Meeting Minutes: Fourth Working Group webinar on the revision on GeoDCAT-AP (SEMIC - A04.02)

Project:	SEMIC	Date and Time:	14/05/2024 10:00 - 12:00
Meeting Type:	Webinar	Location:	Virtual
Coordinators:	Bert Van Nuffelen Jakub Klímek	Issue Date:	05/06/2024

Agenda of the webinar			
10:00 - 10:10	Introduction	<u>Slides 1 - 3</u>	
10:10 - 10:15	DCAT-AP ecosystem	<u>Slides 4 - 8</u>	
10:15 - 10:55	GeoDCAT-AP Issues	<u>Slides 9 - 20</u>	
10:55 - 11:35	Codelists	<u>Slides 21 - 39</u>	
11:35 - 11:50	XSLT & High-Value Datasets	<u>Slides 40 - 49</u>	
11:50 - 11:55	GoeDCAT-AP 3.0.0: Overview of changes	<u>Slides 50 - 51</u>	
11:55 - 12:00	Next steps & closing	<u>Slides 52 - 56</u>	

Meeting Slides	
LINK	

Participants		
Name	Initials	Organisation
Age Sild	AS	Estonian Land Board
Anastasia Sofou	AS	SEMIC Team
Antoine Jacques	AJ	GIM, Belgium

Participants		
Name	Initials	Organisation
Arthur Schiltz	AS	SEMIC Team
Aurelija Sturaite	AS	Construction Sector Development Agency, Lithuania
Bert Van Nuffelen	BVN	SEMIC Team
Daniel Gomez	DG	Cartographic and Geological Institute of Catalonia, Spain
Pavlina Fragkou	PF	DIGIT
Emiel Dhondt	ED	SEMIC Team
Erik Obersteiner	EB	Environment Agency Austria
Fabio Vinci	FV	Epsilon Italia
Florian Barthelemy	FB	SEMIC Team
Francois Prunayre	FP	Cerfrance Haute-Loire, France
Gery Nicolay	GN	Spacebel, Belgium
Hanna Horppila	нн	National Land Survey of Finland
Ine de Visser	ldV	Geonovum, Netherlands
Inga Andriuskeviciute	IA	The National Center for Remote Sensing and Geoinformatics, Lithuania
Jakub Klímek	JK	SEMIC Team
Jitka Faugnerová	JF	Czech Environmental Information Agency
Joachim Nielandt	JN	Digital Flanders, Belgium
Joeri Robbrecht	JR	DG ENV
Jordi Escriu	JE	JRC
Josema Alonso	JA	SEMIC Team
Juliana Karciauskiene	JK	The National Center for Remote Sensing and Geoinformatics, Lithuania
Lena Hallin-Pihlatie	LHP	Finnish Land Survey
Lenka Rejentova	LR	Czech Environmental Information Agency

Participants		
Name	Initials	Organisation
Mayte Toscano	МТ	OGC
Paloma Abad	PA	National Geographic Information Center, Spain
Pascal Derycke	PD	Sciensano, Belgium
Vendula Dastychová	VD	Czech Environmental Information Agency
Veronika Kusova	VK	State Administration of Land Surveying and Cadastre, Czech Republic
Pawel Soczewski	PS	GISPartner, Poland
Štěpán Kafka	SK	Charles University, Czechia

Points discussed and decisions taken

Topic discussed	Outcome	
GeoDCAT-AP		
Issue <u>#81</u> Distributor agent role	Approved : keep the property distributor on Dataset and clarify its meaning.	
Issue <u>#113</u> Revise usage of licences and AcessRights	Approved.	
Issue <u>#82</u> Relax rights max cardinality	Approved : lift cardinality both in GeoDCAT-AP and DCAT-AP.	
Issue <u>#100</u> Relation of spatial resolution on Dataset, Distribution and Data Service	Approved : cardinality of spatial resolution on Dataset to be kept unbounded, wording of 'data is managed' adapted to 'original spatial resolution' or similar.	
Codelists	Approved : feedback is requested over the public review.	

Full Meeting Minutes

	PF welcomes the participants and introduces the topics of the webinar:
Introduction	Introduction
	GeoDCAT-AP Issues
<u>Slides 1 - 3</u>	Codelists

Speaker: Pavlina Fragkou	HVD, GeoDCAT-AP XSLTNext steps
Tagkou	The purpose of this webinar is to close the series on the revision of GeoDCAT-AP.
The GeoDCAT-AP ecosystem <u>Slides 4 - 10</u>	GeoDCAT-AP ecosystem GeoDCAT-AP is primarily a DCAT-AP profile for geospatial datasets. However, the major other inputs are the ISO and INSPIRE standards. Recently, the OGC Standards Working Group (SWG) started working on the GeoDCAT standard.
Speaker: Jakub Klímek, Bert Van Nuffelen	GeoDCAT-AP timeline The revision started with an introductory webinar in February, followed by three working group webinars including this one. In June the editor's draft of GeoDCAT-AP will be released and a public review will ensue. The public review will end in September followed by the official release of GeoDCAT-AP 3.0.0.
	Issue overview A total of 59 issues are open in the GeoDCAT-AP GitHub repository. The first draft of GeoDCAT-AP, which will be published soon, already contains the resolution to 34 of those issues.
	 The remaining issues are classified as follows: 4 issues have a resolution provided, but are not yet implemented; 4x to be discussed today; 2x closed, because no feedback was received; 9x implementation evidence; 3x postponed beyond version 3.0.0; 3x editorial.
GeoDCAT-AP issues Slides 9 - 20	Issue #81 Some of the GeoDCAT-AP agent roles, such as distributor, may not make sense for all DCAT entities, that is Dataset, Dataset Series, Data Service, Distribution.
Speaker: Jakub Klímek	In Flanders, distributor on ISO "dataset" scope code is mapped to geodcatap:distributor of dcat:Distribution instead of dcat:Dataset.
	 SEMIC Proposition The SEMIC proposition is to either: Clarify meaning of the distributor role on dcat:Dataset, or Move the distributor from dcat:Dataset to dcat:Distribution, i.e. Remove distributor from dcat:Dataset, and Map distributor on Datasets always to corresponding dcat:Distribution.

a this tion.
N
the
an f D
is Irce S lic ne f
t. ng
n to
e

Discussion JF is interested in the practical side of publishing a licence. She wonders what would happen to licences that are in national languages that are not English.
BVN replies that the Publications Office has a centralised list which contains the most prominent open licences that are available. Some Member States have reached out to include their national licences. The Publication Office can accommodate national licences in this codelist. However, as this content is of a legal nature it is likely that this text is published on a national portal and ideally the published licence by the Publications Office can point to that licence. The licence does not have to be translated in all official EU languages. Additionally, when published by OP translation services may be offered, but this has to be discussed in accordance with OP.
JF requests clarification on the use of a URL pointing to a published licence in a national data portal. BVN mentions that the register is important in the context of the HVDs because it has to be an open licence, which becomes hard to compare in the case of national licences. However, mapping it to a licence from the EU NAL is allowed.
Issue #82 In INSPIRE metadata, rights are expressed as multiple textual statements. However, in DCAT-AP & GeoDCAT-AP, rights (on Catalogue and Distribution) currently have cardinality 01.
SEMIC Proposition The SEMIC proposition is to either relax rights max cardinality in both GeoDCAT-AP and DCAT-AP, or merge multiple rights statements into one in the mapping from INSPIRE to GeoDCAT-AP.
Resolution The majority of the votes are in favour of lifting the cardinality of dct:rights in both GeoDCAT-AP and DCAT-AP.
Discussion No discussion on this issue was held.
Issue #100 This issue is regarding specifying the definition of spatial resolution on the different levels of Dataset, Distribution, Data Service and Dataset Series, and their relationship.
 SEMIC Proposition The SEMIC proposition to this issue is fourfold: for Distribution, spatial resolution represents the spatial resolution of the described file with a cardinality of 0 to 1; for Data Service, spatial resolution describes the capabilities of the data service with an unbounded cardinality [0n], i.e. in

 which spatial resolutions it can serve data; 3. for Dataset, spatial resolution describes the spatial resolution the data is managed in in the dataset with cardinality 0 to1, i.e. regardless of how it is distributed using distributions; 4. for Dataset Series, it is unclear what spatial resolution of a dataset series means.
Resolution The majority of the votes are in favour of the proposition, however the cardinality of spatial resolution on the class Dataset must remain unbounded (0*). Additionally, the wording 'data is managed in' will be changed to 'source spatial resolution', 'original spatial resolution' or similar.
Discussion JN asks what it means to manage data in a dataset in a certain spatial resolution?
JK replies that a dataset is managed by a publisher and this dataset is kept in a certain spatial resolution. However, in a distribution it can be distributed in different spatial resolutions that could be converted computationally. The spatial resolution of a dataset should be as intended by the data manager or data owner.
JN mentions that the wording of 'manage the data' was unclear to him. He argues that perhaps it could be clearer to specify 'source resolution' or 'original resolution'. JK responds that this suggestion will be included.
AA mentions they have datasets that have different spatial resolutions, for example rural areas have lower resolutions, while cities and towns are more detailed. Therefore, having a maximum cardinality on spatial resolution on Dataset is not ideal and an unbounded cardinality would make more sense. PA, LHP, and JE and DG agree with this.
 Resolved issues The following issues were resolved as no negative feedback was received on the proposed resolution: CRS support in GeoJSON (#6) Required / Recommended properties of supporting classes (#109)
 Closed issues The following issues were closed as no feedback was received: 1. Support 1-to-many mappings for responsible party roles (#39) 2. Relationships between GeoDCAT-AP and DCTERMS agent roles (#57)

Codelists	In this section a common interpretation for the sentence below will be attempted to be established.	
<u>Slides 21 - 39</u>		
Speaker: Bert Van	"The property MUST use as range values codes from {codelist} which are transferred from one specification to another."	
Nuffelen		
	An example of the codelist is the following:	
	ns:codelist1 a skos:ConceptScheme.	
	ns:codelist1 skos:prefLabel "Example Codelist1"@en.	
	codelist1:x2 a skos:Concept.	
	<pre>codelist1:x2 skos:prefLabel "Code x2"@en.</pre>	
	<pre>codelist1:x2 skos:inScheme ns:codelist1.</pre>	
	<pre>codelist1:x2 skos:topConceptOf ns:codelist1.</pre>	
	Most of the code lists are based on a well structured SKOS controlled vocabulary that includes URIs and metadata about such codes or values.	
	Below three examples are given of the use of codelists. Depending on how the reader interprets MUST, they may see different examples as valid, or not.	
	Codelist Qualifier	
	Example	
	The property MUST use as range values codes from EU vocabularies Data theme.	
	Example 1 Example 2 Example 3 _:d dcat:theme nal:AGRI. :d dcat:theme nal:AGRI.	
	_:d dcat:theme naliAGMI:d dcat:theme naliAGMI. _:d dcat:theme inspire:au:d dcat:theme inspire:au.	
	$\checkmark \times \checkmark \checkmark \checkmark \times$	
	In example one the value is from the Data theme codelist, in example 2 the value is from another, INSPIRE codelist, and in the last example, both values are used.	
	The same thought exercise is done but with "the property MAY use as range values codes from EU vocabularies Data theme" instead of MUST as was the case in the previous example.	

Codelist Qualifier		
The property MAY use as range values co	des from EU vocabularies Data the	ne.
_:d dcat:theme nal:AGRI.	xample 2 :d dcatitheme inspirelau.	Example 3 _:d dcat:theme nal:AGRI. _:d dcat:theme inspire:au.
$\checkmark X$	$\checkmark \times$	$\checkmark \times$
BVN presents his own interp regards the first option as tru true and the second as 'uncl	ue. In the MAY case he	e regards the first as
Discussion JF argues that MUST and M she had experience with EU the meaning of MUST.	-	e .
PS has the same interpretat	ion as BVN in the case	e of MUST.
JR mentions that 'shall' indic recommendation, and 'may' permitted. Lastly, 'can' is use	is used to indicate that	it something is
JE agrees with JR as this is framework.	the approach used in	the INSPIRE
JK points the participants to the meaning for MAY/MUST		ains information on
SHACL shapes: as-is The shapes are used for val	idation services and c	onformance testing.
The following pattern is used	d:	

<pre>:Codelist1Restriction a sh:NodeShape ; rdfs:comment "Codelist1 restriction" ; sh:property [sh:hasValue ns:codelist1; sh:minCount 1 ; sh:modeKind sh:IRI ; sh:path skos:inScheme] .</pre>	<pre>:PropertyShape a sh:NodeShape ; sh:property [sh:node :CodelistRestriction ; sh:nodeKind sh:IRI ; sh:path dct:subject ; sh:severity sh:Violation] ; sh:targetClass dcat:Dataset. In case of MANDATORY the severity is a 'Violation',</pre>
	in all other cases the severity is a 'Warning'.
A value has to be within the codelist re that codelist. This is how the pattern w value the error is a violation, in all othe warning.	orks. In case of a mandatory
Multiple systems The question is what happens when m dataset that contains the value from th harvesting the dataset might have a sl	e EU NAL codelist. Each system
Imagine that Portal A harvests accordi codelist. However, Portal B might required codelist. In Portal A the validation will B it will not be validated correctly, and correctly.	ire a value from the INSPIRE be successful, however, in Portal
In case that the INSPIRE value is used opposite is true.	d in the dataset, then the
In the case that both values are used, in the case of MUST.	then it is invalid on both portals
The interpretations of mandatory code	lists can be as follows:
 Interpretation 1: The value space is clo Cardinality and value space co Restricts the freedom of compareinforcing the cardinalities or f codelist values (e.g. to a single) 	nstraints are independent. atible (sub)profiles in terms of further restricting the possible
 Interpretation 2: At least 1 value from t Cardinality and value space co Compatible (sub)profiles may f constraints. 	nstraint are made dependent.

 Cannot be used in case the property is optional (cardinality 0 means optional).
Interpretation 1 is conformant to the way literal value spaces are expressed and is most natural for programming languages and software systems. Interoperation 2 on the other hand is of interest in cases where there is a need for aggregation at the level of properties in the specification.
The validation for each interpretation is also different.
In Interpretation 1 the existing SHACL shapes can be used as-is and the validation results are simple and direct. An error is received when the value is outside of the value space, in that sense violations drive the feedback
In Interpretation 2 the existing SHACL shapes cannot be used as-is. It would require a filtering process as an inherent part of the processing and conformance building. Such a filtering process is non-trivial to standardise and would be imposed on all implementations. Additionally, as the value space is not closed, poor usages are harder to detect. It is, however, possible to create matching SHACL shapes for Interpretation 2.
The last point is that of multiple systems. Both interpretations will have an impact on other related specifications. If Interpretation 1 is used in one of the specifications it will overwrite Interpretation 2 as it is more restrictive. For example, if GeoDCAT-AP used Interpretation 1 and DCAT-AP uses Interpretation 2, then Interpretation 1 will be enforced on DCAT-AP as some datasets will be expressed, and harvested, in both GeoDCAT-AP and DCAT-AP.
The conclusion here is that coordination is required.
SEMIC Proposition The SEMIC proposition is fourfold to accommodate the different types of interpretations:
 A. The property MUST use as range values codes from {codelist} o Interpretation 1: The value space is closed under the codelist and validation results in violations. o All (sub)profiles must avoid conflicts by creating subproperties.
 B. The property MUST have at least one value from {codelist} Interpretation 2: The value space is minimally constrained and validation results in warnings. All (sub)profiles must adopt this interpretation in case they want to restrict the value space.

 C. The property IS RECOMMENDED to use as range values codes from {codelist} Interpretation 1: The value space is closed under the codelist, but other values are tolerated and validation results give warnings. D. The property MAY use as range values codes from {codelist} Interpretation 1: The value space is closed under the
codelist, but other values are accepted. In this case no validation is required. Resolution The proposal will be implemented and can be further assessed by the community during the review.
Discussion IdV thinks it is a difficult discussion and she argues it should be viewed from the level of the Application Profile. For example, a generic AP such as DCAT-AP may not use specific codelists compared to a specialised AP. Therefore, she believes that for example the national level should not be too restrictive, while the domain level should.
 XLST This part of the webinar focuses on the impact of the review of the XSLT. Before the update, in GeoDCAT-AP 2, the property dct:subject was used on which gmd:topicCategory was mapped as can be seen in the example below. After the update the same property of INSPIRE will be mapped to the new GeoDCAT-AP topicCategory property. The input remains the same, only the output is mapped on a different property. Therefore, the effects of the update remain minimal.
<pre><!--- GeoDCAT-AP 2.0.0 output--> <rdf:description> </rdf:description> </pre> <p< th=""></p<>

HVD On June 4th a webinar on HVD reporting will be organised by DG CNECT.
On April 25th a meeting was organised by the Action 2.5 subgroup. As a result two options were introduced to identify HVD in the INSPIRE metadata. Either INSPIRE metadata should be extended to indicate what is a HVD, or datasets should be automatically mapped from the INSPIRE themes to HVD Categories for those datasets that fall under the HVD IR.
The decision on which option to adopt will be held on Friday May 17th. Based on this information it will become clear what needs to happen with the XSLT from SEMIC's side.
Additionally, the HVD Categories NAL code list will be finer-grained and not be limited to the 6 top level categories, but also include second level categories. The release in June will include the IRIs and their English labels. However, in September the translations in the European languages will be added.
XSLT limitations for HVD One of the limitations of the XLST in terms of HVD is with regards to licensing information. DCAT-AP HVD assumes that the licensing information is given in the form of an IRI from the EU licence NAL or that there is at least a mapping towards those licence NAL.
The example below contains some indication towards a creative commons licence. After the transformation in the XSLT it is a Data Service with rights pointing to a rights statement with a description as the IRI as text.
<pre>cgmd:resourceConstraints> cgmd:resourceConstraints> cgmd:resourceConstraints> cgmd:resourceConstraints> cgmd:resourceConstraints> codeList="http://www.isotc211.org/2005/resources/codeL ist.xaiMW_DestrictionCode" codeList="http://coder=strings"></pre>
There are two problems with this. The first is that the licence is not recognised as an IRI and the relation to CC-BY-4.0 is unknown. The relation to CC-BY-4.0 is required by the HVD Implementing Regulation. The XSLT will not be able to produce a correct representation of each licence in every implementation. Therefore, publishers will have to correct this manually in a mapping or with a licence from the codelist.
A second point is concerning identifiers. In the HVD IR there is a requirement that the metadata and service identifiers are persistent and

dereferenceable. In the INSPIRE metadata there may be an identifier in the form of a UUID, however this is not dereferenceable, and there is no guarantee that this is a persistent identifier. The XSLT cannot make the IRI dereferenceable and persistent, as can be seen in the example below. Each publisher should make the identifier available as a URI and make them persistent.



@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix geodcatap: <http://data.europa.eu/930/> .
[] a dcat:DataService ;
geodcatap:resourceType <http://inspire.ec.europa.eu/metadata-codelist/ResourceType/service> ;
dcterms:identifier "229a081d-5c6b-4181-8410-6f07d9b55437" .

Lastly, in INSPIRE currently the simplified INSPIRE metadata exists. This no longer contains independent metadata records for Data Services. The metadata about the Data Service is available by dereferencing the endpoint URL, which the XSLT cannot do. Therefore, it might become difficult to distinguish bulk downloads from APIs which will need to be done manually for reporting purposes.

The XSLT can be found <u>here</u>. The community is encouraged to provide feedback on the limitation under the issue section of this repository.

Discussion

JR mentions that the XSLT does not have enough logic to make the ISO or INSPIRE metadata completely compliant with the HVD requirements. The XSLT would need smarter logic to guarantee compliance. There are two options, map on a national level to DCAT-AP or use the XSLT and accept the fact that some errors may be included. Therefore, JR warns the audience to not be 100% reliant on the XSLT.

LHP mentions that they still have a lot of service metadata. From the presentation she understood that the metadata should also be tagged with "bulk-download" or "API" or distinguish them with another solution. LHP wonders whether data providers should rush to meet the deadline of June 9th or wait until clear guidelines are in place.

JR answers that by the 9th of June the implementation should be in place. However, reporting should be done by the 9th of February. Until then no actions on compliance will be taken. The meeting with DG CNECT on the 4th of June should also bring additional clarity. He expects DG CNECT to be lenient as fully functional implementations cannot be expected yet.

FP mentions that GeoNetwork also planned to embed the <u>conversion</u> which provides conversion to various DCAT profiles.

GeoDCAT-P 3.0.0: Overview of changes	Over the course of the webinars the following main changes were identified and will be implemented in the new version. In total 6 new subproperties will be added.	
<u>Slides 50 - 51</u> Speaker: Jakub Klímek	 dcterms:conformsTo geodcatap:serviceProtocol geodcatap:referenceSystem dct:subject geodcatap:topicCategory dct:type geodcatap:serviceType geodcatap:resourceType geodcatap:serviceCategory The introduction of DatasetSeries, and the properties that are present on a Dataset are also included in this class. Additionally, clarification of usage notes and cardinalities will be provided. Licence and rights mappins and label mappings are also included in the revision. The changes are relatively small and the specification is still very similar to GeoDCAT-AP 2. The additional effort to comply with 	
Wrap-up & next steps Slides 52 - 56 Speaker: Bert Van Nuffelen, Pavlina Fragkou	DCAT-AP 3.0.0 is therefore expected to be minimal. The intention is to start the public review by the end of June and release an official version of GeoDCAT-AP 3.0 in September. The community is encouraged to engage with GeoDCAT-AP and the XSLT on GitHub. Additionally, the first editor's draft will become available soon. The community, the members of DG ENV and the JRC are thanked for their participation in the process and over the course of the four webinars.	