj	ITR	Ministry of Public Administration, Slovenia

Participants		
Name	Initials	Organisation
Jakub Klímek	JK	SEMIC Team
Jorge Sousa	JS	The Administrative Modernisation Agency, Portugal
Judy Wartella	JW	National Park Service, USA
Kees Trautwein	KT	Ministry of internal affairs, Netherlands
Kirsi Mikkonen	KM	Centre for Economic Development, Transport and the Environment, Finland
Kurt Stian	KS	Norwegian Digitalisation Agency
Liivi Karpistsenko	LK	Ministry of Economic Affairs and Communications, Estonia
Lucian Alexandru Onisei	LAO	European Commission
Lukas Kisza	LK	Investment and Development, Czech Republic
Makx Dekkers	MD	SEMIC Team
Marco Combetto	MC	JRC
Matthias Palmér	MP	MetaSolutions AB
Max De Wilde	MDW	IMEC
Melinda Cuzner	MC	Agency for Digital Government, Sweden
Mihai Paunescu	MP	Publications Office of the European Commission
Mikael af Hällström	МаН	Finnish Tax Administration
Miroslav Liska	ML	Ministry of Investments, Regional Development and Informatisation, Slovakia
Morten Borrebaek	MB	Norwegian Mapping Authority
Norman Calleja	NC	Maltese Information Technology Agency
Olli Hurskainen	ОН	Centre for Economic Development, Transport and the Environment, Finland
Pavlina Fragkou	PF	SEMIC Team
Per Nordanlind	PN	Swedish Company Registration Office

Participants		
Name	Initials	Organisation
Peter Bruhn Andersen	PBA	Agency for Digitisation, Denmark
Peter Winstanley	PW	Semantic Arts
Ricardo Sanchez Curiel	PSC	Västra Götaland Regional Council
Riitta Alkula	RA	Digital and population services agency, Finland
Robert Czarny	RC	DG EAC
Sander Van Dooren	SVD	Digital Flanders
Sebastian Sklarß	SS	Init, Germany
Sébastien Albouze	SA	Publications Office of the European Commission
Shanice Callus	SC	Government of Malta
Silvia Baines Zugasti	SBZ	Government of Navarre
Stina Avvo	SA	Ministry of Economic Affairs and Communications, Estonia
Tarja Myllymäki	TM	National Land Survey of Finland
Thomas Holmgren	TH	Swedish Companies Registration Office
Torbjörn UII	TU	Swedish Company Registration Office
Tore Helland	TH	Brønnøysund Register Centre
Viktoria Šunderlíková	VS	Ministry of investment, regional development and information, Slovakia
Zia Alborzi	ZA	National Data Service, Luxembourg

### Points discussed and decisions taken

Topic discussed	Outcome	
Issues on Core Public Event Vocabulary		
Issue #25 Status of the event	Approved  • If URIs are reused.	
Issue #7 Post-fact statistics and knowledge	Approved	
Issues on Core Business Vocabulary		
Issue #26 #24 Non-profits and people as legal entities	Approved	
Issue #44 Other Documents related to Legal Entity	Not approved  • Further input is requested via the GitHub page.	
Issues on Core Location Vocabulary		
Issue #36 AdminUnit class	Approved	
Issue #25 AdminUnit class alignment with INSPIRE	Not approved  • Further discussion on the GitHub page is encouraged.	
Issue #37 A structural way to represent address supplements	Not approved  • Further discussion will be held on the GitHub page.	
Issues on Core Criterion and Core Evidence Vocabulary		
Issue #52 Wrong range for endTime and startTime	Approved	
Issue #47 The usage note for "Evidence - is provided by"	Approved	
Issue #46 & CPSV-AP issue #111 #91 Usage of DCAT: Dataset in CCCEV and CPSV-AP	Approved  Option 1 is approved, option 2 is not approved.	
Issue #51 Qualified relations - AND/OR/NOT combinations of Criterions	Not approved  • Further investigation by SEMIC is necessary.	

	Further discussion on the GitHub page is encouraged.
Issues on CPSV-AP	
Issue #103 #84 #76 Usage of DCAT: Dataset in CPSV-AP	<ul> <li>Approved</li> <li>Align definition with authoritative source.</li> <li>Option 1 is approved.</li> </ul>
Issue #120 Controlled vocabulary for dct:type in LifeEvent and BusinessEvent	Approved  • According to the working group's suggestion of moving the type property to the BusinessEvent and LifeEvent class.
Issue #115 Usage of time:TemporalEntity	Approved  • Further discussion on the GitHub page is encouraged.
Issues on Core Person Vocabulary	
Issue #23 Naming in Person	<ul> <li>Unresolved</li> <li>If no resolution is found by the working group, the issue will be closed.</li> </ul>

#### **Full Meeting Minutes**

### Welcome & Introduction

#### Welcoming word and structure of the webinar

#### Slides 1-8

The subject of this webinar is a discussion on the solutions, minor modifications and further developments of the Core Vocabularies.

### **Speaker:** Pavlina Fragkou

#### Objectives:

- 1. Resolve open issues
- 2. Gather input for improvements
- 3. Share experience

#### Agenda:

- 1. Core Public Event
- 2. Core Business
- 3. Core Location
- 4. Success story: Slovakia
- 5. CCCEV
- 6. CPSV-AP
- 7. Core Person
- 8. Style Guide

The objective of the Core Vocabularies is to provide simplified, reusable and extensible solutions to capture the fundamental characteristics of data.

#### Core Public Event

#### 1) Issue #25

#### Slides 9-12

The issue is about the eventStatus. In the existing diagram it cannot be modelled. An event status with a codelist was requested.

### **Speaker:** Emiel Dhondt

#### **SEMIC Proposition**

SEMIC proposes the addition of an eventStatus property using the RFC 5545 code list, with an appropriate usage note.

#### Resolvement

The group agrees on the incorporation of this proposition if URIs are reused.

#### Discussion

MP asks whether this codelist would include URIs or literals. If it were published by the Publications Office then it would be under the form of URIs.

The question whether it would be published as linked open data is raised by JK.SEMIC confirms this is the case.

SS raises the question if former events need their own status. MP counters this, as querying should be enough to find relevant

information as it was in the past. SS agrees with this.

#### 2) Issue #7

In this issue the community requested properties on interesting statistics after an event has taken place.

#### **SEMIC Proposition**

The expectedNumberOfParticipants property is suggested to be added and the instances of hasParticipation can be counted to calculate attendance.

#### Resolvement

The group agrees on the incorporation of this proposition.

#### **Discussion**

The question is raised why the cardinalities on the properties are all open. However, SEMIC addresses this by the fact that this is true for all Core Vocabularies; it allows customisation when end-users model their own version.

#### **Core Business**

#### **Slides 13-16**

### **Speaker:** Emiel Dhondt

#### 1) Issue #26 #24

This is an ongoing issue concerning LegalEntity not covering all legal entities and has been mentioned several times. Examples are sole traders, self employed people, non-profits, etc.

#### **SEMIC Proposition**

SEMIC proposes to add a legalFormType property using the GLEIF Entity Legal Form code list and republish this as Linked Data through the Publications Office. This code list contains specific legal entities per country and GLEIF can be requested to add missing legal entities.

#### Resolvement

The Working Group agrees on the incorporation of this proposition.

#### Discussion

TU mentions concerns whether GLEIF will remove certain legal entities over time. However the SEMIC team reassures that as long as a legal entity exists, they will not be removed. If a legal entity is not in the list, it can be requested to be incorporated in the next release. GLEIF remains the owner of the data and they are the point of contact for amendments for the list.

#### 2) Issue #44

The AccountingDocument class was added in the last webinar to align with implementing regulation on <u>HVD</u>, however it is currently empty.

#### **SEMIC Proposition**

SEMIC proposes to add a name property and a type property with a legislation-based codelist including financial statements, non-financial statements, management reports and annual financial reports.

#### Resolvement

This issue is currently unresolved, as SEMIC requests input on the relevant GitHub page.

#### **Discussion**

The Working Group suggests incorporating some identifier for the document and an issuing date.

#### **Core Location**

#### **Slides 17-20**

### **Speaker:** Emiel Dhondt

#### 1) Issue #25 #36

A first issue is that some uncertainty exists about what codelist should be used for AdminUnit.code. This is an error because the usage note was related to the wrong property.

A second issue with AdminUnit, which diverges from the INSPIRE guidelines.

#### **SEMIC Proposition**

SEMIC proposes to correct the usage note and add a name property to AdminUnit to align with INSPIRE.

#### Resolvement

The group agrees on the incorporation of this proposition.

#### **Discussion**

JK is interested in whether NUTS data can be used. SEMIC confirms that it can.

MP informed the SEMIC team that RDFs label may not be suitable as it is language dependent. Further investigation is necessary on this topic from SEMIC's side.

AR is interested in the difference between ATU, NUTS & LAU. SEMIC clarifies that they cover different administrative levels.

SEMIC encourages further discussion of this topic under issue tracker #36.

#### 2) Issue #37

In this issue a request was made for a more structural way to represent address supplements.

#### **SEMIC Proposition**

SEMIC proposes to align with INSPIRE by including a class LocatorDesignator which includes a name and a type, using an

INSPIRE code list.

#### Resolvement

The Working Group wishes to further discuss this issue under issue #37.

#### Discussion

SS mentions a need for free text input, which under the current proposition is unavailable. SEMIC responds that LocatorDesingator has a free text field in the property LocatorDesignator.designator.

JK raises the concern that it won't be possible to have multiple designators, multiple names and be able to model their relation in the current proposition. MP suggests to add a blank note to resolve the issue. SEMIC notes that the LocatorDesignator designator and type should be sufficient, to which JK agrees.

MP suggests to look into vCard as it might be a better solution to point to the right designator, this will be further discussed on GitHub under issue #37.

#### Success story: Slovakia

Slides 22-

**Speaker:** Miroslav Liska

#### **Content of presentation**

This presentation covers the Slovakian knowledge graph, with a focus on CCCEV-AP-SK which was created to better disseminate what the vocabulary and semantics are. Lastly there are a few things about ongoing and planned projects using this vocabulary.

#### The Knowledge Graph

They have developed a centralised knowledge graph that consolidates various ontologies, including national and international ones. These ontologies are often derived from standard models, and some examples include RDFs and RDF. When you click on a specific ontology, it displays all its components, including classes and properties presented in triples.

To ensure clarity, the decision was made to use Slovakian ontologies for describing reference data in reference registries, while other ontologies could remain in scope internationally. The challenge they faced was that when creating a data schema, they tended to reuse international ontologies. However, this approach left developers and analysts confused, as they often lost track of which URIs they were using for which classes and properties. For the sake of clarity and consistency, they opted to describe reference data using national ontologies.

The category part of the knowledge graph includes codelists, taxonomies, thesauri, etc. This also includes national taxonomies and the corresponding URIs for the national namespaces. However, a lot of international taxonomies and namespaces are included, like continent,

country, frequency, etc.

The last part of the knowledge graph is the instance data. This is mainly used as clarification for developers. Here they can find how they should implement linked data in their information systems.

Everything within the knowledge graph is linked. The KnowledgeType class is linked to concrete EvidenceTypes linked to this class. On top of this, the data can be instantly requested in JSON-LD or RDF format by a simple push of a button.

The hardest part with regards to this knowledge graph is explaining what the ontologies and taxonomies exactly are.

#### CCCEV-AP-SK

CCCEV-AP-SK is an Application Profile for evidence regulations. Additionally, it shows developers how they should implement this evidence in their information systems.

The structure is relatively simple. It contains the Evidence class, the EvidenceType class, the Agent class with the subclasses Person, Subject, and PublicOrganisation.

They created a lot of regulations that can be found on the methodologies portal of the Slovakian government concerning implementation of such evidence.

#### **Future project**

In the future, they plan to implement a lot more of these regulations. The implementation can be concerning, for example, ID cards, birth certificates, driving licences, etc.

#### **Discussion**

SS wonders about the relationship between the ABR BRegDCAT-AP initiative and the AP from CCCEV in Liska's use case. He questions if the latter might be a better choice for describing evidence and registering requesters in the Member States. SEMIC responds that they will look into this.

MP wants to understand the practical implementation of the examples, particularly regarding the allocation of responsibilities to the involved agencies.

ML offers insight by explaining that they are in the process of establishing an organisation responsible for enforcing data standardisation. The agencies play an active role in defining the precise expressions of these standards through collaborative working group meetings, involving stakeholders, programmers, developers, and others. They carefully examine existing data, identify gaps, and study

the data model's relationship with the evidence, all adhering to the only once principle. Their goal is to match and provide as much information as possible.

MP raises concerns about the sustainability of this approach, particularly for agencies that already have established systems.

ML reassures him by mentioning that the interoperability act mandates each Member State to appoint a national interoperability coordinator, which they have done even before the act will come into effect. This implementation and transition, ML believes, contributes to the sustainability of their efforts.

SS asks about the usage of Europass and Publication Offices Codelists in education evidence, specifically mentioning elements like country codes and the EQR.

ML responds, stating that they have not utilised these codelists yet. Their current focus is on codelists related to open data, such as continents and frequencies. However, he mentions their intention to incorporate Europass and codelists of the Publications Office in the future.

BVN inquires about the effectiveness of the Core Vocabularies and whether they facilitated or hindered the process.

ML explains that they employed Core Vocabularies primarily for content that didn't fall under reference data. This decision was made to avoid confusion among users who might encounter both international and national URIs. However, ML adds that they have plans to map national vocabularies to Core Vocabularies in the future. This mapping will be carried out in a way that aligns with the requirements of data consumers.

#### CCCEV

#### **Slides 23-28**

### **Speaker:** Emiel Dhondt

#### 1) Issue #52

The URIs from W3C are used but they have the wrong range for endTime and startTime.

#### **SEMIC Proposition**

SEMIC proposes to correct this mistake.

#### Resolvement

The group agrees on the incorporation of this proposition.

#### 2) Issue #47

This issue addresses the phrasing of the usage note of isPorividedBy which should in fact be the other way around. Currently it says 'Agents requesting' which should be 'Agents issuing'.

#### SEMIC Proposition

SEMIC proposes to correct the usage note and add supplementary clarification on the properties' usage.

#### Resolvement

The group agrees on the incorporation of this proposition.

#### 3) Issue #46 & CPSV-AP issue #111 #91

The issue has two requests. The first is that CCCEV:Evidence should not be a DCAT:Dataset. The second is that CV:Output should be a DCAT:Dataset.

#### **SEMIC Proposition**

SEMIC proposes three options:

- Represent CCCEV:Evidence no longer as a DCAT:Dataset
- 2) Represent CV:Output as a DCAT:Dataset
- 3) Leave as is

#### Resolvement

Option 1 is well received by the group and will be regarded as the agreed upon proposition. Option 2, on the other hand, is poorly received by the group.

#### **Discussion**

MP asks about the nature of an evidence output in CPSV-AP.

ED responds that, within CPSV-AP, an output can vary. Some individuals may consider an ID card as a valid output, while others may view it as the entry in a registry that references the ID card as the public service's output. The specific definition relies on the implementation in use.

MP observes that such an output is typically treated as an instance and not categorised as a dataset. Due to this perspective, option 2 is deemed unsuitable.

#### 4) Issue #51

In this issue the request for a possibility to use AND, OR or NOT statements allowing combinations of multiple Criterions.

#### **SEMIC Proposition**

SEMIC request input from the community on this issue under the tracker #51.

#### Resolvement

Once further investigated, this issue can be resolved.

#### **CPSV-AP**

#### 1) Issue #103 #84 #76

This issue addresses the fact that the relation with

#### Slides 29-33

**Speaker:** Emiel Dhondt

PublicService and their Dataset is unclear through isDescribedAt. It also addresses the definition of DCAT:Dataset, which is not aligned with the authoritative source (W3C).

#### **SEMIC Proposition**

SEMIC proposes to adapt the definition to align with the authoritative source.

On top of that SEMIC proposes to either change isDescribedAt to:

- 1) dct:isPartOf, since it is the inverse of dct:hasPart
- 2) Rdfs:members to align with how Socrata describes a row.

#### Resolvement

The group agrees on the incorporation of this proposition under option 1 (dct:isPartOf).

#### **Discussion**

Some participants need time to process and will further discuss on GitHub.

MP questions the absence of a connection to DataService.

BVN explains that the dataset's primary objective is to serve as a representation of a catalogue of public services, which is why there is no association with DataService.

#### 2) Issue #120

This issue addresses the request to indicate the business event type and the life event type according to the Controlled Vocabulary Business Events and Life Events, respectively.

#### **SEMIC Proposition**

SEMIC proposes to add a recommendation to use Business Events and Life Events code lists in the usage note.

#### Resolvement

The group agrees on the incorporation of this proposition according to the suggestion of JK to move the type to the BusinessEvent and LifeEvent class and include a corresponding usage note to each one.

#### Discussion

JK suggests relocating the type attribute to specific classes to allow distinct usage notes for separate codelists.

ED indicates that this approach is under consideration, and if there are no objections, it will be implemented. There is no objection from the group and JK's proposal is approved.

## 3) Issue #115

This issue tackles the request to clarify the usage of availabilityRestriction.

#### **SEMIC Proposition**

SEMIC proposes to add:

- 1) a different usage note to increase clarification
- 2) a property "reason" in the TemporalEntity class which would indicate the reason for the deviation of being open or closed with a list of values.

#### Resolvement

This issue is open for further clarification, but regarded as agreed upon.

#### Core Person

#### Slides 34-36

#### Speaker: Emiel Dhondt

#### 1) Issue #23

This issue requests input from the Working Group on how to model names, in particular country specific naming conventions such as double names, maiden names, middle names, etc.

#### **SEMIC Proposition**

SEMIC proposes that if no resolution is found the issue will be closed and will remain unresolved due to lack of momentum.

#### Style guide

#### Slides 37-39

Dhondt

#### Speaker: Emiel

#### What?

Last year SEMIC developed a Style Guide. It provides rules about naming conventions, which syntax to be used, how to manage artefacts and how to organise the creation of a Core Vocabulary.

#### For who?

The Style Guide was designed for editors and reusers of the SEMIC Core Vocabularies and Application Profiles.

#### Some common rules

The Style Guide has some common rules:

- reuse existing concepts as much as possible
- the choice in handling the lexicalisation of concepts shall be clearly defined and consistently implemented
- all UML Element names are case-sensitive and shall follow the CamelCase convention.

#### Style guide compliance

We reviewed our Core Vocabularies to see if we are aligned with the Style Guide. The first issue is that there is no consistent version management when it comes to deprecating. We will make a clear deprecation policy and procedure.

The second is that there is a lack of lexicalisation rules. We will create rules to formalise this.

The third is the reuse of assets. There is no formalised and

	standardised policy or procedure, this will also be tackled in future iterations.
	Self assessment We have created a self assessment system to improve the quality of the solutions. For this, we are creating blog posts about the Style Guide to raise issues, request for contribution and raise comments. It is ongoing work and the aim is to apply this to have better compliance with the specifications.
Wrap-up Slides 40-44	Survey on the adoption of SEMIC specification Working group participants are requested to participate in a survey that will analyse the adoption of SEMIC specifications.
Speaker: Emiel Dhondt	Next steps The next steps in the maintenance of the Core Vocabularies are the following:  - Incorporate the decisions of this webinar  - Act on all actions required to align with the Style Guide  - Release the minor releases of the Core Vocabularies  - Update and release a new consolidated Core Vocabularies  Diagram