

DCAT-AP

Webinar #1
15th of September 2021

Welcome & objectives

Objectives of this webinar

- Work towards a new minor release
- Reduction of the open issues
- Align with the community on DCAT-AP as specification

Overview issues

- 80 issues
 - Most of them have a proposed resolution on github
 - 54 require community alignment
 - 26 out-of-scope (announcements, questions, topics beyond the specification itself, editorial issues)
 - During today's webinar, we will discuss conceptual topics
 - Remaining issues: to be addressed on github.

- Current proposal for solution in draft 2.1.0 on github

<https://github.com/SEMICeu/DCAT-AP/tree/2.1.0-draft/releases/2.1.0>

Agenda

1. Conceptual topics
 1. DCAT-AP as profile of DCAT
 1. Representation/reading/profile extension agreements
 2. Updated UML figure
 2. Dataset/Distribution/Data Service alignment
 3. Alignment with external vocabularies + DCAT-AP study plan
 4. Next steps
2. Annex
 1. SHACL
 2. Specific property requests

DCAT-AP as profile of DCAT

Profile agreements

DCAT-AP as profile of DCAT

Questions like:

- *why is this property mentioned in DCAT for Dataservice not present in DCAT-AP?*
- *why is this property not present on the UML diagram?*

- Challenge:

- DCAT is large with a lot of options
- DCAT offers complex guidelines to extend / build a profile:
e.g. <https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/#qualified-forms>
dct:relation -> specific relations -> qualified relationships

- Solution:

- Representation agreements
- Reading agreements
- Profile extension agreements
- Scope agreements

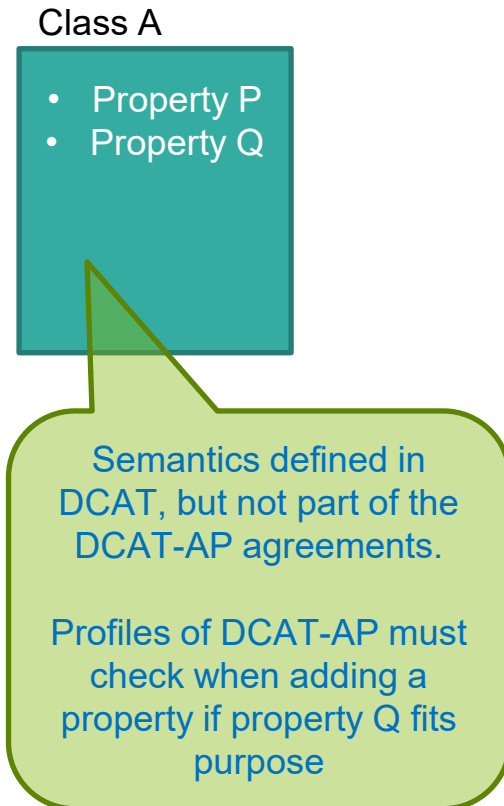
Agreement on master version

Challenge: (re)users claim different masters for DCAT-AP. (pdf/doc/html, UML diagram, SHACL, RDF, ...)

Proposal:

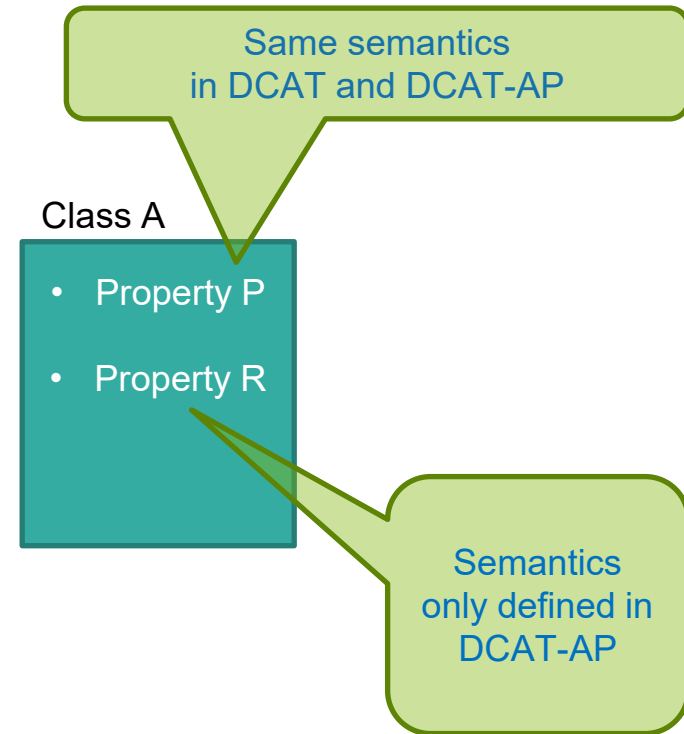
- pdf/doc/html is the master
- Other representations reflect to their best extend the expressed agreements in the master.

Profiling

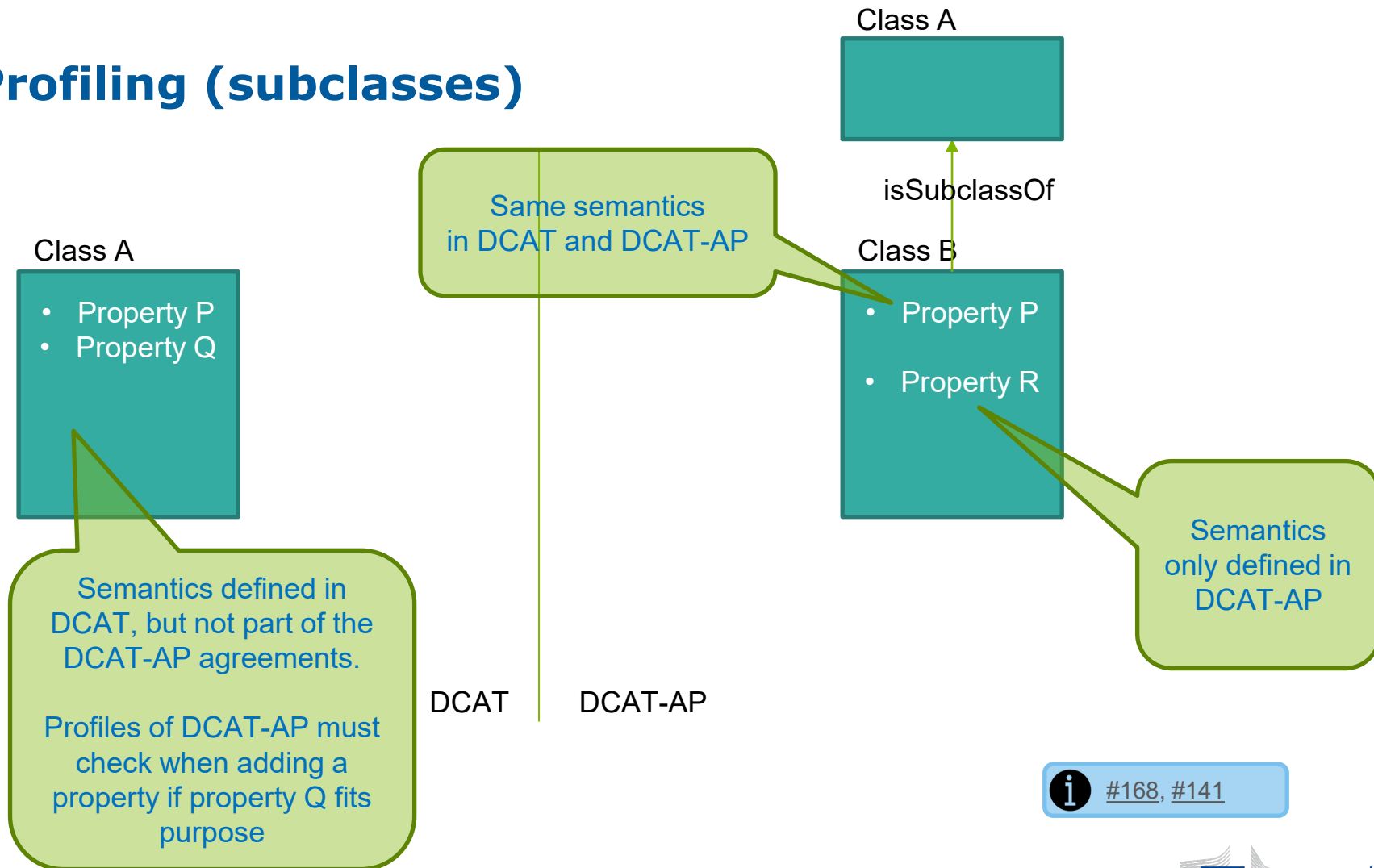


DCAT

DCAT-AP



Profiling (subclasses)



Adoption criteria for properties/classes

Challenge:

- Avoid copying the whole of DCAT as-is
- Allow both specifications have their own lifecycle

Proposal for adoption criteria:

- Bound to a cross border / EU wide alignment use case
- With a semantical precision w.r.t. DCAT (some additional information e.g. specific controlled vocabularies and cardinalities)

Adoption criteria for properties/classes consequences

dcat:Resource

- Does not represent a *business entity*
- In DCAT acts as a mean to condense the specification to create a common class to hook most common properties on
- It only has editorial purposes

Proposal:

- Not to include it in DCAT-AP
- Makes it easier for implementers to focus on the entity as a whole

Adoption criteria for properties/classes consequences

Assessment of optional properties

Class	Properties with no/minimal difference with DCAT	
Catalogue	<ul style="list-style-type: none"> Spatial/geographic themes 	
Catalogue Record		
Dataset	<ul style="list-style-type: none"> Conforms to Documentation Has version Identifier Is referenced by Is version of 	<ul style="list-style-type: none"> Provenance Qualified attribution Related resource Spatial resolution Temporal resolution Was generated by
Distribution	<ul style="list-style-type: none"> Documentation Download URL 	<ul style="list-style-type: none"> Spatial resolution Temporal resolution
Data Service		

Above properties mentioned in DCAT-AP have no/minimal difference with W3C DCAT.
Equal is definition/usage note/cardinality/controlled vocabularies to use

 [#194](#), [#190](#)

Adoption criteria for properties/classes consequences

Assessment of optional properties – should they be maintained?

Proposal:

a) Keep them for now, but create proposals to remove for next iteration.

b) Add following sentence

"Any property that is mentioned in DCAT as applicable to [this] class, but is not explicitly listed in DCAT-AP, is considered an optional field for DCAT-AP."

Qualification of classes and properties

- **Mandatory class:** a receiver *MUST* be able to process information about instances of the class; a sender *MUST* provide information about instances of the class.
- **Recommended class:** a receiver *MUST* be able to process information about instances of the class; a sender *SHOULD* provide information about instances of the class; a sender of data *MUST* provide information about instances of the class, if such information is available.
- **Optional class:** a receiver *MUST* be able to process information about instances of the class; a sender *MAY* provide the information but is not obliged to do so.
-
- **Mandatory property:** a receiver *MUST* be able to process the information for that property; a sender *MUST* provide the information for that property.
- **Recommended property:** a receiver *MUST* be able to process the information for that property; a sender *SHOULD* provide information for that property; a sender *MUST* provide information for that property, if such information is available.
- **Optional property:** a receiver *MUST* be able to process the information for that property; a sender *MAY* provide the information for that property but is not obliged to do so.
-

Qualification of classes and properties

Proposal:

- **Mandatory class:** a receiver *MUST* be able to process information about instances of the class; a sender *MUST* provide information about instances of the class.
- **Recommended class:** a receiver *MUST* be able to process information about instances of the class; a sender *SHOULD* provide information about instances of the class; a sender of data *MUST* provide information about instances of the class, if such information is available.
- **Optional class:** a receiver *MUST* be able to process information about instances of the class; a sender *MAY* provide the information but is not obliged to do so.
- **Deprecated class:** a receiver *SHOULD* be able to process information about instances of that class; a sender *SHOULD NOT* provide information about instances of that class.
- **Mandatory property:** a receiver *MUST* be able to process the information for that property; a sender *MUST* provide the information for that property.
- **Recommended property:** a receiver *MUST* be able to process the information for that property; a sender *SHOULD* provide information for that property; a sender *MUST* provide information for that property, if such information is available.
- **Optional property:** a receiver *MUST* be able to process the information for that property; a sender *MAY* provide the information for that property but is not obliged to do so.
- **Deprecated property:** a receiver *SHOULD* be able to process information about instances of that property; a sender *SHOULD NOT* provide the information about instances of that property.

Qualification of classes and properties

Question: What does the community expects from a mandatory class?

- **Mandatory class:** a receiver *MUST* be able to process information about instances of the class; a sender *MUST* provide information about instances of the class.
- **Recommended class:** a receiver *MUST* be able to process information about instances of the class; a sender *SHOULD* provide information about instances of the class; a sender of data *MUST* provide information about instances of the class, if such information is available.
- **Optional class:** a receiver *MUST* be able to process information about instances of the class; a sender *MAY* provide the information but is not obliged to do so.
- **Deprecated class:** a receiver *SHOULD* be able to process information about instances of that class; a sender *SHOULD NOT* provide information about instances of that class.
- **Mandatory property:** a receiver *MUST* be able to process the information for that property; a sender *MUST* provide the information for that property.
- **Recommended property:** a receiver *MUST* be able to process the information for that property; a sender *SHOULD* provide information for that property; a sender *MUST* provide information for that property, if such information is available.
- **Optional property:** a receiver *MUST* be able to process the information for that property; a sender *MAY* provide the information for that property but is not obliged to do so.
- **Deprecated property:** a receiver *SHOULD* be able to process information about instances of that property; a sender *SHOULD NOT* provide the information about instances of that property.

DCAT-AP as profile of DCAT

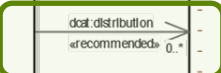
Updated UML figure

 #186,#185,#183,#182,#181,#173,#170,#168,#161



UML figure

- The intent of the figure is to give a visual overview of the specification. It is not a formal definition of DCAT-AP.
- Does not contain every supporting class as class box (readability, maintainability).
- Literal ranges are not included in the figure (e.g. due to xsd:Date or xsd:DateTime).
- If the figure does not contain a notion that exists in DCAT, it does not mean it is prohibited.



UML figure



Use of URIs in the figure

Suppose Dataset has properties
gdpr_risk_category
dataset_popularity_level

And both are mapped on dct:type, then this figure will not show there are 2 distinct properties

Dataset/ Distribution/ Data Service


Feedback

Dataset/Distribution/Data Service

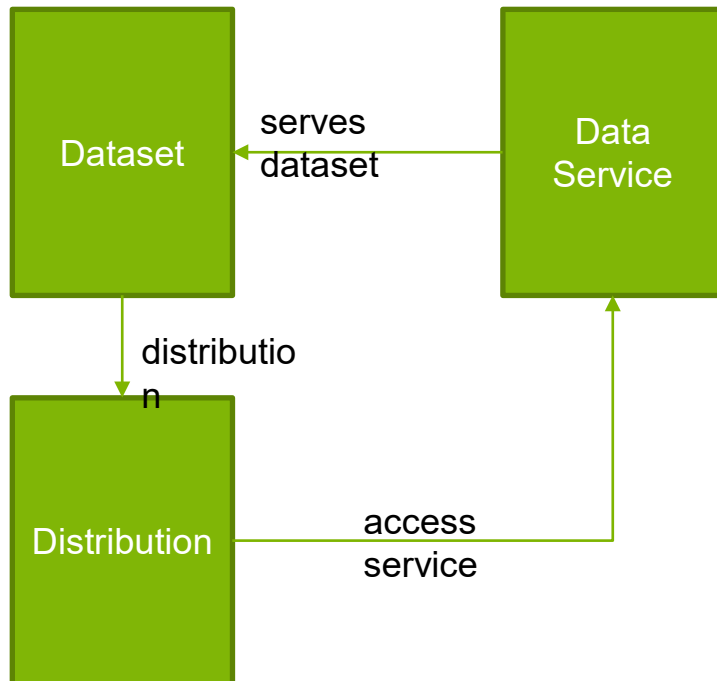


[#202,#180,#144,#142,#123,#109,#100,#91](#)

Guidance on Dataset/Distribution/Data Service

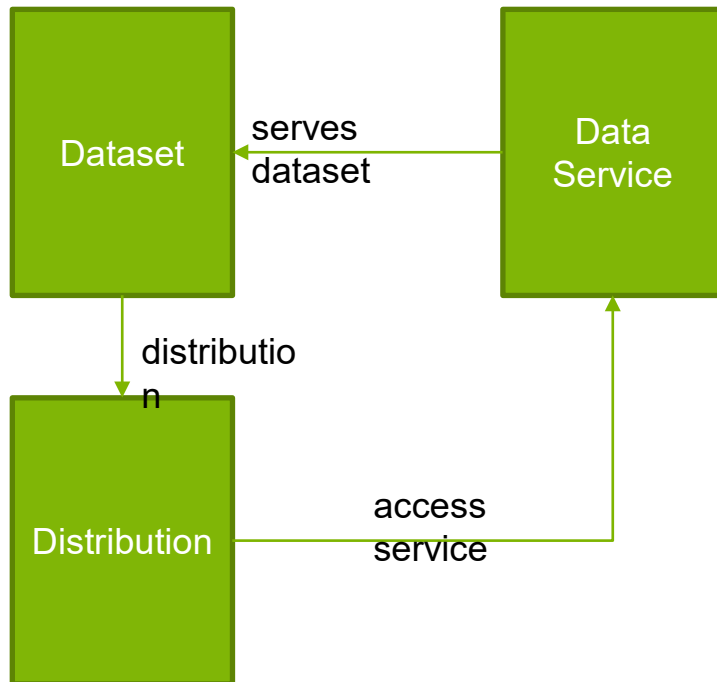
- What is the distinction between Distribution & Data Service?
- Examples/preferred modelling?

Existence dependency



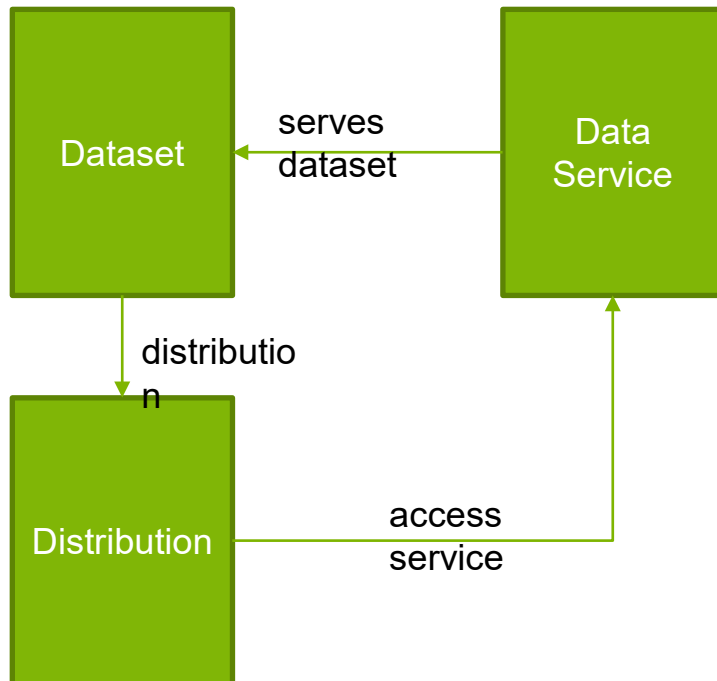
- A distribution cannot exist without a dataset
- A data service can be independent from a dataset or distribution
- DCAT-AP creates a stronger dependency between dataset and distribution: namely datasets without a distribution are incomplete, not desired (distribution is recommended)

Nature of properties



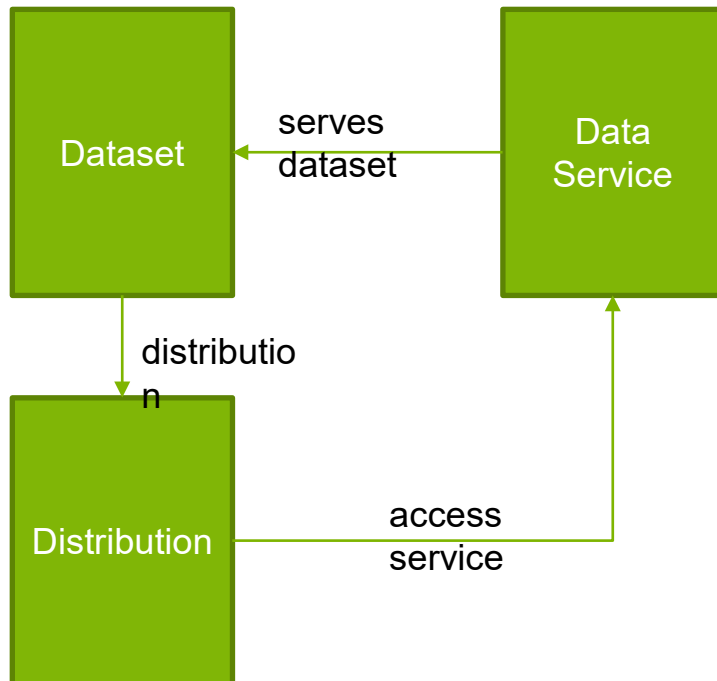
- Distributions properties are often file oriented:
 - Format
 - Byte size
 - Checksum
 - Modification date
 - Download URL/access URL
 - ...
- Relevance of those properties is reduced for services:
 - Services do transformations on the fly
 - Services do format transformations on the fly
 - Services do language transformations on the fly
 - Services do schema transformations on the fly
 - Data supplied by a service is trusted because one only gets access when access is granted (so no need for checksums, etc)

Current practice (from DCAT 1.x)



- APIs and file dumps are all expressed as Distributions

Proposal



- Use Distribution for file-based access to the content of the dataset.

Normally, a distribution contains all datapoints covered by the dataset.

Normally, the content of a distribution cannot be manipulated by the user (download).

- Use Data Service for anything else.

Consequence:

- Datasets without (file-based) distributions are very valid.

Example Distributions

- A RDF file dump
- A CSV downloadable file
- A ZIP file containing a collection of Excel files
- A jpeg image of a heatmap

Example Data services

- A SPARQL endpoint
- A REST-API for public organisations register
- A WFS service
- A download service that allow the user to parameterize the data that should be provided as a dump (e.g. a selection from the ESTAT data cube)

Dataset/ Distribution/ Data Service


Feedback

Alignment with external vocabularies

Alignment with dcterms 2020

- In 2020 dcterms changed for some properties the range restrictions from `rdfs:range` to `dcam:rangeIncludes`
 - Same meaning but disconnected from `rdfs` reasoners. It is up to the users to enforce/activate `rdfs` reasoning
 - DCAT 2.0 did not change any of their statements, and maintained the `rdfs:range`
- Proposal:
 - Follow the DCAT approach

Alignment with DCAT

- DCAT-AP 2.0.0 has not updated its own definitions/usage notes when the alignment was performed.
- Since DCAT-AP is a *strict* application profile of DCAT, the definitions of DCAT are those that provide the semantics.
- Do we consider the usage notes in DCAT-AP as definitions for the term? Or if we apply the distinction as we are now applying to the other Core Vocabularies: definition + usage note.
- Proposal
when creating a html representation apply the CV approach, copy the definition of DCAT into the specification and move any additional information to the (new) usage note.

 [#128,#135,#179,#150](#)

Future alignment with DCAT 3.0

DCAT 3.0 proposes

- Dataset Series
- Dataset Versioning

If you have any comments on these topics we prefer the discussion to happen at the level of W3C.

DCAT-AP Study

DCAT-AP study

- Objective:
 - to propose a strategy and roadmap for the next period (2-3 years)
- Approach:
 - Analyse the functional issues raised on GitHub
 - Analyse national and domain-specific extension
 - Identify potential functional improvements for further versions of DCAT-AP
 - Align with W3C
- Intended result (target: December 2021):
 - Formulate recommendations for further development
 - Propose time plan

DCAT-AP study: profiles

National profiles

- Belgium: [DCAT-BE](#)
- Belgium/Flanders: [DCAT-AP-VL](#)
- Denmark: [DCAT-AP-DK](#)
- Germany: [DCAT-AP.de](#)
- Ireland: [Open Data Technical Framework](#)
- Italy: [DCAT-AP_IT](#)
- Netherlands: [DCAT-DONL](#)
- Norway: [DCAT-AP-NO](#)
- Spain: [Norma Técnica de Interoperabilidad](#)
- Sweden: [DCAT-AP-SE](#)
- Switzerland: [DCAT-AP-CH](#)

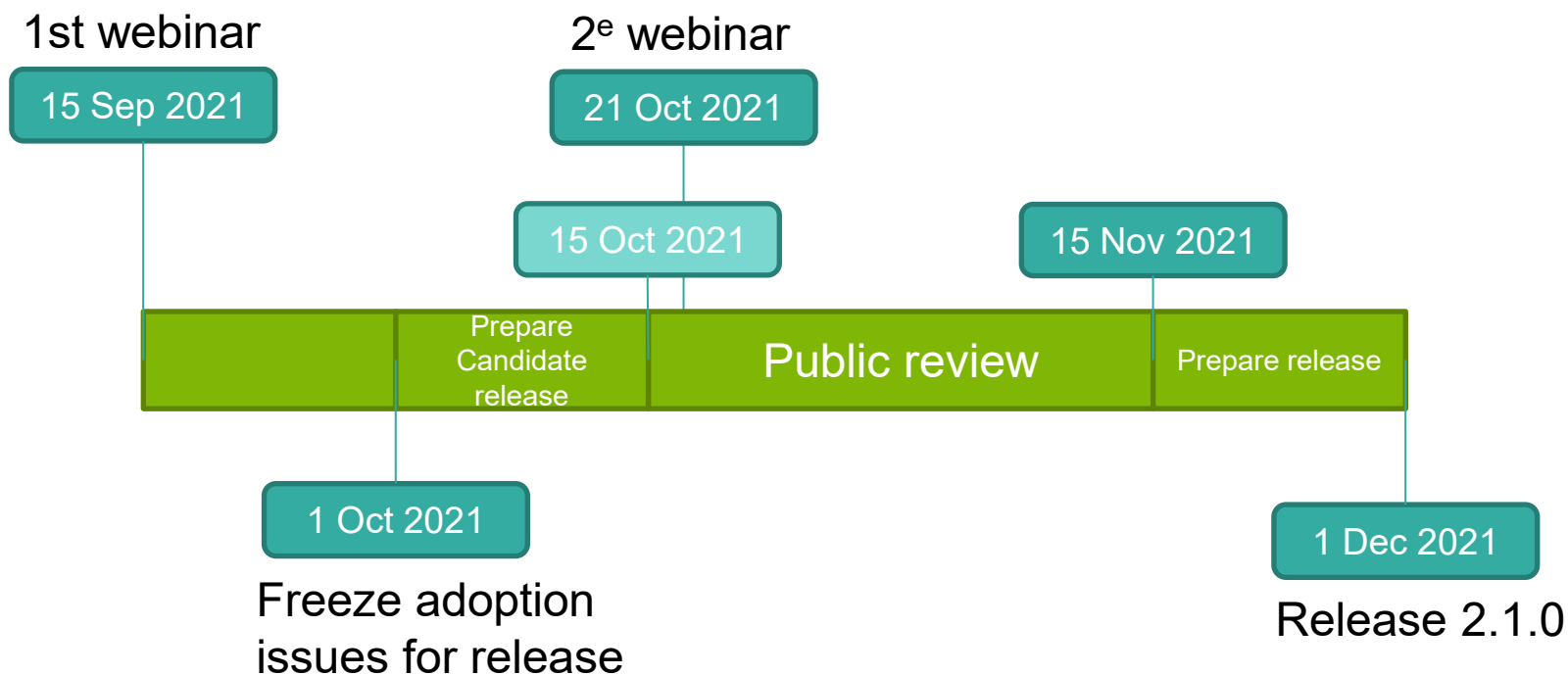
Domain-specific profiles

- Base registers; [ABR-BregDCAT-AP](#)
- Geospatial data: [GeoDCAT-AP](#)
- Joint Research Centre: [DCAT-AP-JRC](#)
- National access points: [napDCAT-AP](#)
- Publications Office of the EU: [DCAT-AP-OP](#)
- Statistics: [StatDCAT-AP](#)

Please let us know if there are any additional profiles that we should consider

Next steps

Planning



Github issues

<https://github.com/SEMICeu/DCAT-AP>

Issues that are '*closed*' (34)

status:wont-fix

- no effect on the specification

status:fixed

- request applied in the draft specification

status:waitingForDecisionW3C

- issue to be handled at W3C

Future-work (12)

- to be handled in next release

Issues that request '*community feedback*' (28)

status:resolution-provided

- resolution provided, but not (yet) in the draft specification

feedback-requested

- issue request community feedback to create a resolution

topic:cataloged_resource

- issue related to the nature of a dataset/data service/distribution.

Announcement



<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/event/core-vocabularies-revision-webinar-6th-october-2021>



<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/semic-conference-2021>

A.O.B.

Questions?



European
Commission

**Thank you for your
participation**

SHACL

State of play

SHACL

<https://www.itb.ec.europa.eu/shacl/dcat-ap/upload>

Option	Meaning
Base Zero (no background knowledge)	All constraints that must be satisfied to be technical coherent, excluding range class membership constraints and usage of controlled vocabularies.
Ranges Zero (no background knowledge)	All range class membership constraints. These are separated from the previous set because this information almost never leads to a validation error. The declaration of a sender that a value V is a member of the class R cannot be contradicted by an independent external source. Therefore to minimise the data volume between sender and receiver, the receiver can rely on that (trust that) the received value is a member of the range class. And therefore these rules should not be executed because they are always valid. Observe that validating or not these constraints is an implementation choice for the exchange between sender and receiver.
Base (with background knowledge)	Case 1 extended with background knowledge. The background knowledge is compiled of all the vocabularies that are used in DCAT-AP.
Ranges (with background knowledge)	Case 2 extended with background knowledge.
Recommendations (with background knowledge)	All constraints concerning recommended properties.
Controlled Vocabularies	All constraints concerning controlled vocabularies. This option will load also the controlled vocabulary in order to validate if a value is a member of the controlled vocabulary
Full (with background knowledge)	The union of cases 3, 4, 5 and 6.

SHACL topics

- Proposal:
 - deprecate the older SHACL files in favor of these.
 - So no further updates on older DCAT-AP versions like 1.2.1

SHACL: validation of controlled vocabularies

- CV have different statuses in DCAT-AP (mandatory, recommended, advised, conditionally)
- Validation of whether a node is a member of a CV can be expressed in different ways (depends on the profile, CV publication, data feed representation that is being validated)
- DCAT-AP contains the idea to motivate to use well-published and well-maintained CV's (goal beyond DCAT-AP as such)
- Make SHACL constraints reflect as close as possible the constraints expressed in the specification.

DCAT-AP enforces minimal data for a codelist
PO will incrementally align with these requirements

SHACL: validation of controlled vocabularies

Currently implemented as:

- *Ranges* check if a value is explicitly denoted as a `skos:Concept` with severity error
- *Controlled Vocabularies* check if a value is a member of a known mandatory Controlled Vocabulary by
 - importing the codelist, and
 - checking if the value has a statement `skos:inScheme`

SHACL: validation of controlled vocabularies

Example

Ex:dataset1 dcat:theme data-theme:EDUC

- ⇒ *Ranges* results in **error**
- ⇒ *controlled vocabulary* result in **success**

SHACL

?
Feedback

Specific property issues

dct:rights/dct:license

Label	Domain	Range	Usage Note	Card
Rights	dcat:Catalogue	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Catalogue.	0..1 (O)
Rights	dcat:Distribution	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Distribution.	0..1 (O)
licence	Dcat:Distribution	dct:LicenseDocument	This property refers to the licence under which the Distribution is made available	0..1 (R)
licence	Dcat:DataService	dct:LicenseDocument	This property contains the licence under which the Data service is made available.	0..1 (O)
licence	dcat:Catalogue	dct:LicenseDocument	This property refers to the licence under which the Catalogue can be used or reused.	0..1 (R)

- License is recommended to stimulate adding legal reuse context
- Historically, only applied at the most granular level: distribution and catalogue

dct:rights/dct:license

Request:

- Make dct:rights recommended/mandatory
- Make dct:license optional

Motivation for request

- a licence is not always possible (cfr Czech Republic legislation)

Proposal:

- Keep as current situation as is, unless there is the need to promote rights to recommended. DCAT-AP does not prevent to provide only dct:rights.

temporalResolution max 1

- Request:
 - Add max cardinality of 1 for temporalResolution
- Motivation Request:
 - Not provided
- Proposal:
 - Unclear if this applies to Dataset or Distribution
 - Keep open unless community has a concrete opinion

Add dct:identifier to Agent

- Request:
 - Add dct:identifier to Agent as (recommended) property
- Motivation:
 - Identifiers ease the maintenance of the catalogue (see <https://ec-jrc.github.io/dcat-ap-jrc/#req-identifiers>)
- Proposal/Discussion:
 - Shall DCAT-AP enforce identifiers on all entities? Are there common guidelines on identifiers?
 - Note that one cannot assume uniform representation identifiers/central catalogue of all identifiers.

Definition identifier Dataset max-cardinality

- Request:
 - Definition states *This property contains the main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue*
 - Restrict max-cardinality to 1
- Motivation:
 - Definition seem to imply there is only one identifier
- Proposal/discussion:
 - No resolution proposal made

Dataset dct:type max-cardinality

- Request:
 - Lift max cardinality restriction of 1 in DCAT-AP
- Motivation:
 - Ability to categorize within multiple codelists
- Proposal/discussion:
 - Expectation management, specification design

Catalog `dct:isPartOf` max-cardinality

- Request:
 - Lift the max cardinality of 0..1 for Catalogue `dct:isPartOf`
- Motivation:
 - Seems more intuitive
 - But no concrete use-case provided
- Proposal:
 - no change

Range spdx:algorithm

- Request:
 - Lift unique allowed value SHA1
- Motivation:
 - Allowance of more secure algorithms
- Proposal:
 - Add the requirement that the algorithm must be one <https://spdx.org/rdf/terms/#d4e1968>

Dct:creator max-cardinality

- Request:
 - Lift the max restriction for dct:creator
- Motivation:
 - Allow DCAT-AP to be used in Open Science Cloud
- Proposal:
 - Definition/usage note change in DCAT-AP: remove *primarily*
 - lift max cardinality
- Consequence:
 - There is not difference anymore with DCAT, so maybe removing it at all?

Temporal range expression

- Request:
 - To extend temporal literal not only to be xsd:Date or xsd:DateTime, but all possible xsd temporal expressions
- Motivation:
 - DCAT allows all cases
- Proposal/discussion:
 - No resolution proposal made

Guidelines & best practices for geographical coverage

- Request:
 - To provide guidelines, best practices and guidelines for geographical coverage
- Motivation:
 - Better harmonized usage
- Proposal/discussion:
 - No resolution proposal made

Guidelines & best practices for dcat:temporalResolution & dcat:spatialResolutionInMeters

- Request:
 - To provide guidelines, best practices and guidelines for temporal resolution and spatial resolution in meters
- Motivation:
 - Better harmonized usage
- Proposal/discussion:
 - No resolution proposal made

Lifecycle management

Background:

- **Lifecycle management** is about to know in which stage of life the Dataset is.
- Catalogue Record – change type (Adms:status) : This property refers to the type of the latest revision of a Dataset's entry in the Catalogue.
- Distribution – status (adms:status): This property refers to the maturity of the Distribution.
 - Completed/Deprecated/Under Development/Withdrawn
- Distribution – availability (dcat-ap: availability) : This property indicates how long it is planned to keep the Distribution of the Dataset available.
 - Temporary/experimental/available/stable

Lifecycle management

Background:

- **Lifecycle management** is about to know in which stage of life the Dataset is.
- Catalogue Record – change type (Adms:status) : This property refers to the type of the latest revision of a Dataset's entry in the Catalogue.
- Distribution – status (adms:status): This property refers to the maturity of the Distribution.
 - Completed/Deprecated/Under Development/Withdrawn
- Distribution – availability (dcat-ap:availability) : This property indicates how long it is planned to keep the Distribution of the Dataset available.
 - Temporary/experimental/available/stable
- DCAT 2.0 has adopted adms:status: The status of the resource in the context of a particular workflow process
 - https://w3c.github.io/dxwg/dcat/#Property:resource_status

Lifecycle management

Proposal:

Definitions of Catalogue Record:change type and Distribution:status might be rewritten to match better the DCAT definition.

- Change type is the status of the catalogue record in the context of editorial flow of the Dataset and Dataservice descriptions
- Distribution status is the status of the distribution in the context of a high level lifecycle

Status: dereferenceable URI dcat-ap:availability

Status: is in the process to be enabled

Status: cv availability

- In order to improve quality of publication, the CV will be now maintained, like any other CV, by the PO.
- impact on the used URIs

Proposal:

- The URIs of the concepts are changed to the pattern <http://publications.europa.eu/resource/authority/distribution-availability/{concept}>.
- The URI of the conceptscheme is changed to <http://publications.europa.eu/resource/authority/distribution-availability>.
- Old URIs will be deprecated

Status:

- To be released on 8 dec 2021 => to be finalized by 1 nov 2021
- Open question on the title: (enlarging the scope)