



European
Commission

DCAT-AP

Implementation guidelines

Semantic & technical

2nd Working Group Meeting

18 March 2016
ISA² Programme Action 2.1

Agenda

- Opening and welcome
- Tour de table
- Proposed guidelines
- Issues related to controlled vocabularies
- Future work on guidelines
- Issues for future DCAT-AP revision
- Next steps
- Closing



Opening and welcome

Introduction of participants - Tour de table





Report on work January-March

Issues voted by Working Group

- MI2: Dataset series
- DI1: Tools for DCAT-AP
- MAI1: Mapping national themes to MDR Data Themes Vocabulary
- UI4: Licence documents and licence URIs
- DI6: Detecting and handling duplicates
- UI1: Contact information
- MI3: Provenance
- UI6: Identifiers for datasets and distributions
- MI7: Extend DCAT-AP
- DI4: Input validation
- Other issues

See also:
https://joinup.ec.europa.eu/asset/dcat-ap_implementation_guidelines/issue/all

Interviews

- Switzerland
 - opendata.swiss
- Sweden
 - [öppnadata.se](https://oppnadata.se)
- Belgium
 - data.gov.be
- Norway
 - data.norge.no
- European Data Portal
 - europeandataportal.eu
- Netherlands
 - data.overheid.nl
- Spain
 - datos.gob.es
- Publications Office
 - open-data.europa.eu
 - publications.europa.eu/mdr/authority/
- Italy
 - dati.gov.it



Proposed guidelines

Dataset series

- No common approach across data providers.
- Suggestions for modelling:
 - Create **multiple Distributions** of a single Dataset if **users** are mostly **interested in the collection** as such
 - Create **different Datasets** if **users** are mostly **interested in the individual members**
 - If user expectations are difficult to determine, create separate Datasets and one combined Dataset with the members as Distributions

Dataset series - versioning

- DCAT-AP allows relating datasets as 'versions' using `dct:hasVersion/dct:isVersionOf` but it is not clearly described in which cases to use these properties.
- Putting a versioning scheme in place and assigning the version number as value to `owl:versionInfo` can be used for indicating precedence/sequence among different versions.
- `adms:versionNotes` can be used for describing the differences between a version and its previous one, or for indicating that a newer version is more valid than an older one.

Tools

- Call for information for tools in various categories:
 - Editors (allowing for manual input and/or input validation)
 - DCAT-AP exporters
 - Mapping of themes
 - DCAT-AP validators

Themes

- The use of the **MDR Data Themes** on all levels (local, regional, National, European) is **encouraged** as it creates coherence.
- If local or national schemes must be used, **mappings** to the MDR Data Themes should be made **available publicly**.
- European Data Portal and Publications Office should work together in co-ordinating mappings to the MDR Data Themes.

Licences

- Licences should always be **identified with URIs**, which should resolve to the description of the licence.
- **Well-known licences** should be used wherever possible.
- If a local or national licence is used, its description should link to a well-known licence on which it is based.
- Observation: Licence types are not commonly provided

Duplicates

- Assign a **stable identifier** to the dataset in the catalogue where the dataset is first published.
- Include this identifier as the value of **adms:identifier**.
- Either **locally minted** identifiers or **external** identifiers, like Datacite, DOI, ELI etc., as long as they are globally unique and stable.
- Harvesting systems should **not delete** or **change** the value of adms:identifier and only use it to compare harvested metadata to detect duplicates.

Contact information

- Provide as a minimum **name** (v:fn) and **e-mail address** (v:hasEmail), and optionally **telephone number** (v:hasTel) and **URL** (v:hasURL).
- Additional recommendations:
 - Stable contact point, i.e. an organisational email address (e.g. feedback@organisation.org), not the email of an individual.
 - Department that handles requests and feedback, if such a department exists, rather than for the organisation as a whole. For example, the URL should point to a contact page or contact form rather than to the homepage of the organisation.

Provenance

- `dct:provenance` is used in few cases for local purposes, e.g. with free text.
- Usefulness in (international) harvesting is questionable and the information may be ignored.
- Detailed provenance requirements may be satisfied with PROV-O (out of scope for DCAT-AP).

Identifiers

- **Stable URIs** should be minted for all entities.
- If possible, URIs should resolve to metadata (303 redirect).
- URIs generated on export must be unique and stable (same URI every time it is generated).
 - In RDF/XML, URI goes in `rdf:about` in `rdf:Description`
 - In JSON-LD, URI goes in `@id` keyword
- If necessary, blank node to be assigned Skolem URIs
- Dataset URI should be copied `dct:identifier`

Extension policy

- Respect the **minimum conformance** requirements as defined in section 6 of the specification of DCAT-AP v1.1
- Further policy elements:
 - May **only narrow**, not widen, usage notes
 - May **add classes and properties** but not when similar elements already exist in DCAT-AP
 - May **change obligations and cardinalities** but mandatory must stay mandatory
 - May **specify additional controlled vocabularies**

Other issues

- `dct:format` and `dcat:mediaType`
 - Only use `dct:format` and do not use `dcat:mediaType`
- Multilingual metadata
 - Always provide language tag for text
- Examples of DCAT-AP-compliant data
 - Gather examples for a wide range of cases from implementations, including edge cases

Issues – controlled vocabularies

Recommendations

- Location
 - Use Geonames always, add terms from MDR NALs (Continents, Countries, Places, Administrative territorial units) if available
- Dataset type
 - Improve MDR Dataset Types NAL
- Access rights
 - Create Access Rights vocabulary



Future work on guidelines

Possible subjects

Technical guidelines

- Service-based data access
- Description of different types of Distributions (files, APIs, nested formats)
- (Multilingual) keywords
- Data versioning
- Examples of use

Organisational guidelines

- Referring to the DCAT-AP in public procurement
- Develop business case for Linked Data

Issues – future DCAT-AP revision

DCAT-AP revision

- Various types of relationships between Datasets
- Reference to legal framework under which the data is collected or published
- Add properties like rights holder, creator
- Add expression of metadata character set, like in GeoDCAT

Next steps

Next steps

- Publication of draft guidelines on Joinup, 25 March
- Working Group members comment and review, 1 April
- Finalisation and publication of guidelines, 8 April
- Further planning of activities in April

Presentation of guidelines and further work in the **workshop in Rome on 13 May 2016**. More information available at:
<https://joinup.ec.europa.eu/node/149828>

Save the date!



12 May 2016
Rome – Italy

Stay tuned at:

<https://joinup.ec.europa.eu/node/148436>



Project Officers Vassilios.Peristeras@ec.europa.eu
Athanasios.Karalopoulos@ec.europa.eu

Visit our initiatives

ADMS ASSET DESCRIPTION METADATA SCHEMA	StatDCAT-AP FOR STATISTICAL DATASETS	GeoDCAT-AP FOR GEOSPATIAL DATASETS	DCAT-AP FOR DATA PORTALS IN EUROPE	CORE PUBLIC ORGANISATION VOCABULARY
CORE PERSON VOCABULARY	REGISTERED ORGANISATION VOCABULARY	CORE CRITERION & EVIDENCE VOCABULARY	CORE LOCATION VOCABULARY	CORE PUBLIC SERVICE VOCABULARY

Get involved



Follow [@SEMICEu](https://twitter.com/SEMICEu) on Twitter



Join the [SEMIC](#) group on LinkedIn



Join the **SEMIC** community on Joinup