



Report

Swedish recommendation for DCAT-AP1.1

DCAT-AP adapted to the needs of öppnadata.se

NOTE: THIS RECOMMENDATION IS SUPERSEDED BY DCAT-AP-SE2.0.0

There is an english version of the specification and further information like harvesting instructions, FAQ, recommendations, examples etc. in swedish.

Author	Matthias Palmér, MetaSolutions AB
Commissioned by	Erik Borälv and Börje Lewin, VINNOVA
License	CC BY-SA 4.0
Date	2016-06-07
Permanent address	http://lankadedata.se/spec/DCAT-AP-SE

Version history

- 2017-09-29 - Improvements
 - Added permanent address
- 2016-06-07 - Improvements
 - Forked from 1.0 document
 - Updated all the changes from the DCAT-AP1.1 specification
 - Removed the top-level Eurovoc appendix
 - Added an appendix describing what has changed since 1.0 with respect to the Swedish adoption
- 2015-02-09 - Improvements
 - Added version history

- Inconsistency between table and discussion regarding mediaType removed
 - Note on additional information being allowed
 - Appendix with top-level Eurovoc concepts
- 2014-09-19 - First stable version

Recommendations

Summary

- Provide a single file in RDF/XML.
- Instances of Catalog, Dataset and Distributions should be URIs, not blank nodes.
- There should be at least one Catalog and at least one Dataset.
- It is recommended that each Dataset should have at least one Distribution.

How to read the tables below

Fixed / Mandatory / recommended / Optional / Avoid

Not all properties are required, the following state abbreviations shows which that should be included always and which can be left out:

S = State abbreviation	Explanation
F	Fixed value, should always be present.
M	Mandatory, at least one value should be present.
R	Recommended, at least one value is preferred.
O	Optional, provide these values if it is easy, ignore otherwise.
A	Avoid, causes confusion and should not be used.

The state abbreviation is represented in the 'S' column. In addition we also include a 'M' column that indicates if there is a maximum limit on the number of occurrences for a single property, e.g. there should be only one date of modification of each dataset.

Important note: This recommendation is focused on what is expected of a DCAT-AP expression in a Swedish context, more specifically from the perspective of the Swedish national portal öppnadata.se. That is, it is fine to provide additional information, but do not expect that information to be visible in the portal. Although the additional information will be available via the API in öppnadata.se in the RDF format. For example, it is fine to provide more specific classification of your dataset using another vocabulary or more fine-granular levels of Eurovoc as long as you also provide an explicit top-level classification. (It need to be explicit as at this

point öppnadata.se do not understand implicit information that requires inference based on OWL, RDFS or SKOS concept hierarchies.)

Dates

Dates are always expressed as datatyped literals according to xsd:date or xsd:dateTime. For example:

```
<rdf:Description rdf:about="http://example.org/distribution1">
  < dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2014-08-29</dct:issued>
  < dct:modified
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-08-29T15:31:10</dct:modified>
</rdf:Description>
```

Namespaces

Namespace	URI
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
dcat	http://www.w3.org/ns/dcat#
dct	http://purl.org/dc/terms/
adms	http://www.w3.org/ns/adms#
vcard	http://www.w3.org/2006/vcard/ns#
foaf	http://xmlns.com/foaf/0.1/
time	http://www.w3.org/2006/time#
xsd	http://www.w3.org/2001/XMLSchema#
schema	http://schema.org/

Swedish changes / adaptations

The 'C' column will indicate that there are Swedish deviations from what is recommended by DCAT-AP1.1. In most cases it is more specific information that is left undecided in the specification. But there are also places where there are major deviations, like not recommending the use of dcat:CatalogRecord.

dcat:Catalog

Property	Comment and value	S	M	C
----------	-------------------	---	---	---

rdf:type	http://www.w3.org/ns/dcat#Catalog	F	1	
dcat:themeTaxonomy	<i>Do not provide this value. There is no need to point to the concept scheme, individual datasets points to the concepts, that is enough. Even if provided, there is no rule to let information on the catalog to influence datasets.</i>	A		x
dct:title	Language literal	M		
dct:description	Language literal	M		
dct:publisher	A foaf Agent	M	1	
dcat:dataset	Points to the dataset	M		
dct:issued	Date literal, see above.	R	1	
dct:language	A URI corresponding to a language, see list at: http://publications.europa.eu/mdr/authority/language/	R		
dct:license	This is the license of the catalogue, not the datasets. The value should be a URI, such as a CC-license: http://creativecommons.org/licenses/by/3.0/ Only CC-BY and CC-0 allowed. <i>Change: Restriction to these two licenses.</i>	R	1	x
dct:modified	Date literal, see above.	R	1	
foaf:homepage	A URL to a webpage.	R	1	
dct:rights	A RightsStatement according to ODI, see below.	O	1	
dct:spatial	A reference to first or second order administrative division in geonames, e.g.: http://sws.geonames.org/2673722 <i>Change: Use of geonames instead of the recommended MDR Countries Named Authority List</i>	O		x
dcat:record	Intermediate construction of little use, hence avoid. <i>Change: No CatalogueRecords.</i>	A		x
dct:hasPart	Related catalogue	O		
dct:isPartOf	Related catalogue	O		

dcat:Dataset

Property	Comment and value	S	M	C
rdf:type	http://www.w3.org/ns/dcat#Dataset	F	1	
dct:type	<i>Change: underspecified in the specification at this point, ignore for the time being.</i>	A	1	x
dct:title	Language literal	M		
dct:description	Language literal	M		
dct:publisher	An organization as a foaf:Agent, see below	R	1	
dcat:contactPoint	A contact as a VCard expression, see below	R		
dcat:distribution	A distribution, see below	R		
adms:sample	A sample distribution with a subset of the data	O		
dcat:keyword	Language literal	R		
dcat:theme	Use the DCAT theme vocabulary, i.e. 13 themes: http://publications.europa.eu/resource/authority/data-theme/[AGRI ECON ...]	R		
dct:identifier	Main identifier as a literal	O		
adms:identifier	Alternative identifier as a literal	O		
dct:issued	Release data, an xsd:dateTime	O	1	
dct:modified	Date of modification, an xsd:dateTime	O	1	
owl:versionInfo	A literal	O	1	
adms:versionNotes	A literal with language	O		
dct:language	A URI corresponding to a language, see list at: http://publications.europa.eu/mdr/authority/language/	O		
dcat:landingPage	A URL to a webpage	O		
dct:accrualPeriodicity	The frequency of update, e.g. http://publications.europa.eu/resource/authority/frequency/[ANNUAL MONTHLY ...]	O	1	
dct:conformsTo	A URI to a formal specification of the dataset	O		

dct:spatial	A reference to first or second order administrative division in geonames, e.g.: http://sws.geonames.org/2673722 <i>Change: Always use geonames instead of the recommended to use MDR Countries Named Authority List if possible and otherwise fallback to geonames</i>	O		x
dct:temporal	A dct:PeriodOfTime, see below	O		
dct:accessRights	<i>Change: Vocabulary missing (specification says it should be developed soon), assuming namespace:</i> http://publications.europa.eu/resource/authority/dataset-access/[NON-PUBLIC PUBLIC RESTRICTED]	O	1	x
dct:source	Reference to another dataset.	O		
dct:hasVersion	Reference to another dataset.	O		
dct:isVersionOf	Reference to another dataset.	O		
dct:relation	Reference to a any kind of related resource	O		
dct:conformsTo	URL to implementing rule or specification	O		
foaf:page	URL to documentation	O		
dct:provenance	A literal with language <i>Change: restricting to literals rather than pointing to resources via URLs.</i>	O		X

dcat:Distribution

Property	Comment and value	S	M	C
rdf:type	http://www.w3.org/ns/dcat#Distribution	F	1	
dcat:accessURL	A URL to a place to access the distribution.	M		
dct:description	A literal with language	R		
dcat:mediaType	The file format given as an IANA mediatype, use format instead, avoid. <i>Change: Disallowed, dct:format is enough.</i>	A		x
dct:format	A internet media type as a string. <i>Change: Allow only internet media types, registered at IANA or unregistered in the 'x.' tree.</i>	R	1	x

dct:license	This is the license of the distribution. The value should be a URI, such as a CC-license: http://creativecommons.org/licenses/by-nc/3.0/ <i>Change: Do not require a ADMS licence type to be expressed (via dct:type) on each license URI. Since we are encouraging reuse of established licenses this should be documented elsewhere, not repeated with each use.</i>	R	1	x
adms:status	According the ADMS status vocabulary, e.g.: http://purl.org/adms/status/Completed	O	1	
dcat:byteSize	Literal som xsd:integer	O	1	
dcat:downloadURL	A direct download URL to the distribution	O		
dct:issued	Date of publication, xsd:date or xsd:dateTime	O	1	
dct:modified	Date of modification, xsd:date or xsd:dateTime	O	1	
dct:rights	A RightsStatement according to ODI, see below.	O	1	
spdx:checksum	Checksum instance, see class below.	O		
dct:title	A title for the distribution as a literal with language.	O		
foaf:page	URL to documentation resource	O		
dct:conformsTo	URL to established schema to which distribution conforms	O		

foaf:Agent

Property	Comment and value	S	M	C
rdf:type	http://xmlns.com/foaf/0.1/Agent	F	1	
dct:type	According the ADMS publisher type vocabulary, e.g.: http://purl.org/adms/publishertype/Company	R		
foaf:name	Name of person or organization as a literal	M		
foaf:homepage	Webpage of organization as a URL.	O		x
foaf:mbox	Email to organization, as a URI.	O		x

spdx:Checksum

Property	Comment and value	S	M	C
rdf:type	spdx:Checksum	F	1	
spdx:checksumValue	Checksum value	M	1	
spdx:algorithm	Checksum algorithm, e.g.: spdx:s#checksumAlgorithm_sha1	M	1	

ODI RightsStatement

Since no indication was given in the DCAT-AP specification about which RightStatement expression to use all the fields in the following table are specific to the Swedish version.

Property	Comment and value	S	M	C
odrs:AttributionText	Literal with language code	R		x
odrs:AttributionURL	Link to document	R	1	x
odrs:copyrightNotice	Literal with language code	O		x
odrs:copyrightStatement	Link to document	R	1	x
odrs:copyrightYear	A literal for the year.	R	1	x
odrs:copyrightHolder	A foaf:Organization with name, mbox and homepage, similar to foaf:Agent above.	R	1	x
odrs:jurisdiction	Country URI, use geonames.	R		x
odrs:reuserGuidelines	Link to document	R	1	x

vCard expression

Since no indication was given in the DCAT-AP specification about which vCard expression to use all the fields in the tables below are specific to the Swedish version.

vCard contact

Property	Comment and value	S	M	C
rdf:type	http://www.w3.org/2006/vcard/ns#Organization or	M	1	x

	http://www.w3.org/2006/vcard/ns#Individual			
vcard:fn	Formatted name as a literal	M	1	x
vcard:hasEmail	Email as URL, that is using the mailto protocol.	M	1	x
vcard:hasTelephone	vCard Telephone	R		x
vcard:hasAddress	vCard Address	R		x

vCard Telephone

Property	Comment and value	S	M	C
rdf:type	A telephone type, either cell, pager, audio, etc, e.g.: http://www.w3.org/2006/vcard/ns#Cell And a context, either work or home, e.g.: http://www.w3.org/2006/vcard/ns#Work	M	2	x
vcard:hasValue	A telephone number according to the tel: protocol, e.g.: tel:+4608123456	M	1	x

vCard Address

Property	Comment and value	S	M	C
rdf:type	http://www.w3.org/2006/vcard/ns#Address	F	1	x
vcard:street-address	Literal value	M	1	x
vcard:postal-code	Literal value, only digits	M	1	x
vcard:locality	Literal value, e.g. Stockholm	M	1	x
vcard:country-name	Literal value, e.g. Sweden	M	1	x

dct:PeriodOfTime

Property	Comment and value	S	M	C
rdf:type	http://purl.org/dc/terms/PeriodOfTime	F	1	
schema:startDate	Startdate, xsd:date or xsd:dateTime	M	1	

schema:endDate	Enddate, xsd:date or xsd:dateTime	M	1	
----------------	-----------------------------------	---	---	--

Discussion

Non-dereferenceable vocabularies

According to DCAT-AP a long list of vocabularies are suggested to be used, although they are not possible to be dereferenced. This is true for several vocabularies:

- Filetypes - <http://publications.europa.eu/resource/authority/file-type/RDF>
- Languages - <http://publications.europa.eu/resource/authority/language/SWE>
- etc.

Recommendation: Use the vocabularies anyway (unless otherwise recommended below) but contact someone responsible and kick them hard.

dcat:mediaType vs. dct:format

According to DCAT dcat:mediaType should be used for IANA recognized formats and dct:format for non-recognized IANA formats. The examples shows that the range is a string, e.g. "text/csv". In DCAT-AP, the range for both dcat:mediaType and dct:format is a controlled vocabulary, not a string. Hence, the distinction between mediatype and format is non-existent, unless the user is supposed to know which of the accepted values correspond to acceptable IANA mimetypes. Also, there is no mapping from the URI provided in the vocabulary to the actual IANA mediatype, only identifiers that seem to correspond to common file-extensions, e.g. "DOC", "TAR", "CSV" etc. In combination with the URIs of the vocabulary not being de-referencable the vocabulary is close to useless.

Recommendation: use IANA codes directly (as strings, not URIs) via the dct:format property. The dcat:mediaType property should be avoided. The filetype vocabulary is used to get a suitable subset of IANA mimetypes to provide as suggestions in a user interface. The user interface should also include the option to provide your own mimetype if a suitable cannot be found among the suggested.

Range of dct:spatial

According to the specification MDR continents, countries and places should be used when possible, geonames otherwise. This division is not only problematic from a technical perspective it is also hard to maintain over time. From a technical perspective it is very impractical to merge all four lists into one large list and sort out the duplicas as geonames are very large. To sort out the duplicas on the fly in a search approach lets us avoid to download large sets of data. However, the basic problem of sorting out the duplicas remains as it is not trivial to match with 100% accuracy. A manual matching process could be done but would soon get outdated. From a maintainance perspective it is also problematic if the MDR lists where ever to change, leading to a situation where old data would use geonames to indicate a place and while newer would use MRD places vocabulary.

The only solution if the combination of MDR vocabularies and geonames is to be kept is to push the problem to the end user. That is, the end user will have to first try to find a suitable location in one dropdown / search and if that fails use another. This approach would perhaps be acceptable if the person doing the metadata markup were a technical savvy person or at least someone with a deeper understanding of the standard, but this is not to be expected. Hence, this approach is not acceptable when building user interfaces that target the common user who is expecting a smooth user interface.

Recommendation: Always use Geonames.

Range of dct:Publisher

The vocabulary suggested (MDR Corporate bodies Named Authority List¹) is far from relevant in a national perspective, which is also noted in the DCAT-AP specification.

Recommendation: Rather than establishing a finished vocabulary we encourage each publisher to include a FOAF description in the form of a foaf:Organization with a few relevant properties such as foaf:name, foaf:homepage and maybe a dct:identifier for the organisation number (or a more relevant property).

License and Rights

The vocabulary specifies that for a Catalog and Distribution a LicenseStatement is recommended and a RightsStatement is optional. The difference is that a LicenseStatement is a legal document while the RightsStatement is more vague and can provide information about access etc. Unfortunately the DCAT-AP is rather vague about the expression of both. The information about LicenseDocument seems rather confusing since it is stated that a dct:type property should point to the “type of license”, this seems to indicate that an intermediate blank node should be introduced that points out the license using dct:type... This is not how it is done in the examples of ADMS nor in DCTerms.

Recommendation license: Use the CC-licenses URLs directly as they are instances of cc:License which is a subclass of dct:LicenseDocument. If needed another license can be used if a suitable URL can be found for that license.

Recommendation rights: Use the RightsStatement from the ODI, e.g. the Open Data Rights Statement Vocabulary at <http://schema.theodi.org/odrs/>.

¹ <http://publications.europa.eu/mdr/authority/corporate-body/>

APENDIX 1 Important changes from DCAT-AP1.0

The changes listed below are only those that correspond to differences between DCAT-AP1.0 and 1.1. Furthermore, changes that are not relevant for those that already complied to the Swedish adoption of DCAT-AP1.0 is left out. For instance, any changes made to CatalogRecord or changes that correspond to loosening cardinality restrictions or the change from mandatory to recommended (or to optional) are left out (but listed in the table above of course). A full list is provided in the specification.

Changes on Catalog instances

New properties

spdx:checksum
foaf:page
dct:isPartOf

Changes on Dataset instances

Property changes

adms:Contactpoint -> dcat:ContactPoint
adms:version -> owl:versionInfo

New properties

adms:sample
dct:source
dct:accessRights
dct:source
dct:hasVersion
dct:isVersionOf
dct:relation
dct:conformsTo (OBS, missing in changelog in the specification document)

Range (vocabulary) change

dct:spatial
dcat:theme
dct:accrualPeriodicity

Changes on Distribution instances

New properties

foaf:page
dct:conformsTo (OBS, missing in changelog in the specification document)