European Commission

25 NOVEMER 2022

SEMIC Webinar

DIGIT.D2 - Interoperability.

interoperable europe

Objective of this third webinar

- Discussing open DCAT-AP issues
- Initiating the adoption of W3C DCAT 3.0 into DCAT-AP

Agenda

1. **Welcome**
2. **Context of DCAT-AP**
3. **Discussion on open issues**
4. **Short break**
5. **Alignment with W3C DCAT 3.0**
6. **Wrap up and next steps**

Workshop practicalities

Please mute your microphones

You can also share your questions for the Q&A session via the chat

The workshop will be recorded

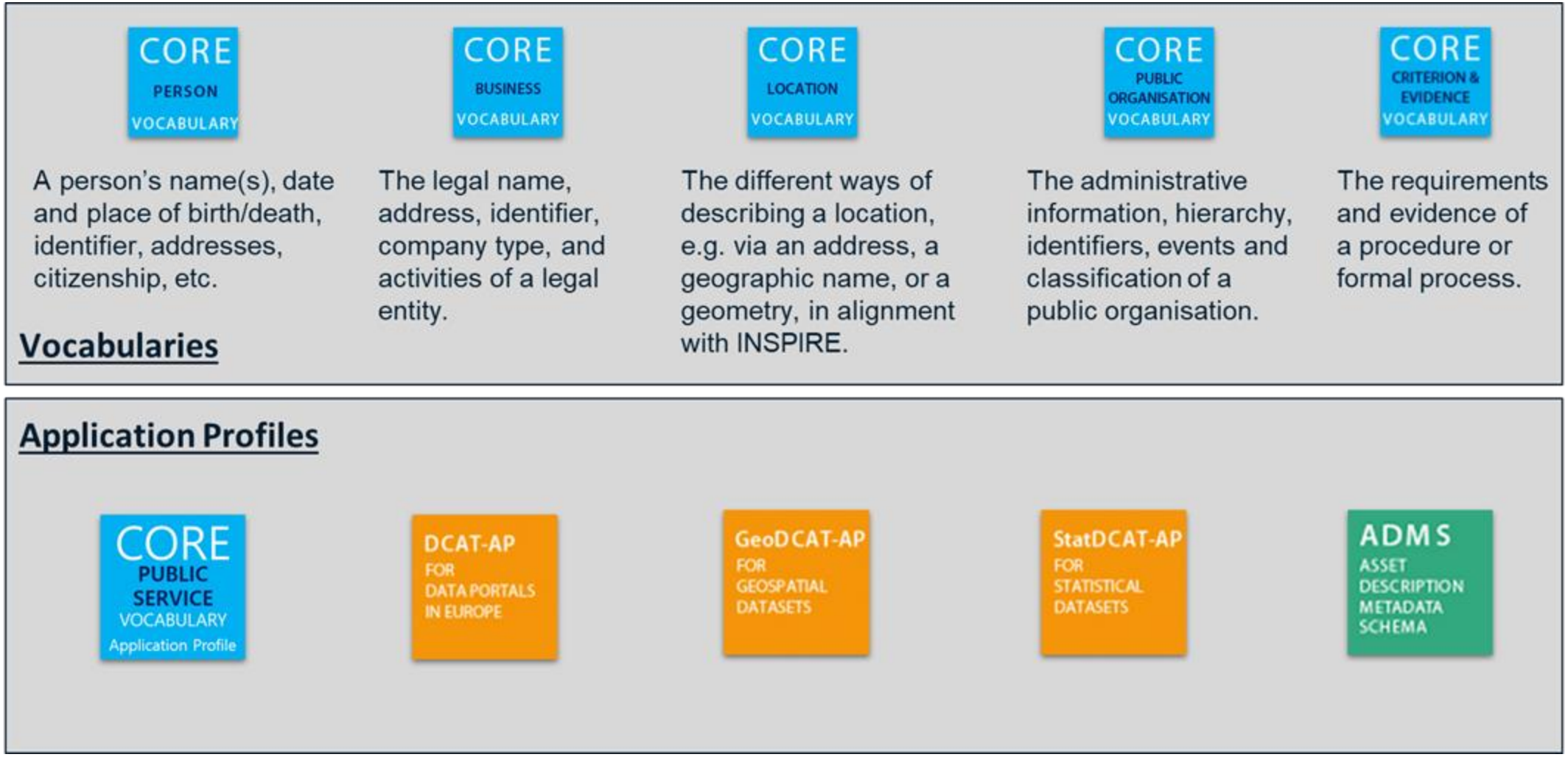
Context of the SEMIC assets

Introduction to SEMIC

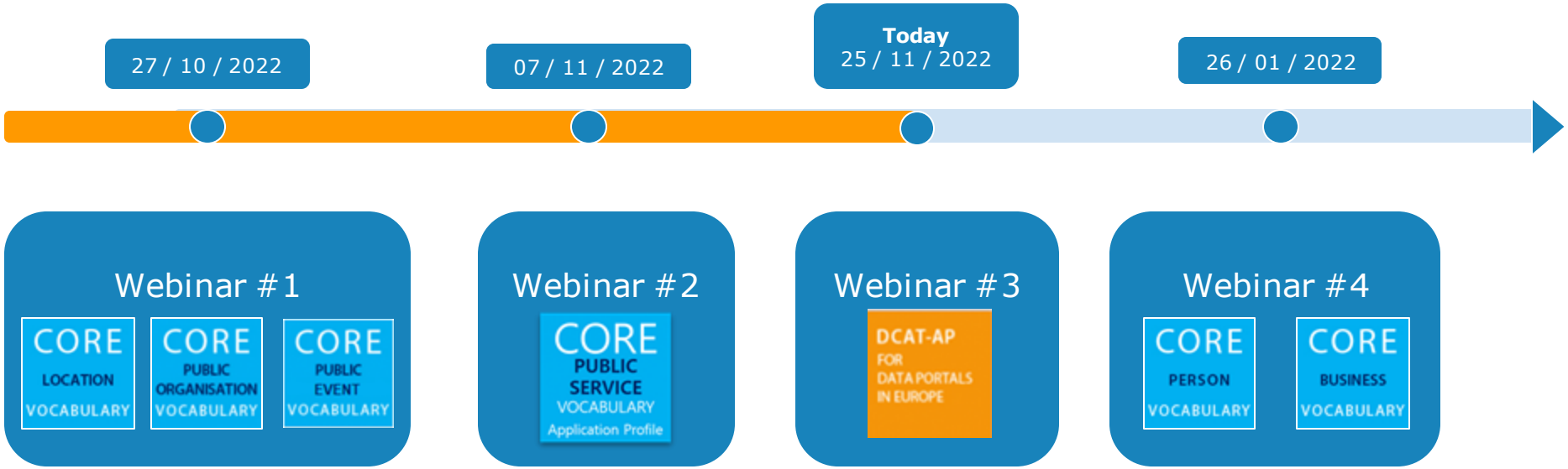
The objectives of the SEMIC action is to promote Semantic Interoperability amongst the EU Member States by:

- Promoting, share and reuse of semantic assets, experience and tools and facilitating agreements in key areas.
- Identifying opportunities for alignment on semantic definitions, metadata and reference data sources with special focus on identification and definitions of Core Concepts / Vocabularies.
- Raising awareness on the importance of data and metadata management.

Current SEMIC assets



Timeline webinars



Objectives of DCAT-AP

- Supporting the discovery of/access to (open) data in a cross-border and cross-domain environment, by harvesting data from distributed portals.

- In the form of an application profile of W3C DCAT, by
  - expressing constraints and usages on DCAT properties and classes, and
  - including additional properties and usages of controlled vocabularies

  In such a way that the metadata descriptions are maximally harmonised across Europe, and provide a reliable source for the European Data portal.

- To serve communities better extensions exist: BregDCAT-AP, GeoDCAT-AP, StatDCAT-AP.

New release DCAT-AP

- Bug fixes
  - Only on GitHub, not during this webinar
- Addressing some open issues
- HTML representation
  - Only on GitHub, not during this webinar
- Alignment with W3C DCAT 3.0

ADMS

- SEMIC has reached out to purl.org and W3C to enable "refreshing" the existing content
- Initial preparations have started for a refreshed release of ADMS.
- Link: https://github.com/SEMICeu/adms

DCAT-AP issues

Agenda

- Format for Data Services
- Spatial issues
- SHACL issues
- Legal information
- Various short issues (if time permits)
- Reminders

Data service > Format (issue #217, also W3C #1055 #1381, Breg-DCAT-AP)

Mentioned use-cases

- **Use Case 1:** filter/search for *the data of datasets in a desired representation*
- **Use Case 2:** filter/search for *data services that provide a dataset in a desired representation*
- **Use Case 3:** (automatically) *detail the* (representation) *capabilities of a data service*

Use case 1

Filter/search for the data of datasets in a desired representation

As-is

- If there is a distribution within the desired format → found
- If not, then the associated data service could possibly provide it
  - → Unknown
  - → When the complete access to the data is via a service then the metadata does not provide any clue.

Use case 2

Filter/search for data services that provide a dataset in a desired representation

As-is

- Not possible today
- alternative search possible: which are the services that conform to a desired protocol
  - e.g. instead of searching for RDF, one searches for SPARQL endpoints

Use case 3
(automatically) detail the (representation) capabilities of a service

As-is

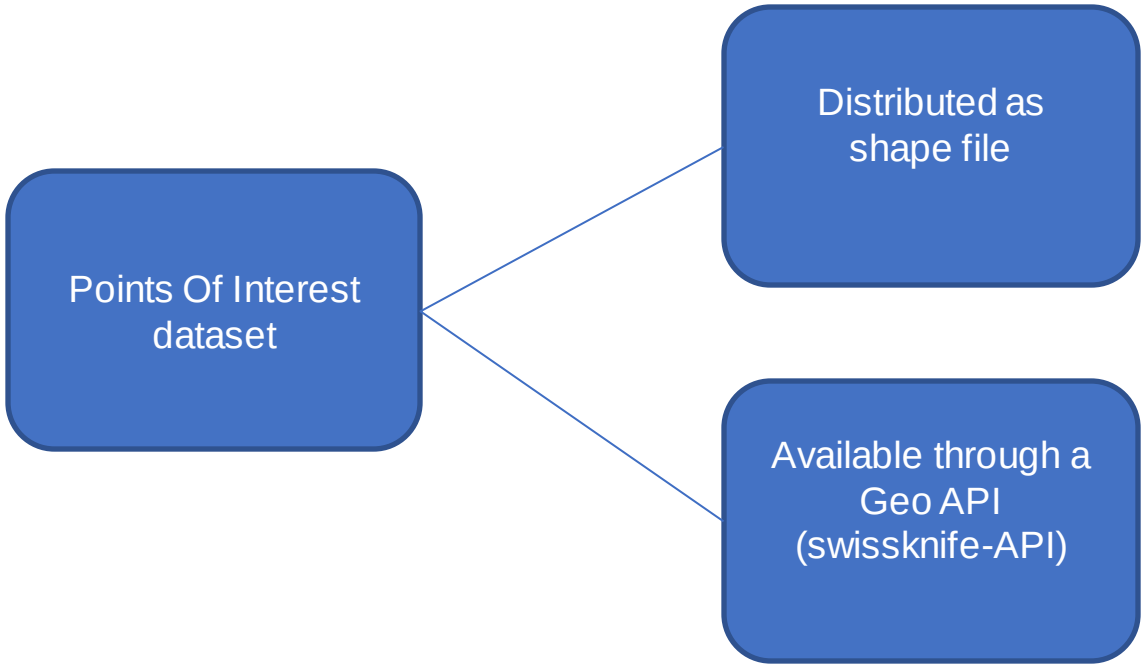
Property endpointDescription refers to

- Static technical descriptions: OpenAPI, API documentation, …
- Self descriptive services: Hydra, OGC capabilities services, …

→ No data extraction/mappings are currently provided to link both worlds.

→ **Challenge:** how to keep "metadata description" and "service capabilities" in sync?

Example



- Single snapshot
- In well defined shape schema (column names are fixed)

- Multiple technical formats: JSON, XML
- Multiple Geo specific formats: WKT, GeoJson, GML
- Multiple CRS
- Multiple file formats and packaging formats: Geopackage, File geodatabase, Shape, …

Approach

- Do the use cases, as described, address real usage patterns?
- If so, which use case should be considered?

Proposal: format definition – possible options

A data service indicates a format:

1. The structure that **is** returned by querying the endpointURL
   (implies max 1 value, and no other possibilities)
2. The structure that **can be** returned by querying the endpointURL
   (implies multiple values)
3. The structure that **by default is** returned by querying the endpointURL
   (implies max 1 value)

Range of locn:geometry is locn:Geometry (reopen issue #177)

Observation

- In DCAT 2.0: rdfs:Literal
- DCAT 3.0: aligned with Core Location 2.0 (locn)

Proposal

To align with DCAT 3.0

Impact

Backward incompatibility with DCAT 2.0 (Literal → Class)

Cardinality constraints spatial representations (issue #175)

A dataset can have multiple *geographical coverages,* each of them can be represented in various ways:

- By preference: using a bounding box or centroid
- Alternatively using a broad range of geometry descriptions.

The cardinality constraint 0..1 on these spatial representations expresses that at most 1 representation for a single geographical coverage could be given.

However: it is possible to encode the bounding box in different serialisations WKT, GML, … using distinct CRS. This is similar to multi-language texts.

Cardinality constraints spatial representations (issue #175)

Proposal A

- Add usage note to explain that
  - Different serialisations of only the same polygon (geometry) is allowed
  - Serialisations should be CRS aware
- Lift the cardinality from 0..1 → 0..n
- Impact on the catalogue maintainer (current situation)

Proposal B

Limit the serialisations to a single representation:

- Including all characteristics (including the used CRS)
- All impact on the data publishers

Proposal C (variant of B)

Impose a set of accepted serializations with all characteristics (including the used CRS) but the publisher is free to choose.

They should provide only provide one (as they are anyhow fully equivalent and thus can be considered redundant).

- Reduced impact on the data publishers (than proposal B)
- Reduced impact on the catalogue maintainer

SHACL reformulation (issue #48,pull request #224)

Current SHACL shapes combine multiple constraints into one expression

Condensed representation

Drawbacks

- No reference to individual constraint possible
- Translations are not easy to make
  - **NOTE:** sh:message overwrites the internal message by the engine cfr https://www.w3.org/TR/shacl/#results-message.

Proposal

Current

```
:DataService_Shape
  a sh:NodeShape ;
  sh:name "Data Service"@en ;
  sh:property [
      sh:minCount 1 ;
      sh:nodeKind sh:Literal ;
      sh:path dct:title ;
      sh:severity sh:Violation
  ],
  ...
```

Proposal

```
:DataService_Shape
  a sh:NodeShape ;
  sh:name "Data Service"@en ;
  sh:property :DataService_Property_dct_title_nodekind ;
  sh:property :DataService_Property_dct_title_minCount ;
  ...

:DataService_Property_dct_title_nodekind
  sh:nodeKind sh:Literal ;
  sh:path dct:title ;
  sh:severity sh:Violation.

:DataService_Property_dct_title_minCount
  sh:minCount 1 ;
  sh:path dct:title ;
  sh:severity sh:Violation.
```

1. Named constraints
2. Each constraint is unique
3. Dereferenceable constraint names (not for next release)

SHACL: codelist validation (issue #218)

Some codelists are not published as SKOS

- Geonames,
- IANA media types
- …

Some codelists are too large to be downloaded for validation

- Geonames

Proposal

- To do a textual validation based on domain namespace rather than on membership of a Controlled Vocabulary.

Legal information (issue #209)

Observation

Many data catalogue providers do not provide quality legal information in data.europa.eu.

Proposal

To add a stronger statement about the FAIR principles, improve section 5.4 on licences with a recommendation to use the NAL http://publications.europa.eu/resource/dataset/licence if the MS does not provide guidance from itself.

Justification

- Licences are expressions of legislation, which is MS specific. MS would probably like to use their MS specific URIs rather than EU URIs.
- Unless an EU legal framework is developed that is capturing data and service licences in machine readable way, DCAT-AP can only highlight the importance, but not impose the use of a central EU NAL.

Replace controlled vocabulary for Distribution status (issue #228)

Proposal

Replace controlled vocabulary adms:status with NAL distribution-status for the Class Distribution
(was ADMS status vocabulary)

Impact

Replacement of the URIs in the data catalogues

Mitigation actions

Add a SHACL to detect the use of the adms:status concepts

dcat:themeTaxonomy (issue #207)

In current specification

- Section 5.2 contains a table with the mandatory codelists to use for some properties.
- Because dcat:theme imposes the use of the NAL data-theme, at least the value for this property is the NAL data-theme.
  - This is a value constraint and not a codelist constraint

Proposal

- Remove the value constraint on dcat:themeTaxonomy in the table of section 5.2
- Adapt the usage note for the property to:
  - *This property refers to a knowledge organization system used to classify the Catalogue's Datasets. It must have at least the value NAL:data-theme as this is the manatory controlled vocabulary for dcat:theme.*

Bytesize (issue #214)

Observation

- Issue #214 = Question on the usage of numeric xsd:types in DCAT
- DCAT 3.0 changed range from
  - xsd:decimal to xsd:nonNegativeInteger.

Proposal

To align with DCAT 3.0

Reminder: identifiers guideline (issues #223, #187, #141)

In Q2 2022, webinars on the usage of identifiers took place.

The guidelines propose a way to use and manage identifiers so that the network of harvesting catalogues can make more reliable decisions and references.

→ Proposal is ready for review

- Issue #223 → reference to guideline
- Issue #187
- Issue #141

Reminder: Cardinalities mismatch between DCAT and OWL representation in DCAT (issue #116)

Refiled as issue in W3C DCAT

Proposal accepted by W3C and closed

Proposal

To close this issue

Reminder: How to create DCAT profiles (issue #163)

Importing the dcat.ttl definition, with an explicit subclass relation of dcat:Catalog to dcat:Dataset, creates unintended SHACL validation errors.

Posted as issue in #1387 in W3C DCAT

Not resolved.

Short break (5’)

DCAT 3.0 alignment

Motivation

DCAT 3.0 is estimated to become W3C recommendation Q1 2023

- Prepare alignment with the community
- Ability to align with DCAT 3.0 as soon as possible after adoption as a W3C recommendation.
- Address Issue #203 and #155

DCAT 3.0 compared to DCAT 2.0

Addition of DatasetSeries

- DatasetSeries, first, last, prev, inSeries

Addition of versioning support

- Previousversion, hasVersion, hasCurrentVersion, version, versionNotes

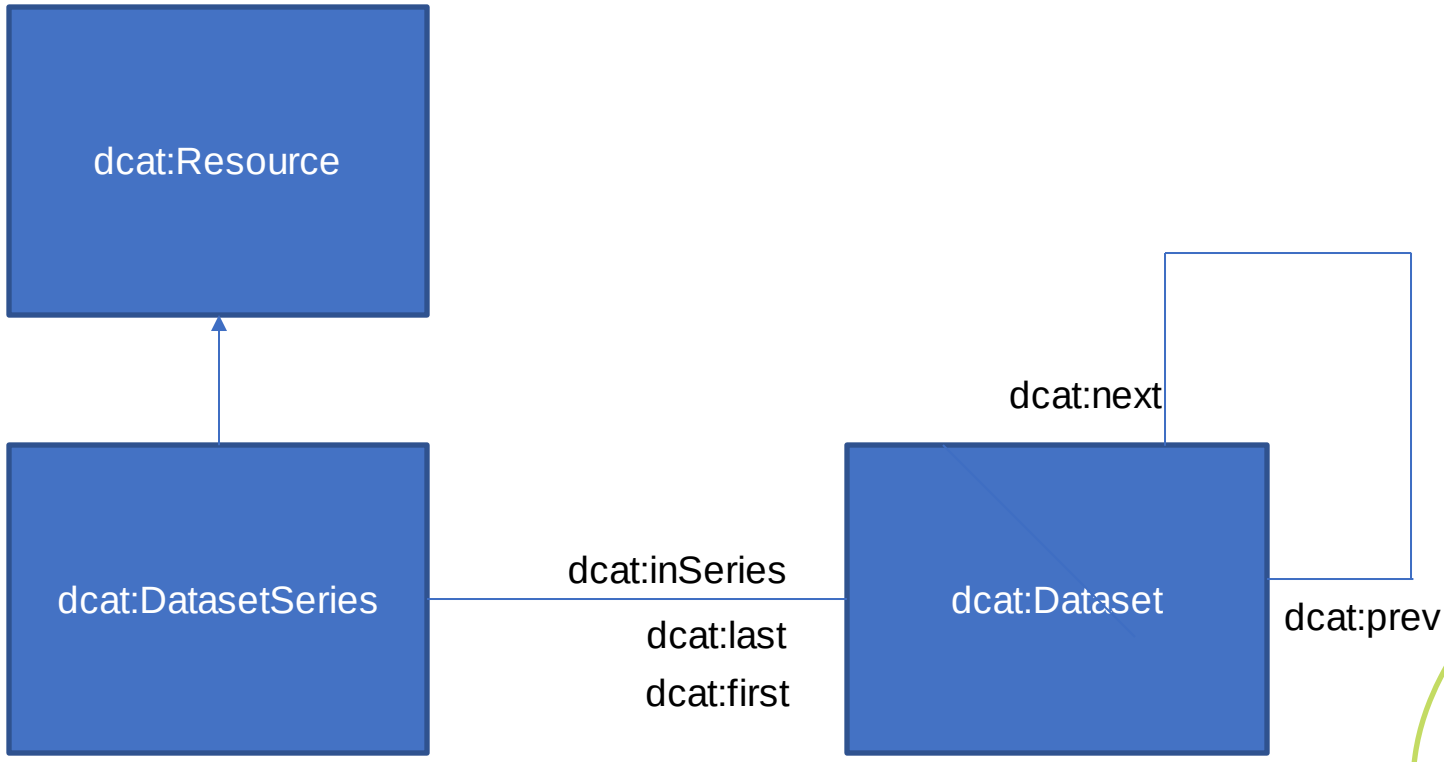
Addition of properties

- https://w3c.github.io/dxwg/dcat/#Property:catalog_resource,
- https://w3c.github.io/dxwg/dcat/#Property:resource_has_part
- https://w3c.github.io/dxwg/dcat/#Property:resource_replaces
- https://w3c.github.io/dxwg/dcat/#Property:resource_status
- https://w3c.github.io/dxwg/dcat/#Property:distribution_checksum

Addition of Checksum

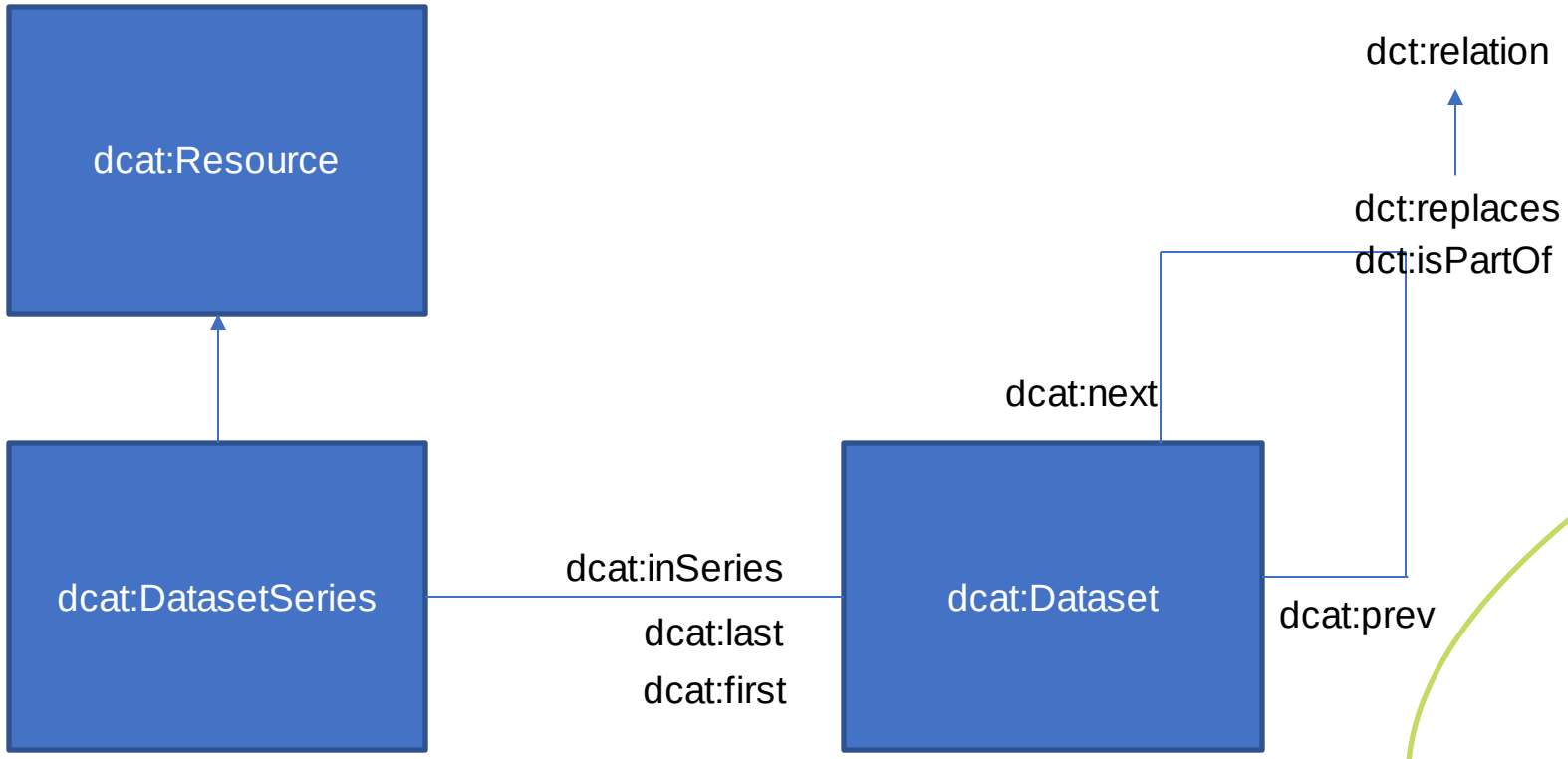
- https://w3c.github.io/dxwg/dcat/#Class:Checksum
- https://w3c.github.io/dxwg/dcat/#Property:checksum_algorithm
- https://w3c.github.io/dxwg/dcat/#Property:checksum_checksum_value

Introducing DCAT 3.0 dataset series



A collection of datasets that are published separately, but share some characteristics that group them.

Introducing DCAT 3.0 dataset series



A collection of datasets that are published separately, but share some characteristics that group them.

Adopting dataset series in DCAT-AP

Contribute #240

As it considers new terminology the **impact** on DCAT-AP is **backwards compatible** (i.e. pure extension).

Issue **#240** has been created to **collect proposals** for minimal set of properties/constraints: e.g.

- Presence of a title
- At least one element in the series
- Preferred chain is forward linking (first → next → next)

Contribute #240

Alternative methods to handle dataset series

Not the intent **to exclude alternative** methods (e.g. using the so-called soft-typing approach) but will **discourage** these stronger

→ In DCAT-AP a dataset series is an entity that is **explicitly tagged** as dcat:DatasetSeries

→ **Impact:** If data.europa.eu would visualise a dataset series in its portal it can base it on dcat:datasetSeries.

Any other approach would be considered as a collection of tightly interlinked datasets and distributions.

Contribute #241

Versioning

Property	DCAT 3.0	DCAT-AP 2.1.1	Proposal
Previous version	dcat:previousVersion		Adopt as there is no counterpart in DCAT-AP
Has current version	dcat:hasCurrentVersion		Adopt as there is no counterpart in DCAT-AP
Has version Is version of	dcat:hasVersion dcat:isVersionOf	dct:hasVersion dct:isVersionOf	• Deprecation of the old URI mapping and replace it with the new URIs. • Create supporting SHACL shape to detect deprecated URI usage
version	dcat:version	owl:versionInfo	• Deprecation of the old URI mapping and replace it with the new URIs. • Create supporting SHACL shape to detect deprecated URI usage
Version Notes	adms:versionNotes	adms:versionNotes	Adopt as is

Contribute #242

Status (adms:status)

DCAT 3.0 adopts adms:status as **property** for **dcat:Resource**

DCAT 3.0

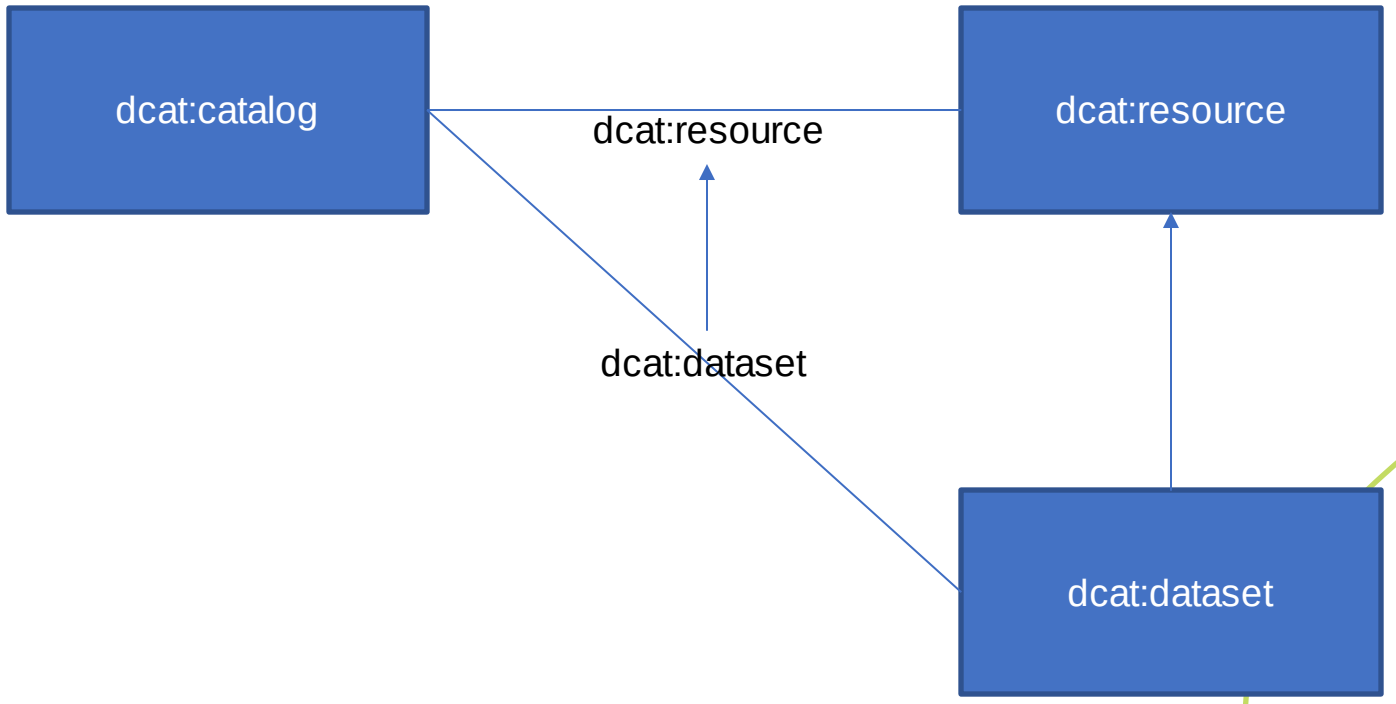
- *The status of the resource in the context of a particular workflow process [VOCAB-ADMS]. (domain dcat:Resource)*

DCAT-AP 2.1.1

- *Catalogue Record:* The status of the catalogue record in the context of editorial flow of the dataset and data service descriptions
- *Distribution:* the status of the distribution in the context of maturity lifecycle. It MUST take one of the values Completed, Deprecated, Under Development, Withdrawn.

Proposals for wider adoption can be provided in issue

Resource (dcat:resource)



This is the most general predicate for membership of a catalog. Usage of a more specific sub-property is recommended when available.

Contribute #243

Resource (dcat:resource) : impact assessment

1. Could be used to encode the requirement that a catalogue should be non-empty.
   - the SHACL must import the subclass relationship in order to guarantee correct validation

2. Unclear how to indicate the dataset series are in the catalogue
   - No specific subproperty exists
   - dcat:dataset, dcat:dataservice, dcat:catalog exists
   - Connected with issue #163 on how to support profiling of DCAT

Contribute
#244

Checksum

Contribution from DCAT-AP into DCAT.

Proposal

If any changes are made to the definitions in DCAT 3.0

→ adopt them.

Towards DCAT 3.0 alignment

summary

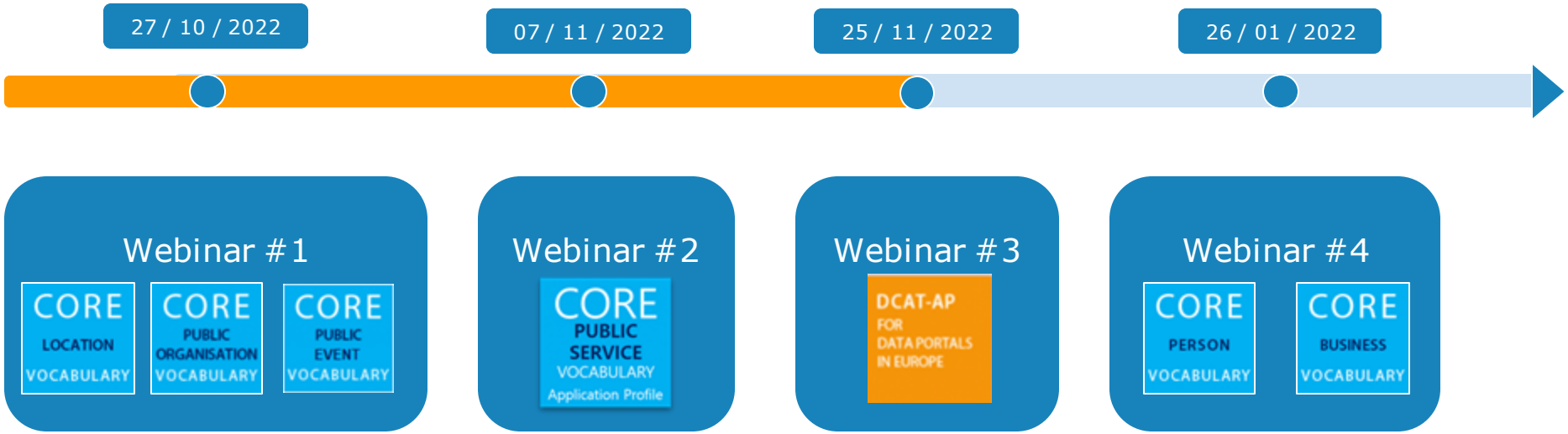
Contribute #239

- DatasetSeries can be adopted with minimal impact
- Versioning backwards incompatible changes
- Checksum and adms:status offer new possibilities
- Open discussion on how to denote the content of the catalogue

All your contributions are welcome on GitHub (overview issue #239)

Wrap-up and next steps

Next steps



In the meantime, you are invited to share your feedback on the issues on [GitHub.]


Thank you

interoperable europe

innovation ∞ govtech ∞ community

Stay in touch

(@InteroperableEU) / Twitter

Interoperable Europe - YouTube

Interoperable Europe | LinkedIn

DIGIT-INTEROPERABILITY@ec.europa.eu

https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe

European Commission

25 NOVEMER 2022

Informal meeting on HVD

DIGIT.D2 - Interoperability.

interoperable europe

Contribute #231

Supporting HVD directive

summary (policy side)

- Directive (EU) 2019/1024
  - Implementing Act is being finalised: expected adoption Q1 2023
  - Prepared in collaboration with the MS policy officers (PSI working group)

- (High level) a HVD dataset should
  - be described in metadata
  - have a downloadable distribution
  - have an API
  - information must be present according to some schema.

- MS should report on the progress

Supporting HVD directive

summary (DCAT-AP side)

Contribute #231

- Provide means to indicate that a dataset is a HVD dataset
  - Using dcat:theme/dct:type to indicate using a concept from a controlled vocabulary managed by PO.
- Provide means to indicate that a dataset adheres to a schema
  - Use dct:conformsTo