



DCAT Application Profile for Data Portals in Europe

WORKING GROUP

VIRTUAL MEETING 2

2013-04-17

Meeting Minutes



DCAT Application Profile – Working Group Virtual Meeting 2

Venue	Virtual Meeting	Meeting date	2013-04-17
Author	MDK	Meeting time	16:00 – 18:00
Reviewed by	SG	Issue date	2013-04-18
Status	for review	Version	0.02

ATTENDEES

Name	Abbreviation	Organisation
Leda Bargiotti	LB	Publications Office
Antonio Carneiro	AC	Publications Office
Alessandra Cornero	ACo	Formez PA
Christopher Gutteridge	CG	University of Southampton
Gregor Eibel	GE	Federal Chancellery of Austria
Dominik Klauser	DK	Federal Chancellery of Austria
Kai Koistinen	KK	National Land Survey of Finland
Florian Marienfeld	FIM	Fraunhofer FOKUS
Andrea Perego	AP	European Commission – Joint Research Centre
Emilio Garcia	EG	Telefonica
Thodoris Papadopoulos	TP	Greek Prime Minister's Cabinet , Strategic Planning Bureau
Giovanni Lalle	GL	Senato della Repubblica
Martin Alvarez	Mal	CTIC
Simon Dustowski	SD	Fraunhofer FOKUS
Valerie Brasse	VB	Eurocris
Ioannis Kanellopoulos	IK	European Commission - Joint Research Centre
Bart Rousseau	BR	City of Ghent - E-Strategy
Carlos Iglesias	CI	Independent Consultant
Hans Overbeek	HO	KOOP
Martin Kaltenböck	MK	Semantic Web Company (SWC)
Sander van der Waal	SW	The Open Knowledge Foundation
Thimo Thoeye	TT	City of Ghent - E-Strategy
Fadi Maali	FM	DERI
Phil Archer	PA	W3C
Enric Staromiejski	ES	Everis
Makx Dekkers	MD	Independent metadata expert
Nikos Loutas	NL	PwC EU Services
Stijn Goedertier	SG	PwC EU Services
Michiel De Keyzer	MDK	PwC EU Services

AGENDA

Item	Owner	Subject
1	AC	Welcome and overview
2	MD	<p>Collaboration in the Working Group</p> <ul style="list-style-type: none"> • Public mailing list archive • Issue tracker • Conference call system • Contributor licence agreement (You need to be a member of this project and logged in if you want to be able to read and sign the contributor agreement.)
3	SG	<p>Use Cases</p> <ul style="list-style-type: none"> • EPrints Repository (usecase) • Subscriptions, alerts and recommendations • Systems as data consumers • Multilingual metadata
4	MD	<p>DCAT Application Profile: Draft 2 - issues:</p> <p><i>Scoping issues:</i></p> <ul style="list-style-type: none"> • extension patterns and Use of "mandatory" confusing • Enable syndication of dataset metadata? and URI sets and DCAT-AP <p><i>Classes/Properties/Cardinalities:</i></p> <ul style="list-style-type: none"> • DCAT-AP and Spanish profile (NTI) and proposals to increase interoperability between the two profiles • Use of org:Organization cf. foaf:Organization • Dataset quality rating • Dataset versions • Max cardinality of mandatory/optional properties and Cardinality of spatial and temporal coverage • Contextual metadata and data set provenance • Alternative representations of spatial coverage • Assumption about digital assets? • Dataset: dcat:landingPage • Dataset: dcat:keyword • Dataset: dct:issued and dct:modified • Catalog: dct:spatial • Catalog: dct:license • Catalog: dct:issued and dct:modified <p><i>Editorial issues:</i></p> <ul style="list-style-type: none"> • Section 5 does not define "minimal" and Mandatory versus minimal properties • Description of CERIF • Models for describing datasets or similar information
5	MD	<p>Controlled vocabularies:</p> <ul style="list-style-type: none"> • Requirements for Controlled Vocabularies • Controlled vocabularies used in data portal • Licensing of and alternatives to EuroVoc • NUTS instead of MDR Countries / Places • Use of CERIF standard vocabularies

		<ul style="list-style-type: none"> • Registers operated by OP's MDR and DCAT-AP
6	AC	Wrap-up, actions, next working group meeting (2013-05-08)

DISCUSSION

1. Welcome and overview 2. Collaboration in the Working Group

Discussion

- AC welcomed everybody and thanked the members of the working group for their contributions.
- MD explained the different tools that are used within the WG:
 - [Public mailing list archive](#): all communications sent to the mailing list are archived online and publicly accessible. The mailing list is used for general announcements and discussion and contributions by the Working Group members.
 - [Issue tracker](#): topics that require further discussion will be formally logged as issues in the issue tracker.
 - [Conference call system](#): to access the conference call system, you use the link published on the respective virtual meeting page on Joinup. You need to enter as a guest. Then you can connect by phone by dialling a valid number or insert your telephone number and let the system call you. You will be asked to enter the room number, which is also available on the respective meeting page on Joinup.
 - [ISA Contributor Agreement](#)¹:
All Working Group members are invited to electronically sign the ISA Contributor Agreement on Joinup. Signing the contributor agreement allows the EU to make the specification available under the ISA Open Metadata Licence v1.1².
To sign the contributor agreement you need to be registered on Joinup and request membership to the DCAT Application Profile project on Joinup. Only then you will be able to access and sign it. Membership can be requested on the [DCAT Application Profile webpage](#) on Joinup.
If there are any problems with signing you can contact MDK.
- EG explained that they do not have an official authorisation to sign the Contributor Licence Agreement ([CLA](#)). However, everything they contribute is publically available because it is part of their legal framework. Therefore EG does not find it really necessary to sign the CLA.
 - This will be further discussed offline.

¹ https://joinup.ec.europa.eu/asset/dcat_application_profile/legaldocument/isa-contributor-agreement-v11
(You need to be a member of this project and logged in if you want to be able to read and sign the contributor agreement.)

² <https://joinup.ec.europa.eu/category/licence/isa-open-metadata-licence-v11>

3. Use Cases

Discussion

SG explained that several new requirements and user scenarios were contributed by the working group. These were added to [draft 2](#) of the DCAT Application Profile (DCAT AP):

- BH contributed [his requirements for cross-lingual search](#). This included the possibility to have multilingual descriptions and multilingual datasets, but also having the possibility to have "multilingual links", f.i. for the landing page (in Belgium you can have a landing page in different languages, each having their own URL).
 - MD said that with regards to the multilingual links, it is not the URL that has a language but the information behind the URL. A link can never have a language.
 - PA explained that it is possible to have a language in the hypertext of the page, using hreflang.
 - MD asked PA to provide some more information on hreflang.
 - MD said that he will include some information on hreflang in the specification.
- AM contributed [a requirement](#) explaining that developers wanting to use datasets for building applications, need information on the interoperability and quality of the data.
 - SG asked AM to further refine this use case.
- EF contributed a [user scenario](#) about the cross-portal search for datasets related to immigration, unemployment, and immigration control legislation.
 - SG said that these user scenarios help clarifying the objective of the DCAT AP. Furthermore, they can also help assessing the usefulness of the controlled vocabulary that will be used to indicate the dcat:theme.
- SG explained that the WG members from EuroCRIS also contributed a requirement with regards to [contextual metadata and provenance](#). TT of the Ghent open data portal also confirmed that this was important to him.
 - SG asked both to further refine this requirement.
- AP contributed a requirement for the [cross-border spatial search for datasets](#).
 - SG has formulated these requirements in terms of a user scenario.
 - AP will further refine this scenario.
- WG members from EuroCRIS asked to add a user scenario for additional services such as [subscriptions, alerts and recommendations](#).
 - SG has put this use case in a user scenario.
- CG formulated the requirement for data consumers to be able to report required corrections to datasets. Without this one loses one of the big benefits of open data: crowd-sourced error detection.
 - SG asked CG to further clarify his requirement. Does he require a contact person?
 - CG clarified that a contact point would be needed for data consumers to report errors.
- SG explained that more use cases and user scenarios could still be contributed and invited the members of the WG to share them through the mailing list.

4. DCAT Application Profile: Draft 2 - issues

Discussion

- MD explained that [draft 2](#) of the DCAT Application Profile (DCAT AP) has been released on Joinup. Members of the WG are invited to review and submit any comments.

Scoping

The discussion below refers to the following points:

1. [extension patterns](#) and [Use of "mandatory" confusing](#)
 2. [Enable syndication of dataset metadata?](#) and [URI sets and DCAT-AP](#)
- MD said that the initial intention for the DCAT AP, was to restrict to what is available in [DCAT](#) and only select these elements and properties that are needed. However, some comments were submitted, asking whether we should really restrict to DCAT or if extensions could be added (point 1).
 - Several member of the working group explained that they already did extensions on DCAT for usage in their data portal.
 - PA explained that even if new properties and classes are defined, the specification will still be an Application Profile of DCAT.
 - MD said that if there is a need in the working group to allow extensions, this should be allowed. This will widen the original scope of this work.
 - The working group agrees.
 - CG explained that adding a contact point for a dataset would enable the benefit of crowd-sourced error detection.
 - EG asked whether it would be more useful to have the organisation as a contact for a dataset instead of a contact person.
 - HO agreed.
 - SG mentioned that ADMS also foresees to add a [contact point](#) for an interoperability asset. The EU Open Data Portal also has this information.
 - MD explained the issue "[Enable syndication of dataset metadata?](#)". It was agreed that the Application Profile would not include a mechanism for data exchange.

Classes/Properties/Cardinalities:

The discussion below refers to the following points:

1. [DCAT-AP and Spanish profile \(NTI\)](#) and proposals to increase interoperability between the two profiles
2. [Use of org:Organization cf. foaf:Organization](#)
3. [Dataset quality rating](#)
4. [Dataset versions](#)
5. [Max cardinality of mandatory/optional properties](#) and [Cardinality of spatial and temporal coverage](#)
6. [Contextual metadata and data set provenance](#)
7. [Alternative representations of spatial coverage](#)
8. [Assumption about digital assets?](#)
9. [Dataset: dcat:landingPage](#)

10. [Dataset: dcat:keyword](#)
 11. [Dataset: dct:issued and dct:modified](#)
 12. [Catalog: dct:spatial](#)
 13. [Catalog: dct:license](#)
 14. [Catalog: dct:issued and dct:modified](#)
- MD explained that MC [contributed](#) the Spanish Technical Interoperability Standard for the Reuse of Information Resources (Annex III – Catalogue’s Document And Information Resource Metadata) also known as NTI. A comparison between this profile and draft 1 of the specification of the European DCAT AP has been prepared and was shared with the WG. From the analysis, some issues were raised in the issue tracker (point 1).
 - MD suggested to make the following properties optional instead of mandatory (point 9-12):
 - Dataset – dcat:landingPage
 - Dataset – dcat:keyword
 - Dataset – dct:issued
 - Dataset - dct:modified
 - Catalog – dct:spatial
 - The WG agreed.
 - MD suggested making the following properties mandatory instead of optional (point 13-14).
 - Catalog – dct:license
 - Catalog – dct:issued
 - Catalog – dct-modified
 - BR commented that he is not sure about making Catalog – dct:license mandatory this information is not always available. The availability of this information can for instance depend on the status of publication of the dataset.
 - MD replied that this property is not a property of Dataset but of the Catalog.
 - CG commented that if license information is not available, you can always make the license “no specific license”.
 - The working group agrees on making the properties mandatory.
 - MD suggested to broaden the scope for publisher by using foaf:Agent instead of org:Organization (issue 2).
 - MD explained the issue related to [dataset versioning](#), raised by AP (point 4).
 - PA explained that [ADMS](#) includes a way of versioning assets, which could be referred to for datasets.
 - FM clarified that the DCAT specification currently does not support versioning of datasets.
 - MD suggested waiting for the [DCAT](#) Working Group to come with a proposal with regards to the versioning of datasets and thus not include it in the specification for the DCAT AP.
 - PA said that it is very likely, nearly certain, that DCAT will not include next/prev/last. AP disagreed in not including it in the specification.
 - TT explained that the Ghent University is looking to apply versioning on the triple level. More information can be found via the following link:
http://ruben.verborgh.org/publications/vandersande_ldow_2013/
 - NL said that from his understanding of the use cases, supporting version information is not required.

- SG said that from the experience he has with ADMS, probably not many data portals will have accurate information on the version chain of datasets. Moreover, using versioning information implies that every new version of a dataset is requires a new description of the dataset, so a new metadata record. SG thinks that this would create unnecessary overhead, without serving the primary use case of the application profile.
- TP agreed with SG.
- MD asked AP to write a note on the use case for having versioning information.
- MD said that for the moment, versioning will be left out the DCAT AP.
- MD explained that in the current draft there are no cardinalities. This will be added in the next draft (point 5).
- MD asked TT to provide a use case on the [contextual metadata and data set provenance](#) so it could be included in the next draft (point 6).
 - TT agreed.
- MD explained the issue raised by AP on "[Alternative representations of spatial coverage](#)": In some cases it could be possible to have other values for this property than what is included in the proposed controlled vocabularies, f.i. coordinates. To cover this, AP referred to the [Core Location Vocabulary](#) (point 7).
 - The WG agreed.
- MD explained the issue "[Assumption about digital assets?](#)": Not all datasets are digital and thus, will not have a downloadURL or accessURL. Many datasets of the relevant governmental data sets are not available on the web (point 8).
 - CG explained that he proposed two use cases for this.
 - MD explained that DCAT defines datasets as digital datasets.
 - FM replied that DCAT is not narrowing datasets to digital datasets only.
 - FM suggested adding a note in the specification that says that not all datasets should be available online.
 - SG agreed on not restricting datasets to those directly available on the Web. Some valuable governmental data is not freely available on the Web, yet it still makes sense to have it listed on data portals.
 - FM explained that if a dataset is not available on the web, the description of the distribution could explain how it could be obtained. Furthermore, the distribution could include a property dcat:accessURL to a Web page explaining how to obtain the dataset.
 - FIM said that he thinks it brings added value to only focus on online datasets
 - CG disagreed with FIM
 - MD said that if a distribution has no accessURL and downloadURL it is assumed to not be online. The way on how the dataset distribution in this case could be obtained should be added in the description of the distribution.
 - FIM agreed with MD.
 - SG suggested adding a use case for non-digital data.

Editorial issues

The discussion below refers to the following points:

1. [Section 5 does not define "minimal"](#) and [Mandatory versus minimal](#)

- [properties](#)
2. [Description of CERIF](#)
 3. [Models for describing datasets or similar information](#)
- IP commented that the difference between mandatory, optional and minimal should be further clarified (point 1).
 - Several members from the WG agreed that the way this is currently described in the specification, is confusing.
 - AC commented that especially the difference between mandatory and minimal is confusing.
 - MD explained that the mandatory classes and properties are those that a data provider should provide. However if only the minimal properties is the set of properties that is needed for a receiver to be able to process the data. The minimal properties are a subset of the mandatory properties.
 - NL and SG suggested only keeping mandatory and optional.
 - FM said that having the division is ok but the visualisation of this division in the document should be improved.
 - MD agreed on improving the visualisation.
 - MD said that he will add a description of CERIF to the specification (point 2).
 - HS proposed to also take the work of [Discovery](#) into consideration for section 2.1 (point 3).
 - MD agreed to add this in the specification.

5. Controlled vocabularies

Discussion

The discussion below refers to the following points:

1. [Requirements for Controlled Vocabularies](#)
 2. [Controlled vocabularies used in data portal](#)
 3. [Licensing](#) of and [alternatives](#) to EuroVoc
 4. [NUTS instead of MDR Countries / Places](#)
 5. [Use of CERIF standard vocabularies](#)
 6. [Registers operated by OP's MDR and DCAT-AP](#)
- MD explained that a list of requirements (point 1) will be elaborated for the controlled vocabularies included in the DCAT AP specification. Examples of these requirements are:
 - URIs should be dereferenceable.
 - A persistence policy should exist for the controlled vocabulary.
 - The controlled vocabularies should be multilingual (NL).
 - AC explained that the [licensing](#) policy for [EuroVoc](#) is currently under revision and will most likely be updated so the thesaurus is openly available. More information on this will be available for the [next meeting](#) (point 3).
 - With regards to [alternatives](#) to [EuroVoc](#), AC explained that they do not want to impose EuroVoc as the only controlled vocabulary (point 3).

- MD discouraged looking for alternatives for [EuroVoc](#) , since this would probably be more complex than actually using [EuroVoc](#).
- MD said that a comment was made that [EuroVoc](#) uses different URIs for different languages of a concept.
 - SG replied that this comments is probably not correct.
 - LB said that they will check internally.
- MD asked the members of the WG to contribute the controlled vocabularies they use in their data portal (point 2).
- MD explained the issue "[NUTS instead of MDR Countries / Places](#)" and asked the question to the Publications Office whether the name authority list is granular enough compared to [NUTS](#) (point 5).
 - LB answered that for the purpose of this exercise the name authority list will probably be better suited than [NUTS](#).
- AP explained that [INSPIRE](#) has developed a number of code lists, of which some them could be useful for the WG. AP will share more details on this with the WG.
- MD explained that with regards to the [registers operated by OP's MDR and DCAT-AP](#), any suggestions can be made to the [Publications Office](#) and they will then take this in consideration (point 6).
- AC asked the Working Group members to continue the discussion on the controlled vocabularies online and indicate which controlled vocabularies are used on their data portal:
https://joinup.ec.europa.eu/asset/dcat_application_profile/issue/which-vocabularies-are-used-your-data-portal

6. Wrap-up, actions, next working group meeting (2013-05-08)

Discussion

- AC concluded the meeting:
 - The meeting minutes of this meeting will be available on: <https://joinup.ec.europa.eu/node/64618>
 - Next meeting: 2013-05-08. The agenda and connection details will be made available on: <https://joinup.ec.europa.eu/node/64744>

ACTIONS & DECISIONS

Action	Owner	Date
EG and MD to discuss signing the CLA.	MD, EG	2013-05-08
MD to improve the visualisation in the specification of the division between mandatory, optional and minimal properties.	MD	2013-05-08
AP to write a note on the use case for having versioning information in the DCAT AP.	AP	2013-05-08

MD to add cardinalities to the properties.	MD	2013-05-08
TT to provide a use case on contextual metadata and data set provenance .	TT	2013-05-08
MD to add a description of CERIF to the specification.	MD	2013-05-08
MD to include the work of Discovery (a metadata ecology for education and research. Discovery is advocating the adoption of open metadata for the furtherance of scholarship and innovation) as related work in section 2 of the specification.	MD	2013-05-08
AM to further refine the use case on having information on interoperability and quality of datasets.	AM	2013-05-08
AP to further elaborate the use case on cross-border search for datasets.	AP	2013-05-08
MD to elaborate a list of requirements for the controlled vocabularies included in the DCAT AP specification.	MD	2013-05-08
PA to provide more information on hreflang to MD	PA	2013-05-07
MD to include information on hreflang in the specification	MD	2013-05-08
Members of the WG to contribute the controlled vocabularies they use in their data portal.	WG members	2013-05-08
LB to check whether in EuroVoc different URIs are used for different languages of a concept.	LB	2013-05-08
AP to share information on the code lists developed in the context of INSPIRE with the WG.	AP	2013-05-08

Decision

It is possible to define extensions on DCAT for the Application Profile.

The following properties will be made optional instead of mandatory:

- Dataset – dcat:landingPage
- Dataset – dcat:keyword
- Dataset – dct:issued
- Dataset - dct:modified
- Catalog – dct:spatial

The following properties will be made mandatory instead of optional:

- Catalog – dct:license
- Catalog – dct:issued
- Catalog – dct:modified

Version information will currently be left out the DCAT AP.

Refer to the [Core Location Vocabulary](#) to have other values for spatial coverage than the values included in the proposed controlled vocabularies.