DCAT Application Profile for Data Portals in Europe

WORKING GROUP

VIRTUAL MEETING 3

2013-05-10

Meeting Minutes

DCAT Application Profile – Working Group Virtual Meeting 2			
Venue	[Virtual Meeting](#)	**Meeting date**	2013-05-10
Author	MDK	**Meeting time**	16:00 – 18:00
Reviewed by	SG	**Issue date**	2013-05-15
Status	for review	**Version**	0.02

ATTENDEES

Name	Abbreviation	Organisation
Leda Bargiotti	LB	Publications Office
Antonio Carneiro	AC	Publications Office
Kai Koistinen	KK	National Land Survey of Finland
Florian Marienfeld	FlM	Fraunhofer FOKUS
Andrea Perego	AP	European Commission – Joint Research Centre
Emilio Garcia	EG	Telefonica
Giovanni Lalle	GL	Senato della Repubblica
Martin Alvarez	Mal	CTIC
Valerie Brasse	VB	Eurocris
Carlos Iglesias	CI	Independent Consultant
Hans Overbeek	HO	KOOP
Phil Archer	PA	W3C
Enric Staromiejski	ES	Everis
José Manuel Canabal	JMC	Red.es
Enrico Francesconi	EF	Italian National Research Council
Alejandra Sánchez Magento	ASM	National Centre of Geographic Information
Bert Van Nuffelen	BVN	TenForce
Szabolcs Szecaks	SS	EC – DG DIGIT
Makx Dekkers	MD	Independent metadata expert
Stijn Goedertier	SG	PwC EU Services
Michiel De Keyzer	MDK	PwC EU Services

AGENDA

Item	Owner	Subject
1	AC	Welcome and overview
2	AC	Acceptance of [meeting minutes](#) of previous WG meeting
3	MD	Collaboration in the Working Group • [Public mailing list](#) archive • [Issue tracker](#)

8	AC	- 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.)
4	MD	Application Profile - issues *Editorial* 1.4 Structure - Requirements for Controlled Vocabularies (Comment 14303) - Typo: schemas.org *Introduction (section 1)* - Document the use of other vocabularies and/or specifications *Terminology (section 5)* - Mandatory, recommended and optional redux - Excluded versus Optional classes *Classes and properties (section 6 & 7)* - Dataset versions - Managing optional properties for foaf:Agent - Walk through cardinality of properties (including issue on identifiers) - Datasource fees - Mandatory properties for Distribution too limited ? *Accessibility and multilingual aspects (section 10)* - Multilingual Aspects *Deployment (section 11)* - URI sets and DCAT-AP - Enable syndication of dataset metadata? - Clarification about the expected operational environment for DCAT-AP *Licensing* - Various licenses per dataset is problematic (i.e. one per distribution) - NoLicenseDefined
5	MD	Controlled vocabularies (section 8): - Mapping Austrian, German and Spanish controlled vocabularies for theme with EuroVoc - Use of CERIF standard vocabularies - Requirements for Controlled Vocabularies - Investigate alternatives to Eurovoc - Which vocabularies are used on your data portal?
6	SG	Use cases (section 3)
7	MD	Public Review period
8	AC	Wrap-up, actions, next working group meeting (2013-06-12)

1. Welcome and overview
2. Acceptance of meeting minutes of previous WG meeting
3. Collaboration in the Working Group

Discussion

- The [meeting minutes of the 2nd WG Virtual Meeting](#) were accepted.
- Everyone is invited to sign the Collaborator Licence Agreement (CLA) on Joinup. This will allow to publish the outcome of the working group under an open licence:
  https://joinup.ec.europa.eu/node/63578
  - If there are any technical problems with signing the CLA, please contact Michiel De Keyzer.
  - If there are any organisational problems to sign the CLA, please contact Makx Dekkers, Stijn Goedertier or Michiel De Keyzer.

4. Application Profile - issues

Discussion

Editorial
The discussion below refers to the following points:
1. [1.4 Structure](#)
2. [Requirements for Controlled Vocabularies (Comment 14303)](#)
3. [Typo: schemas.org](#)

- MD said that the structure of section 1.4 will be updated following the suggestions included in the issue (point 1)
- MD will update the title of section 8.2 from "Proposed vocabularies" to "Vocabularies to be used" (point 2).
- MD will correct the typo in the section about schema.org ("schemas.org").

Introduction (section 1)
The discussion below refers to the following points:
1. [Document the use of other vocabularies and/or specifications](#)

- MD explained that the DCAT Application Profile also reuses properties from vocabularies other than DCAT (point 1). This is currently not represented in the text. MD will mention the used vocabularies in the specification.

Terminology (section 5)
The discussion below refers to the following points:
1. [Mandatory, recommended and optional redux](#)
2. [Excluded versus Optional classes](#)

- MD suggested using "mandatory", "recommended" and "optional" properties and not mentioning of minimum/minimal properties anymore (point 1).
  - In the issue it is explained that definitions at section 5 are not consistent as it looks like mandatory has different meanings for classes and properties once the "minimal" concept has been

introduced in mandatory properties a receiver MAY refuse to process the information for a particular resource if the information is not provided".

- o MD suggested introducing the same nuance in the mandatory classes definition to make both of them consistent.
- o MD suggested removing the definition for an optional class and a section 6.2 "Optional classes" because this is currently empty because there are no optional classes for now.
- o The WG agreed on the suggestions.
- MD suggested removing the definition and section for excluded classes (point 2).
  - o The WG agreed.

Classes and properties (section 6 & 7)
The discussion below refers to the following points:
1. Dataset versions
2. Managing optional properties for foaf:Agent
3. Walk through cardinality of properties (3a) (including issue on identifiers (3b)
4. Datasource fees
5. Mandatory properties for Distribution too limited ?

- For the versioning of datasets (point 1), MD proposed to add adms:version and adms:versionNotes to the Application Profile and at the moment not add any version pointers (for instance xhv:last, xhv:next and xhv:previous).
  - o The WG agreed.
- MD explained the issue: it was not really clear from the current version of the Application Profile, which properties are optional for foaf:Agent, skos:Concept and skos:ConceptScheme (point 2) .
  - o MD proposed to, for the time being, delete the sections mentioning optional properties for the external classes.
  - o The WG agreed.
- MD explained that he has added to cardinalities to all properties included in the Application Profile (point 3a). For mandatory properties the minimum cardinality would be 1 and the maximum cardinality 1 or *. For optional properties the minimal cardinality would be 0.
  - o The WG agreed.
- MD explained the issue raised by CI (point 3b) on having distinct identifiers for native identifiers (dct:identifier) and other identifiers (somenamespace:otherIdentifier).
  - o MD said that having two different types of identifiers could be confusing and proposed to not do it.
  - o AP commented that having two identifiers does not need to have big impact because we they can be linked with owl:sameAs.
  - o CI commented that it could be more confusing not being able to distinguish between native system IDs and other external ones.
  - o SG refers to the discussion on ADMS Identifier (http://www.w3.org/ns/adms#adms:identifier).
  - o SG said that he believes having a separate identifier would overly complicate matters. It would also not be needed, as data portals could mint their own URIs based on local identifiers.
  - o PA said that there can probably can pop-up more use cases for

having different identifiers. This is exactly where adms:Identifier was created for: to cover the situation where different kind of identifiers are relevant for different purposes.
- o MD proposed to further discuss this issue online.
- o EG said that all catalogs currently have their own identifier, but in order to have some kind of interoperability we need to have a harmonised identifier, for instance on the European level.
- o MD suggested EG to elaborate a proposal for an additional identifier.
- o LB said that they will share the Publications Office experience on identifiers with the European Data Portal.
- MD explained the issue on adding fee information for datasets as a property to the specification (point 5).
  - o PA commented that it is likely, from the discussion in the DCAT Working Group, that a fee for what DCAT calls a distribution could be useful too.
  - o MD suggested adding a property indicating the fee for the dataset.
  - o AP said that the fee is not related to the dataset itself but more to the licence applying to the dataset.
  - o PA said that he can provide information on an external property that could be reused for this property. PA suggested to use bibtex:hasPrice but stressed that this will need further investigation.
  - o SG said that adding fee information is not really a use case for the DCAT Application Profile and this could overcomplicate things.
  - o MD said that in the world of e-commerce, a price is not linked directly to a product. A price is related to an offer, offering 1 or more products. This means that it is possible that a particular piece of fee information, does not apply to everyone.
  - o MD suggested leaving this out of the specification.
  - o The WG agreed.
  - o AP suggested adding a property indicating the licence type.
  - o SG suggested using adms:licenceType.
  - o MD said that he will provide a proposal based on what is in ADMS for licence type, including the controlled vocabularies ADMS uses for that.
  - o EG commented that we do need to think about interoperability of licenses.
  - o MD asked EG to submit a proposal on that to the working group.
6. MD explained the issue "Mandatory properties for Distribution too limited ?" (point 5) and suggested to not change anything to the specification based on this issue.
  - o The WG agreed.

Accessibility and multilingual aspects (section 10)
The discussion below refers to the following points:
1. Multilingual Aspects

- MD explained that based on the issue "Multilingual aspects" (point 1) no changes will be made to the specification.

Deployment (section 11)
The discussion below refers to the following points:
1. URI sets and DCAT-AP

2. [Enable syndication of dataset metadata?](#)
3. [Clarification about the expected operational environment for DCAT-AP](#)

- MD explained that in the context of the issue "[URI sets and DCAT-AP](#)" (point 1) some best practices will be added to the specification.
  - ○ AP said there is confusion about the possible use for URIs for datasets and the metadata on datasets and proposed to add something to the specification to solve this.
  - ○ MD asked AP whether he could propose a solution.
  - ○ AP said that he will make a list of different solutions and that he will send this to the WG.
- MD said that to solve the issue "[Enable syndication of dataset metadata?](#)" (point 2) a field has been added that indicates the change type in a catalogue record.
  - ○ AP said that he will raise an issue on the provenance of the provider of the metadata because this is different than the catalogue record.
- To clarify the expected operational environment for the DCAT-AP (point 3), MD suggested to add some text in the document from the input provided by SG, NL and CI. This way it should be clear that we do not expect that owners of a portal change the way their system is organised but that they should provide a way of exporting their metadata into the DCAT-AP RDF format.

Licensing

The discussion below refers to the following points:
1. [Various licenses per dataset is problematic (i.e. one per distribution)](#)
2. [NoLicenseDefined](#)

- MD explained that if there are multiple distributions with different licences, an aggregator may need to select one distribution, e.g. the one with the "most open" licence, and forget about the rest (point 1).
- With regards to the issue "[NoLicenseDefined](#)" (point 2), MD proposed to use the controlled vocabulary for licence type for indicating that there is no license, if this is the case.
  - ○ SG disagreed because 'licence type' should be a property of a licence, whereas "NoLicenceDefined" should be a value for the 'licence' property of a dataset.
  - ○ This topic will be further discussed offline.

5. Controlled vocabularies (section 8):

Discussion

- The controlled vocabularies for theme received from Austria, Spain and Germany were mapped to the EuroVoc thesaurus. The result of this exercise can be found via the following link: [https://docs.google.com/spreadsheet/ccc?key=0AtYBrl3GPikydEppeERJb2Fx VDQzMzBZMjBnWS1KN1E#gid=0](#)
- MD said that a link will be added to the CERIF standard vocabularies. An explanation will also be added to the document that other vocabularies could be used.
- MD asked whether the WG agrees with the proposed requirements for the controlled vocabularies in the specification.

- o The WG agreed.
- AC explained that the URIs for the EuroVoc terms will be made dereferenceable in the near future. The Publications Office is currently looking at the best way to do this.
- MD said that if we look at the potential vocabularies for theme, none of them really meets the requirements.
- MD explained that for the proposed controlled vocabularies, the vocabularies already identified in the context of the ISA Programme where reused as much as possible.

6. Use cases

Discussion

- SG gave a short summary of the additional use cases that were contributed to the working group for instance about publishing legislation as open data. More information on this can be found in the corresponding issue: https://joinup.ec.europa.eu/asset/dcat_application_profile/issue/provide-your-use-case-and-user-scenario
- There is also a use case motivating the syndication of metadata records that are created, updated, or deleted after a particular point in time. This has led to the addition of a "change type" property to the CatalogRecord class. https://joinup.ec.europa.eu/asset/dcat_application_profile/issue/enable-syndication-dataset-metadata#comment-14246

7. Public review period

Discussion

- MD explained that after having processed the comments from this WG Virtual Meeting, a new version of the specification will be made available. This will be the final draft and will go into public review.
- The public review period will last for 4 weeks, and it is the intention to enable the wider public to have a look at the specification and submit any comments.
- It is also possible to comment on the proposed controlled vocabularies.
- We invite the members of the WG to promote this work and to stimulate people from outside the WG to review the work and give any comments.

8. Wrap-up, actions, next working group meeting (2013-06-12)

Discussion

- AP asked whether there is anything foreseen for this working group to map existing schemas to the DCAT-AP. AP said whether he can provide some documentation for this and an XSLT to transform XML and asked whether this could be included in the specification.

- o MD replied that this is surely useful information and that it could for sure be contributed. However, MD suggested keeping it separate from the specification because this goes beyond the goal of this WG.
- The next WG Virtual meeting will take place on 2013-06-12 from 16:00 till 18:00 CEST. All details, including agenda and connection information, will be made available in advance on the following page: http://joinup.ec.europa.eu/asset/dcat_application_profile/document/dcat-application-profile-wg-virtual-meeting-2013-06-12

ACTIONS & DECISIONS

Action	Owner	Date
Sign the Contributor Licence Agreement.	WG Members	2013-05-24
The structure of section 1.4 will be updated following the suggestions included in the issue "1.4 Structure".	MD	2013-05-13
The title of section 8.2 from "Proposed vocabularies" to "Vocabularies to be used".	MD	2013-05-13
MD to correct the typo in the section about schema.org ("schemas.org").	MD	2013-05-13
MD to mention the fact that there is also re-use from other vocabularies than DCAT and mention the particular vocabularies.	MD	2013-05-13
MD to add a pointer to adms:Identifier to the specification.	MD	2013-05-13
EG to elaborate a proposal to add a harmonised identifier.	EG	2013-05-10
LB to share the Publications Office experience on identifiers with the European Data Portal.	LB	2013-05-24
MD to provide a proposal for the type of licence based on what is in ADMS for licence type, including the controlled vocabularies ADMS uses for that.	MD	2013-05-13
EG to submit a proposal for a set of properties for licenses so licenses across countries could be compared.	MD	2013-05-10
MD to clarify the expected operational environment for the DCAT-AP so it is clear that it is not expected that owners of a portal change the way their system is organised but that they should provide a way of exporting their metadata into the DCAT-AP RDF format.	MD	2013-05-13
AP to provide a list of different solutions to solve the confusion about the possible use for URIs for datasets and the metadata on datasets.	MD	2013-05-10

MD & SG to discuss the issue [NoLicenseDefined](#) offline.	MD / SG	2013-05-13
MD to add a link to the CERIF standard vocabularies and an explanation that other vocabularies could be used.	MD	2013-05-13
Promote the work done and to stimulate people from outside the WG to review the work and give any comments.	WG members	2013-06-12

Decision

- The [meeting minutes of the 2nd WG Virtual Meeting](#) were accepted.
- The definition for an optional class and a section 6.2 "Optional classes" will be removed because this is currently empty because there are no optional classes for now.
- The definition and section for excluded classes will be removed
- adms:version and adms:versionNotes will be added to the Application Profile No version pointers (for instance xhv:last, xhv:next and xhv:previous) will currently be added.
- The sections mentioning optional properties for the external classes will be removed.
- Information about fees will be left out of the specification.
- The WG agreed on the list of requirements for controlled vocabularies included in the specification.