



# DCAT-AP

## FOR DATA PORTALS IN EUROPE

**BRECHT WYNS**  
PWC EU SERVICES

**AUTHORS: BRECHT WYNS, MAKX DEKKERS, NIKOLAOS  
LOUTAS, VASSILIOS PERISTERAS, ATHANASIOS  
KARALOPOULOS**

# TOPICS OF TODAY

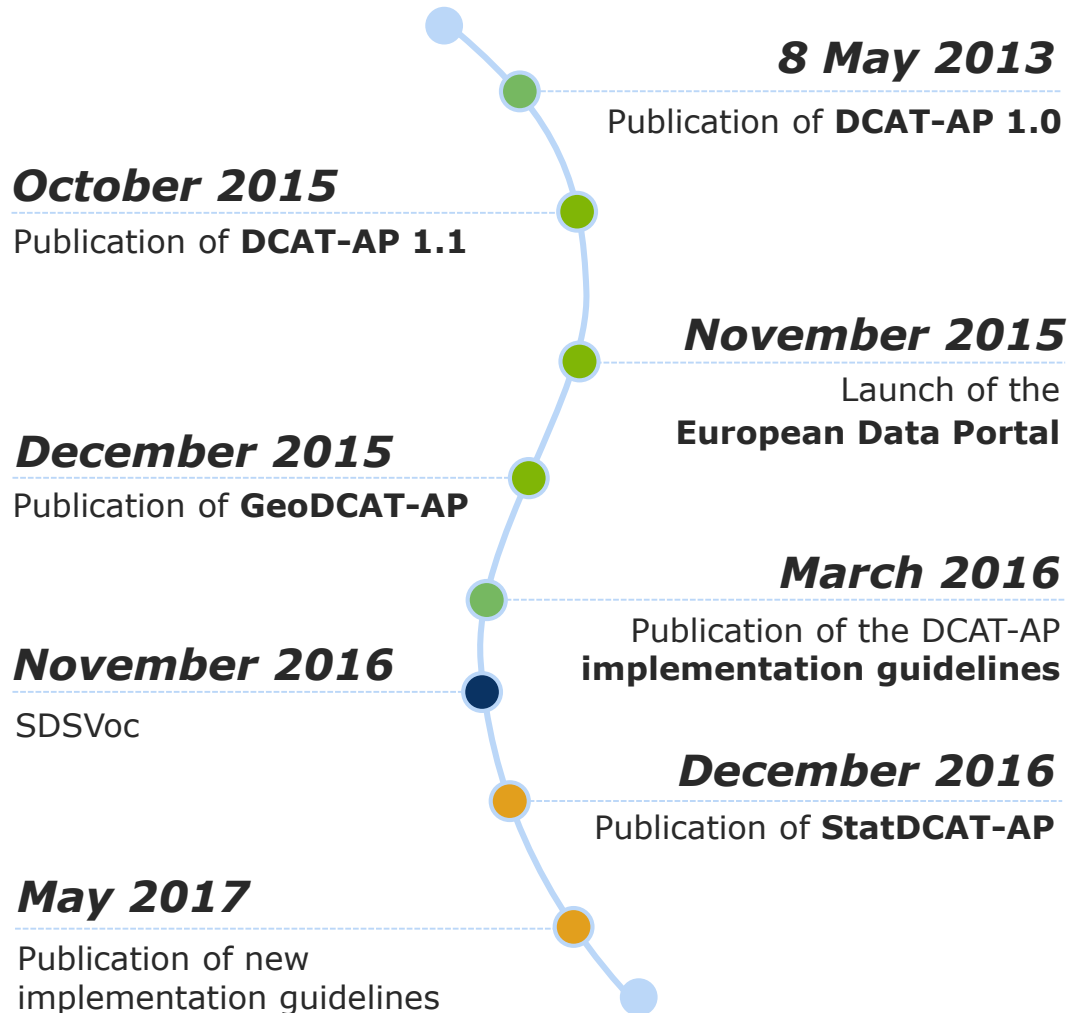


1. Introduction
2. Implementations
3. Extensions
4. Issues with DCAT(-AP)
5. Get in touch

# INTRODUCTION TO DCAT-AP

- Describing **public sector datasets** in Europe
- Enable **cross data-portal search**
- Improve discoverability **across borders** and **sectors**
- Working group of **experts**
- **Open** collaborative process
- Based on DCAT
  - Mandatory, recommended and optional **classes & properties**
  - **Controlled vocabularies**

# DCAT-AP IS AN ONGOING INITIATIVE



## SUCCESSFUL IMPLEMENTATIONS

- **15+** data portals
- **Local extensions:**  
e.g. DCAT-AP\_IT, DCAT-AP-NO, DCAT-AP-NL
- **Tools:** validators, harvesters and exporters  
Available via [Joinup](#)



## TWO EXTENSIONS: GEODCAT-AP & STATDCAT-AP

### GeoDCAT-AP

- An **extension** to DCAT-AP for **geospatial metadata**
- Take into account the **standards** currently used in the geospatial community: INSPIRE & ISO19115

### StatDCAT-AP

- Fully conformant **extension** of DCAT-AP for the exchange of metadata for statistical datasets
  - Describing **statistical data standards** in the DCAT-AP format
  - **extending the DCAT-AP** with additional descriptive elements for statistical data

## ISSUES WITH DCAT(-AP)



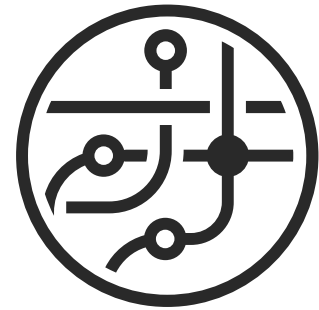
1

ISSUES FOR GUIDELINES & DCAT-AP



2

ISSUES FOR 3<sup>rd</sup> PARTIES



3

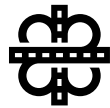
ISSUES FOR DCAT

# ISSUES WITH DCAT(-AP) GUIDELINES & DCAT-AP

Existing guidelines are available via <https://joinup.ec.europa.eu/node/148075/>



MDR Data Themes



Extending DCAT-AP



Licence documents



Provenance



Contact information



Identifiers for datasets & distributions



Dataset series



Duplicates



DCAT-AP Tools



## ISSUES WITH DCAT(-AP) GUIDELINES & DCAT-AP

Topics for additional guidelines or updating DCAT-AP, e.g:

- Contradiction dct:spatial and dct:Location
- De-referencing vocabularies
- URIs for organisations
- Agent roles
- Expressing data quality
- ...
- Organisational & legal challenges

Share your input via [https://joinup.ec.europa.eu/asset/dcat-ap\\_implementation\\_guidelines/issue/all](https://joinup.ec.europa.eu/asset/dcat-ap_implementation_guidelines/issue/all)

## ISSUES WITH DCAT(-AP) DCAT

- **Relationships between Datasets**, incl. versioning, time sequence, parent/child and grouping of collections: use of relation types
- **Rights and licences** for datasets: relationship with licences on catalogue and distributions
- **Service-based data access**: modelling of non-file distributions and set of properties to enable machine-processing
- **Relationship between Distributions**: similarity criteria
- **Packaging of distribution files**: expression of format of included files
- **Scientific data** and **data citation**

# RELATIONSHIPS BETWEEN DATASETS



## Evolutionary relation

**Versioning.** Each new version has newer data and replaces the precedent one, which turns obsolete and remains there for reference only



## Time series

**Sequence of points** in time observations



## Datasets with **parts**

**Parent** datasets with multiple **child** sub-datasets



## Collections

**Grouping** datasets under one umbrella, based on different dimensions, e.g. geography, or use cases

# RIGHTS AND LICENCES

- In the DCAT model, rights and licences are assigned to **catalogues** and to **distributions**, not to datasets.
- In actual implementations, rights and licences may be associated with the **dataset**, applying to all distributions of the dataset.

## SERVICE BASED DATA ACCESS

- Many datasets are not published as files, but can be accessed through APIs or SPARQL endpoints
- Definition of Distribution in DCAT mentions that “Examples of distributions include a downloadable CSV file, an **API** or an RSS feed”.
- DCAT only seems to focus on **files**, for example by defining format and media type which are not relevant for APIs or end points.
- Specific information is necessary to access APIs and end points, e.g. methods and schemas, and the current version of DCAT does not include properties to express those types of information.

## RELATIONSHIPS FOR DISTRIBUTIONS

- DCAT definition of Distribution is ambiguous  
*Represents a specific available form of a dataset. Each dataset might be available in different forms, these forms might represent different formats of the dataset or different endpoints. Examples of distributions include a downloadable CSV file, an API or an RSS feed*
- Do all distributions contain the same data?
- May distributions contain different slices of a dataset?  
E.g. files for individual years in a multi-year dataset.
- Need to develop clear criteria to determine whether two data files or feeds can be distributions of a single dataset or of different datasets.

## PACKAGING OF DISTRIBUTION FILES

- Distributions are often made available in a packaged or compressed format  
E.g. a group of XLS files packaged in a ZIP file, compression of large files
- DCAT requires the package format to be expressed in dct:format
- It might be useful for an application to know what formats are contained in the package
- As a consequence, it might be useful if DCAT considered ways to indicate various levels of packaging. An example of an approach is in the way ADMS defines Representation Technique (see <https://www.w3.org/TR/vocab-adms/#representation-technique>).

## SCIENTIFIC DATA & DATA CITATION

In order to meet the requirements of the JRC Data Catalogue (<http://data.jrc.ec.europa.eu/>), which include support to data citation, JRC developed an extension of DCAT-AP for research data, covering the following requirements:

- a) ability to indicate dataset **authors**
- b) ability to describe data **lineage**
- c) ability to give potential data consumers information on how to **use the data** ("usage notes")
- d) ability to link to scientific **publications** about a dataset
- e) ability to link to **input** data (i.e., data used to create a dataset)



# CONTACT US

**Brecht Wyns**

PwC EU Services  
[brecht.wyns@be.pwc.com](mailto:brecht.wyns@be.pwc.com)

**Makx Dekkers**

AMI Consult  
[makx@makxdekkers.com](mailto:makx@makxdekkers.com)



**Nikolaos Loutas**

PwC EU Services  
[nikolaos.loutas@be.pwc.com](mailto:nikolaos.loutas@be.pwc.com)

**Athanasios Karalopoulos**

European Commission – ISA<sup>2</sup> Programme  
[athanasios.karalopoulos@ec.europa.eu](mailto:athanasios.karalopoulos@ec.europa.eu)



Promoting semantic interoperability in Europe

# STAY CONNECTED!

## PROJECT OFFICERS

Vassilios.Peristeras@ec.europa.eu

Suzanne.Wigard@ec.europa.eu

Athanasios.Karalopoulos@ec.europa.eu

## GET INVOLVED

- Follow @SEMICeu on Twitter
- Join the SEMIC group on LinkedIn
- Join the SEMIC community on Joinup

## 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