Approach towards napDCAT-AP specification

Report of guidelines and best practices for the definition of a roadmap towards napDCAT-AP

Status: Final

Version: v 1.0

Date: 27 June 2022

Legal disclaimer

The sole responsibility for the content of this document lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein. All images are provided by the respective partners (unless otherwise noted) and are approved for reproduction in this publication.

Document information

Project acronym	NAPCORE
Full project title	National Access Point Coordination Organisation for Europe
Grant Agreement No.	MOVE/B4/SUB/2020-123/SI2.852232
Activity no. and title	SWG 4.4 – Work Item 2.2: Approach towards napDCAT-AP specification
Author(s)	Marco Comerio, Cefriel; Mario Scrocca, Cefriel; Antonia Azzini, Cefriel
Co-author(s)	Peter Lubrich, BASt; Hilde Austlid, Vegvesen
Related to Milestone no.	none
External Milestone	no

Document history

Version	Date	created/ modified by	Comments
0.1	04.04.2022	Mario Scrocca Antonia Azzini Marco Comerio	First draft of guidelines and best practices from the literature and published extensions.
0.2	14.04.2022	Mario Scrocca Antonia Azzini Marco Comerio	First draft of guidelines and best practices from the interviews with experts.
0.3	20.05.2022	Mario Scrocca Antonia Azzini Marco Comerio	Revision and refactor of best practices according to the defined roadmap.
0.4	27.06.2022	Mario Scrocca	Integrate feedback from SubWG4.4 partners
1.0	27.06.2022	Mario Scrocca	Final release

Action requested

☐ To be revised by partners involved in the preparation of the document

☐ For review/ approval by the Core Alignment Team

☐ For approval by the NAPCORE Steering Committee

Abstract

The report describes DCAT-AP and its available extension eliciting a set of guidelines and best practices for the definition of a roadmap towards napDCAT-AP.

Abbreviations

Abbreviation	Meaning
EC	European Commission
CAT	Core Alignment Team
NAP	National Access Point
NAPCORE	National Access Point Coordination Organisation for Europe
SC, SCOM	Steering Committee
SCS	Steering Committee Support
SWG	Sub-working Group
WG	Working Group
WP	Working Programme

Table of contents

List of figures

1. Introduction

This report represents the outcome of the Work Item 2.2 in the NAPCORE Sub-Working Group 4.4 and discusses the DCAT-AP specification and its available extensions to identify an approach towards napDCAT-AP based on guidelines, best practices, and artifacts to be produced.

The methodology followed for carrying out the Work Item is described in Figure 1. The following inputs were considered:

- Document from the literature describing DCAT-AP and its extensions
- Artefacts produced for the publication of DCAT-AP and its extensions
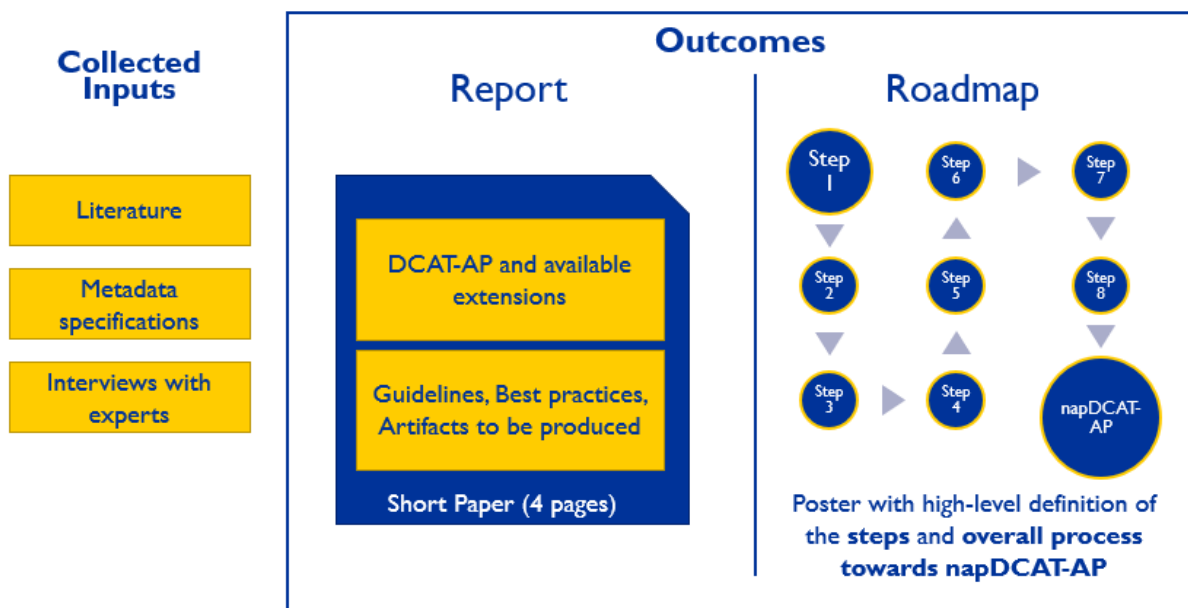- Interviews with DCAT-AP and SEMIC experts[1]

Figure 1: Methodology for the definition of a roadmap towards napDCAT-AP

The analysis of the collected inputs enabled the definition of two documents representing the outcomes of the Work Item:

- A *report*, represented by this document, describing DCAT-AP and its available extensions (Section 2), and eliciting guidelines (Section 3), best practices (Section 4), and artifacts (Section 5), to be produced for the definition of a DCAT-AP extension;
- A *roadmap,* describing through a diagram the defined steps and the overall process towards napDCAT-AP (Section 6).

2. DCAT-AP and available extensions

This section discusses DCAT-AP and the available extension for the application profile. The section reports an overview of national extensions of DCAT-AP and a description of GeoDCAT-AP and statDCAT-AP as widely accepted specifications in the corresponding domain, respectively, geospatial data and statistical data.

---

[1] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic

Approach towards napDCAT-AP specification (Report)

2.1.   DCAT and DCAT-AP

The Data Catalog Vocabulary (DCAT) [DCAT Version 2][2] is an RDF vocabulary designed to describe data catalogs using a standardized set of classes and properties. Using DCAT, publishers increase interoperability and discoverability enabling the possibility for applications to easily consume metadata from multiple catalogs.

DCAT Application Profile (DCAT-AP) [DCAT-AP Version 2.1.0][3] specifies, considering DCAT as the base vocabulary, a profile for data portals in Europe to favour the aggregation, exchange, search and automated processing of metadata. DCAT-AP specifies cardinalities and obligations (mandatory, recommended, and optional) for DCAT elements to be provided. Moreover, it defines recommendations for controlled vocabularies and additional properties to be used for metadata. The intended DCAT-AP scope is cross-border and cross-domain. For this reason, applications within a national domain, or applications in a specific domain may have different requirements and therefore may want to define extensions to the basic profile. If the DCAT standard is extended correctly and according to the common rules, even the backward compatibility can be preserved, which is to be aimed for.
In the following paragraphs, we discuss relevant available extensions defining national profiles and domain-tailored profiles.

The working group responsible for DCAT-AP specification and maintenance has more than 50 participants from 14 countries, the participation is open and regular seminars are hold for dissemination and discussion about future directions. All the material about DCAT-AP is openly available online:
- the process and methodology followed for the definition of the model [Process and Methodology][4];
- the released specifications and the associated artifacts [DCAT-AP Repository] [5];
- the change and release management policy [Change and Release][6].

The DCAT-AP repository is also used to collect and track issues reported by members of the working group but also by external users.

2.2.   National Extensions

Different national extensions of DCAT and/or DCAT-AP have been defined to support specific requirements for catalogues within a state. A set of these extensions is listed in this section to showcase different approaches in the extension of DCAT-AP.

[DCAT-AP BE][7] extends DCAT-AP by adding a theme taxonomy to categorise datasets in Belgian data portals.

---

[2] [DCAT Version 2] W3C Recommendation (2020). https://www.w3.org/TR/vocab-dcat-2/
[3] [DCAT-AP Version 2.1.0] DCAT-AP Working Group (2021). https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/210
[4] [Process and Methodology] ISA Programme (2013). https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/process-and-methodology-developing-semantic-agreements
[5] [DCAT-AP Repository] DCAT-AP Working Group (accessed online 2022). https://github.com/SEMICeu/DCAT-AP
[6] [Change and Release] ISA Programme (2017). https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/document/change-and-release-management-policy-dcat-ap
[7] [DCAT-AP BE] http://dcat.be/

[DCAT ES][8] is defined for Spain through a legal document and extends DCAT, thus presenting some divergences with DCAT-AP. It defines a theme vocabulary for datasets in Spanish and recommends the usage of a Spanish geographic vocabulary.

[DCAT-AP DE Version 2.0][9] for Germany, extends DCAT-AP defining additional properties to specify metadata, e.g., properties to describe the quality process, the geocoding, the legal details to access a resource, an additional checksum algorithm. Moreover, it includes recommendations defined in GeoDCAT-AP.

[DCAT AP IT Version 1.1][10] for Italy, extends DCAT-AP by defining additional properties (e.g., to specify the rights holder of a resource) and changing cardinalities and obligations (e.g., some properties, like *creator*, *identifier* and *downloadURL*, are made mandatory).

[DCAT AP NO Version 2.1.1][11] for Norway, extends the BRegDCAT-AP[12] extension for base registries of DCAT-AP, thus defining the usage or classes from the [Core Public Service Vocabulary][13] and the [European Legislation Identifier Ontology][14]. Moreover, it defines additional properties, e.g., *hasQualityNote* allows through a controlled vocabulary [DQVNO][15] to describe the data quality.

2.3.   GeoDCAT-AP

GeoDCAT-AP [GeoDCAT-AP Version 2.0.0][16] is the recommended DCAT-AP extension for metadata describing geospatial datasets and data services.

The objective of GeoDCAT-AP is not to replace existing recommendations for datasets in this domain, but to specify canonical mappings enabling the representation of metadata in compliance with DCAT-AP. The binding with DCAT-AP entities enables owners of geospatial information to increase the interoperability of metadata through the RDF format.

GeoDCAT-AP describes mappings to DCAT-AP for the union of metadata defined by ISO 19115 (geographic information metadata) and the INSPIRE Directive (2007/2/CE). The GeoDCAT-AP extension includes additional properties with respect to the DCAT-AP standard model [Additional GeoDCAT-AP elements][17] and a set of transformation guidelines.

The defined mappings have been organized into two groups, called *core* and *extended*. The *core* group includes mappings to support the extraction and reuse of geospatial metadata conforming to DCAT-AP from data portals and APIs following the INSPIRE/ISO 19115 directives.

The *extended* group includes the *core* mappings, defining additional transformation guidelines for geospatial data, for example, coordinate reference systems or spatial resolution.

---

[8] [DCAT ES] https://www.boe.es/eli/es/res/2013/02/19/(4)

[9] [DCAT-AP DE Version 2.0] https://www.dcat-ap.de/def/dcatde/2.0/spec/

[10] [DCAT-AP IT Version 1.1] https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0

[11] [DCAT-AP NO Version 2.1.1] https://data.norge.no/specification/dcat-ap-no/

[12] [BRegDCAT-AP] https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap/release/210

[13] [Core Public Service Vocabulary] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/core-public-service-vocabulary

[14] [European Legislation Identifier Ontology] https://op.europa.eu/en/web/eu-vocabularies/eli

[15] [DQVNO] Norwegian Profile for the Data Quality Vocabulary (DQV). Norwegian Digitalisation Agency https://data.norge.no/vocabulary/dqvno#

[16] [GeoDCAT-AP Version 2.0.0] GeoDCAT-AP Working Group (2020). https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/

[17] [Additional GeoDCAT-AP elements] https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/#geodcat-ap-classes-and-properties

2.4. StatDCAT-AP

StatDCAT-AP [StatDCAT-AP Version 1.0.1][18] is a DCAT-AP extension supporting the definition of metadata for statistical datasets and enhancing the interoperability of statistical data sources in open data catalogues adopting DCAT-AP.

StatDCAT-AP supports the description of statistical datasets in any format, such as the Statistical Data and Metadata eXchange specification ([SDMX][19]), the RDF Data Cube Vocabulary [RDF Data Cube][20] and CSV.

StatDCAT-AP provides a specification that is fully conformant with DCAT-AP and defines a small number of additions to model relevant metadata for the statistical domain. In particular, StatDCAT-AP allows describing the multidimensional structure of statistical datasets representing the numerical variables (or measurements) on different dimensions, e.g., geographic, temporal or specific dimensions.

The definition of StatDCAT-AP adopted the same [Process and Methodology][4] used for DCAT-AP. The process, based on the definition of a Working Group, submitted drafts for external public review to involve the main stakeholders. Different public channels were adopted to disseminate and share the work both during the process and after the publication of the extension, e.g., for issue tracking.

3. Available guidelines to extend DCAT-AP

A public document identifies guidelines to extend DCAT-AP to satisfy local or domain-specific requirements [How to extend DCAT-AP][21]. The rationale of the defined guidelines is to allow the extensibility of the application profile while preserving interoperability in a wider European and cross-domain environment.

The general criteria for any extension is to respect the minimum conformance requirements of DCAT-AP, thus allowing interoperability of metadata represented using the extended profile with metadata adopting DCAT-AP.

The following rules detail how to apply the general criteria:

1. Extensions must not widen but may only narrow down the usage notes as specified in DCAT-AP, so that all information provided according to the extension remains valid for DCAT-AP.
2. Extensions may add classes that are not specified for DCAT-AP; however, an extension should not add classes that are similar to DCAT-AP classes.
3. Extensions may add properties that are not specified for DCAT-AP; however, an extension should not add properties that are similar to DCAT-AP properties.
4. Extensions may change the cardinalities for properties defined for DCAT-AP respecting the following rules:
   - *Mandatory* properties in DCAT-AP must be mandatory in the extension;
   - *Recommended* properties in DCAT-AP may be declared optional or mandatory in the extension;
   - *Optional* properties in DCAT-AP may be declared recommended or mandatory in the extension.
5. Recommended and optional properties in DCAT-AP may be removed from the extension.

---

[18] [StatDCAT-AP Version 1.0.1] StatDCAT-AP Working Group (2019). https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/statdcat-application-profile-data-portals-europe/release/101

[19] [SDMX] https://sdmx.org/

[20] [RDF Data Cube] https://www.w3.org/TR/vocab-data-cube/

[21] [How to extend DCAT-AP] DCAT-AP Working Group. https://joinup.ec.europa.eu/release/dcat-ap-how-extend-dcat-ap

6. Extensions must include all the mandatory controlled vocabularies as listed in the specification of DCAT-AP.
7. Extensions may add mandatory controlled vocabularies.

4. Best practices from available DCAT-AP extensions

A report of best practices for DCAT-AP extension is not publicly available, however, they are partially documented by existing extensions. The following best practices were extracted from the analysed DCAT-AP extensions and considering the additional inputs collected, in particular, the performed interviews with experts.

4.1. Domain model

The following best practices were identified for the definition of the domain model of a DCAT-AP extension:

- The report "Process and methodology for developing semantic agreements" [Process and Methodology][4] defined in the context of the *Interoperability Solutions for European Public Administrations* (ISA) *Programme*, describes a process and methodology for the development of conceptual models and vocabularies fostering semantic interoperability between information systems. This work is an elaboration of the standardisation process and methodology of the World Wide Web Consortium (W3C)[22].
  The process involves setting up a Working Group and submitting drafts of the specification for internal and external public review, the methodology focuses on the identification of elements to be covered. Relevant steps to be performed are: (i) analyse the domain and the use cases addressed by the model to be developed, (ii) reach a consensus by developing semantic agreements among different stakeholders, (iii) define and implement a structured and clear model, and (iv) publish the model for public revision.
- A DCAT-AP extension should identify and prioritise new metadata elements considering the potential contribute that they could generate regarding data discovery.
- The definition of a DCAT-AP extension should consider existing data portals, for example, if similar properties defining specific value sets are commonly used, it is worth considering them to ensure that a reasonably generic vocabulary is defined and that a mapping can be defined.
- To support a harmonised evolution of DCAT, DCAT-AP and the related extensions the collaboration of actors involved in the respective Working Groups is encouraged.
- The definition of the conceptual model can be supported by UML-based tools relying on a standardised graphical notation and enabling an RDF export of the modelled extension. A refinement of the vocabulary can be performed using tools for ontology editing.

4.2. RDF implementation

The following best practices were identified for the definition of the RDF implementation of a DCAT-AP extension.

Best practices for the identification and implementation of classes and properties:

- The extension of DCAT-AP should consider the vocabularies reused by DCAT and DCAT-AP as preferred sources to identify new metadata elements.
- Additional metadata terms should be selected from existing well-known and well-maintained vocabularies.

---

[22] https://www.w3.org/

- The reuse of other DCAT-AP extensions, e.g., GeoDCAT-AP for geospatial metadata, should be recommended.

Best practices for the identification and implementation of controlled vocabularies:

- The usage of controlled vocabularies, preferably the EU Vocabularies maintained by the Publications Office, should be encouraged specifying a range of admissible values for metadata properties. Technically, there can be more than one controlled vocabulary for each property, but it is preferrable to avoid this design decision to facilitate the implementation of the specification by portal designers.
- The definition of new controlled vocabularies should preferably be based on the [SKOS][23] vocabulary.
- To restrict the range of a metadata property it is preferrable to define a specific sub-property, to enlarge the range of a metadata property a mapping mechanism should be defined to guarantee interoperability.

4.3.   Documentation and guidelines

The following best practices were identified for the documentation and guidelines associated with a DCAT-AP extension:

- The documentation of a DCAT-AP extension should contain an explicit definition of cardinalities for metadata properties (mandatory, recommended, optional).
- To improve interoperability of metadata it is recommended to define guidelines and conventions for the format of identifiers that should be used as value for specific metadata properties.
- The definition of canonical mappings to/from other relevant standards in the considered domain is highly recommended as annex of the DCAT-AP extension.
- An open license for the extension is preferable to promote its adoption by a large number of stakeholders.
- The definition of the governance is responsibility of the stakeholders working on the DCAT-AP extension. A governance process for the proposal, revision and acceptance of modifications should be identified and preferably stated in a document (cf., for example the one defined for DCAT-AP [Change and Release][6]).
- To facilitate the implementation of the DCAT-AP extension in data portals it is recommended to define implementation guidelines (cf., for example the one defined for DCAT-AP [DCAT-AP Implementation Guidelines][24]) listing also compatible tools for the harvesting, validation, and export of metadata compliant with the specification. The guidelines could also support implementers in the assessment of the metadata quality.

4.4.   Hosting and publication

The following best practices were identified for the hosting and publication of a DCAT-AP extension:

- A public repository (e.g., on GitHub) should be set up to collect and host all the artifacts composing the release of the DCAT-AP extension.

---

[23] [SKOS] Simple Knowledge Organization System https://www.w3.org/2004/02/skos/
[24] [DCAT-AP Implementation Guidelines] https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-implementation-guidelines/about

Approach towards napDCAT-AP specification (Report)

- The repository should support a mechanism for the maintenance of the extensions: versioning, hosting of working draft for future releases, collection and tracking of issues.
- A long-term hosting solution should be guaranteed and a *permanent identifier* (e.g., W3ID[25] or data.europe.eu[26]) should be assigned to the DCAT-AP extension to enable the possibility of changing the hosting solution while preserving the same identifier.
- The hosting solution should provide access to the extension in different formats (extension distributions) and may implement a content-negotiation mechanism to serve them according to the request made by the user.
- The hosting solution should support a URI strategy to access both the latest release and the previously published versions of the extension.
- It is recommended to index the DCAT-AP extension in the Linked Open Vocabularies (LOV) portal[27] to improve its discoverability.

Best practices for the hosting and publication of controlled vocabularies will be further analysed by SubWG4.4 Task 4.4.3.2 "Hosting of controlled vocabularies".

5. Artefacts to be produced for a DCAT-AP extension

This section discusses the artifacts to be produced for the publication and release of a DCAT-AP extension. The set of artifacts is derived considering already published DCAT-AP extensions and inputs from interviews with experts.

The suggested distributions for the release of a DCAT-AP extension are:

- HTML (and, optionally, PDF) documentation
- RDF serialisations: RDF/XML, Turtle, JSON-LD, etc.
- UML diagram in the original format (e.g., EAP-File) and additionally a graphical SVG representation

It is important to describe for each distribution of the DCAT-AP extension what is directly encoded/represented (e.g., in the diagram) and what is documented elsewhere.

The documentation of the DCAT-AP extension should contain different contents, for example:

- a description of the terminology used,
- the definition of the relevant class and properties,
- the definition of mandatory/recommended/optional constraints,
- the identification of controlled vocabularies to be used,
- the specification of canonical mappings with respect to other relevant metadata standards in the considered domain,
- a changelog considering the different version of the extension.

Finally, to support the automated execution of quality checking procedures, it is recommended to produce SPARQL queries and/or shapes adopting the Shapes Constraint Language ([SHACL][28]) that serialise the constraints defined in the documentation of the DCAT-AP extension (e.g., the cardinality constraints).

---

[25] https://w3id.org/
[26] https://data.europa.eu/URI.html
[27] https://lov.linkeddata.es/
[28] [SHACL] https://www.w3.org/TR/shacl/

6. Roadmap towards napDCAT-AP

The defined roadmap towards the publication of napDCAT-AP v1 is represented by the document "Approach towards napDCAT-AP specification (Roadmap).pdf" and is composed of the following five steps:

1. Definition of requirements for napDCAT-AP
2. Definition of the napDCAT-AP domain model
3. RDF implementation of napDCAT-AP
4. Documentation and guidelines for napDCAT-AP
5. Hosting and publication of nap-DCAT-AP

The defined roadmap references and complements this report.

The outcome of the first step of the roadmap is described in the report produce by the SubWG 4.4 WI 2.1 "Requirements Analysis for a new Metadata Specification".

7. Conclusions

This report presented the outcomes of the analysis of different inputs for the definition of guidelines and best practices to extend DCAT-AP. These results have been used to define the roadmap for the definition of napDCAT-AP as an extension of DCAT-AP for National Access Points that is represented by the document "Approach towards napDCAT-AP specification (Roadmap).pdf". The roadmap references the identified guidelines and best practices also considering the SubWG 4.4 Work Plan for the definition of the steps to be performed and their dependencies. The presented roadmap and this report will guide the next steps of SubWG 4.4 Work Plan for the specification and publication of napDCAT-AP v1.