



ENVRI-FAIR Task Force 1

Service catalogue tutorial session

Rossana Paciello¹ (rossana.paciello@ingv.it), **Daniele Bailo**¹ (daniele.bailo@ingv.it), **Valerio Vinciarelli**² (Valerio.vinciarelli@ingv.it)

¹ EPOS ERIC Executive Coordination Office – IT Unit

² Istituto Nazionale di Geofisica e Vulcanologia (INGV)



ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068



Prerequisites for attending this tutorial

RDF

- <https://www.w3.org/RDF/>
- <https://www.w3.org/TR/turtle/>
- <https://www.w3.org/TR/shacl/>

DCAT-AP

- <https://www.w3.org/TR/vocab-dcat-2/>
- <https://www.w3.org/2017/dxwg/>
- <https://github.com/epos-eu/EPOS-DCAT-AP>

METADATA EDITOR

- <https://epos-eu.github.io/DCAT-Metadata-Editor/>

CERIF

- <https://eurocris.org/services/main-features-cerif>



Agenda

- Proposal for an ENVRI catalogue of services
- DCAT-AP overview
- EPOS-DCAT-AP extension
- DCAT Metadata Editor – Live demo
- ENVRI-FAIR service catalogue demo



Objectives and outcomes from TF1 work

- Creation of a catalogue of services within ENVRI-FAIR
- Catalogue should be EOSC-ready
- Catalogue should not map all metadata, but describe services that provide access to data

Generic RI architecture

Type of services, metadata, and data formats from the RIs:

Services

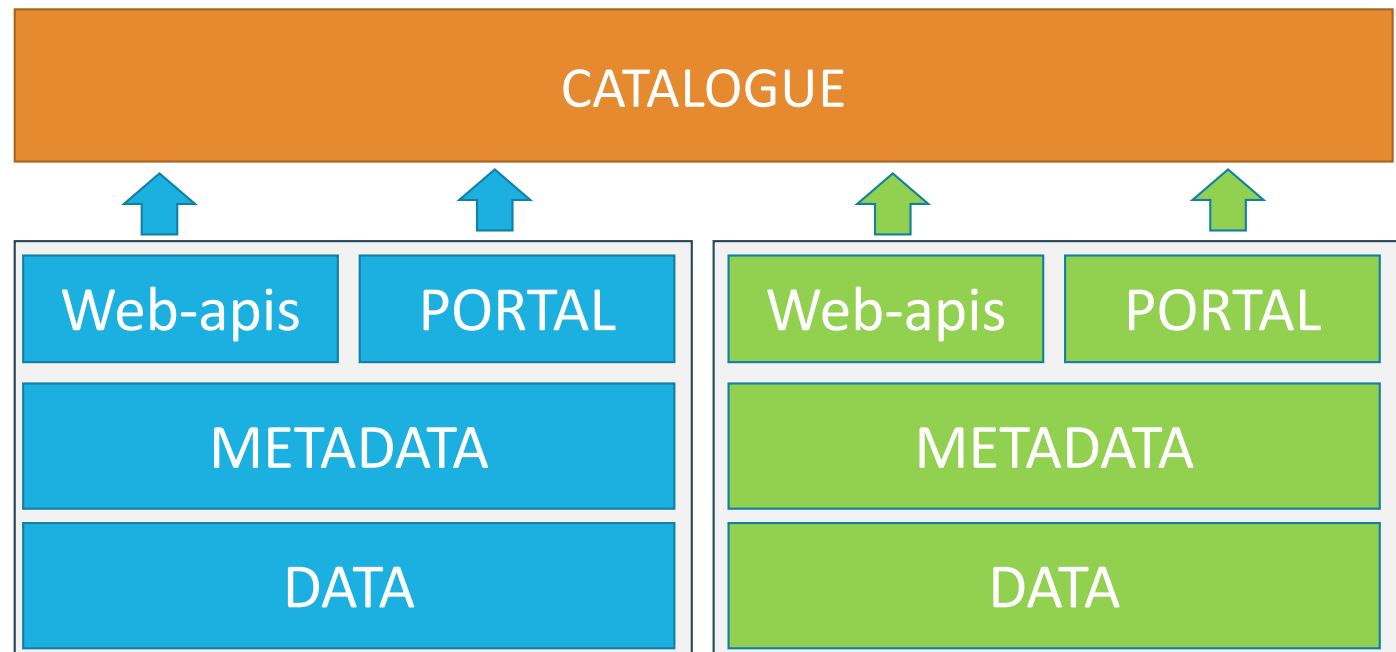
- RESTFUL WEBAPIS
- ERDDAP services
- Opendap

Metadata

- OAI-PMH ISO 19139
- Geonetwork (restful apis+oai-pmh)

Data

- DATA (csv, JSON, netcdf, ASCII, nasa-ames)
- METADATA



Proof of Concept for an ENVRI CATALOGUE OF SERVICES

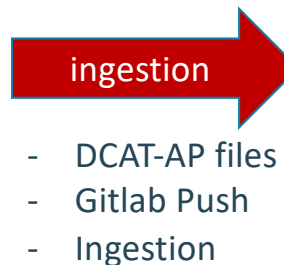
RI Services

- Examples:
- Standards (OGC)
 - RESTful
 - De facto standards
 - ...others

Metadata description

Description of services in DCAT-AP extension

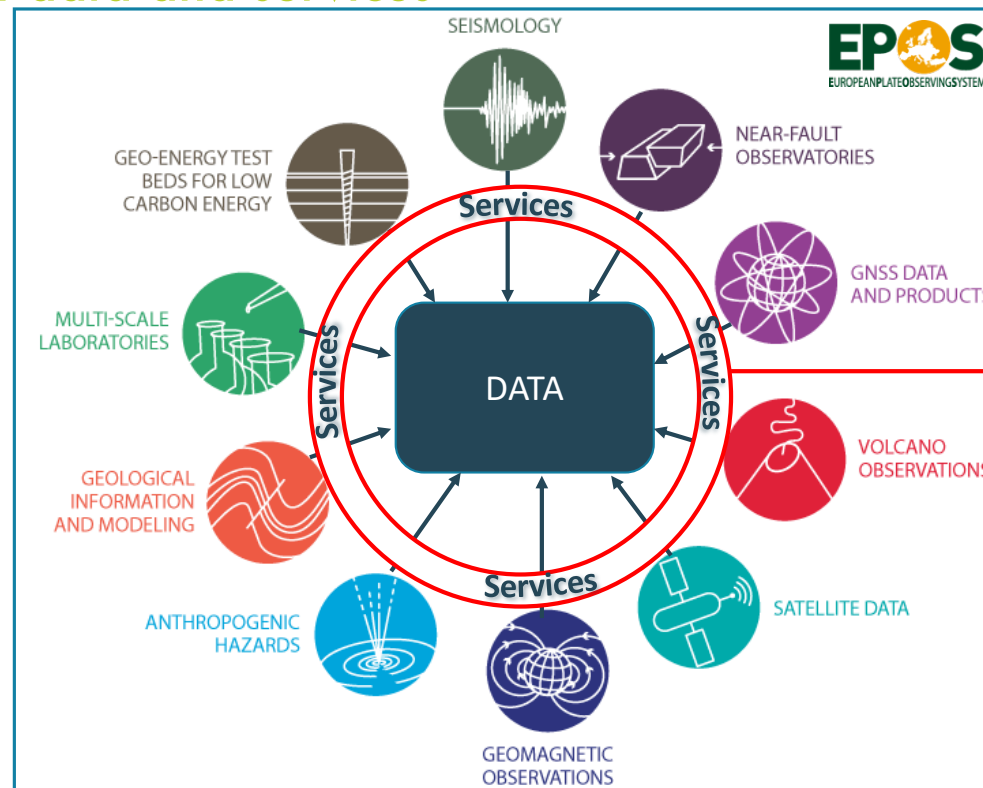
ENVI-FAIR catalogue



Services from RIs listed into the ENVRI-FAIR catalogue

RIs in solid-Earth subdomain

Integration of data and services



- OGC Web Services
- RESTful APIs
- Open Search APIs
- Community standards

How to describe **data** and **services**?

DCAT APPLICATION PROFILE

DCAT-AP

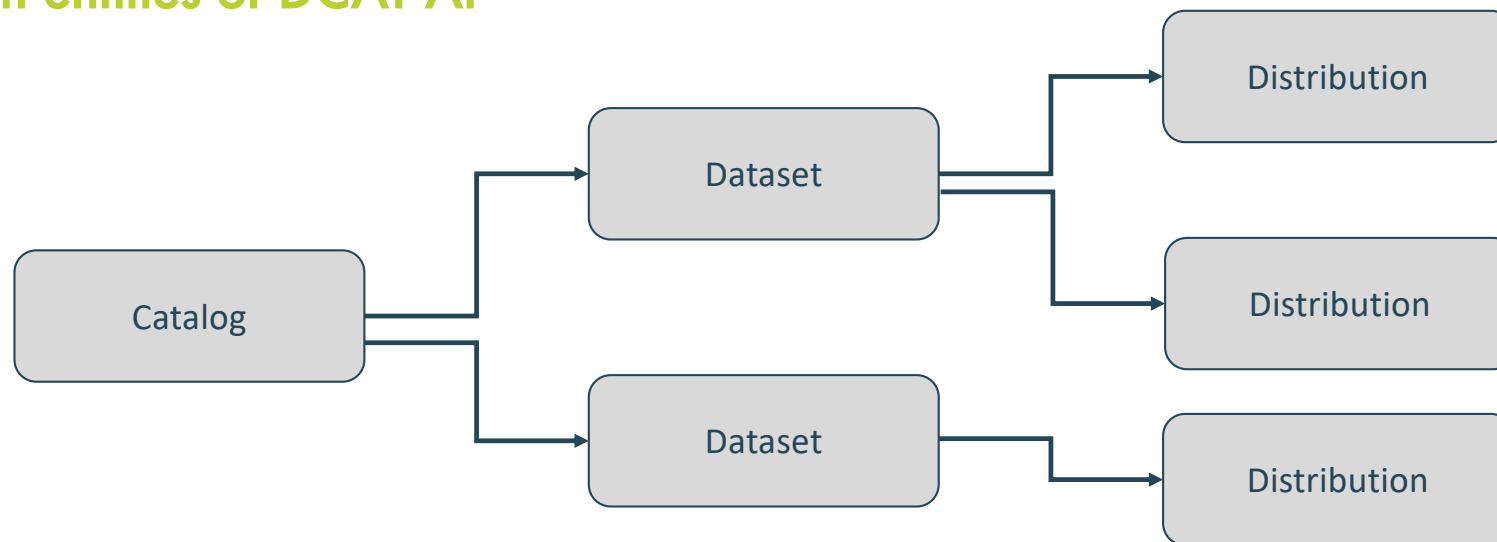
- Specification for data portals in Europe based on the Data Catalogue Vocabulary (DCAT) developed by W3C.
- It aims to facilitate data findability, cross-reference and interoperability between data catalogues on the web.

Why using DCAT-AP?

- The data model is represented in RDF supporting the principles of Linked Open Data (LOD)
- Not meant to replace domain/community specific standards, but to make them interoperable
- Easily extendable through Application Profiles

DCAT-AP

Main entities of DCAT-AP



A Catalog contains one or more Datasets

A Dataset has one or more Distributions

Resource Description Framework (RDF)

- A language for representing information about web resources
- RDF represents information using set of **Triples**
- **Triple** in RDF is analogous a sentence in natural language (NL)
 - Every NL sentence contains 3 parts: subject, predicate and object
 - Every **Triple** in RDF contains *<subject, predicate, object>*
 - **Subject** – a resource which may be identified with a URI
 - **Predicate** – a URI-identified reused specification of the relationship
 - **Object** – a resource or literal to which the subject is related

Resource Description Framework (RDF)

EXAMPLE

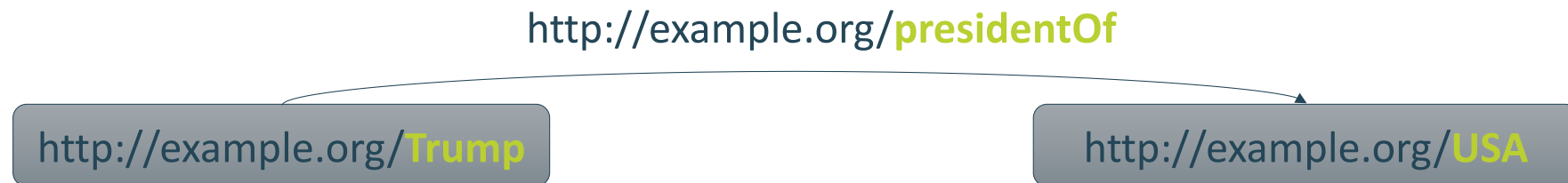
Natural Language Sentence: *Trump is the president of USA*

RDF Triple:

<http://example.org/Trump> <http://example.org/presidentOf> <http://example.org/USA>
subject *predicate* *object*

RDF as a Graph

- Subjects and objects are nodes of the graph
- Predicates form the connecting arcs



RDF serialization formats

- An RDF graph can be serialized in different formats:
 - RDF/XML
 - N-Triples
 - RDF/Turtle
 - Notation 3 (N3)

- Examples:

RDF/XML format:

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:ex="http://example.org/">
<rdf:Description rdf:about="http://example.org/Trump">
  <ex:presidentOf rdf:resource="http://example.org/USA" />
</rdf:Description>
```

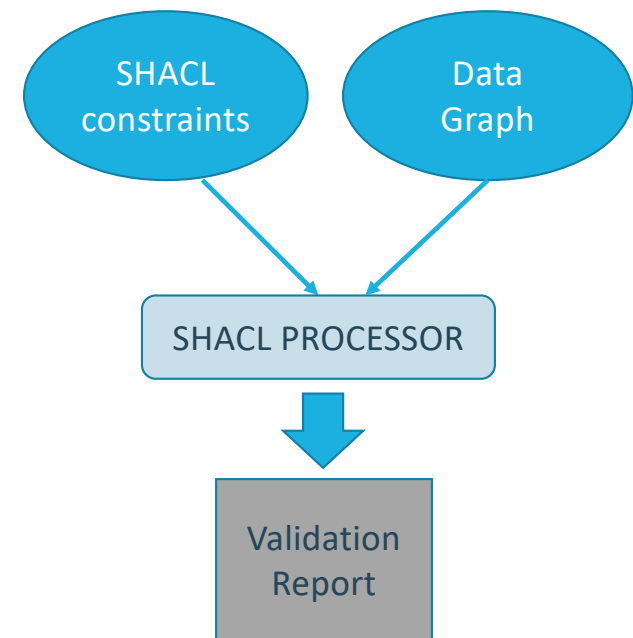
RDF/Turtle format:

```
@prefix ex: <http://example.org/> .
<http://example.org/Trump> ex:presidentOf ex:USA .
```

Shapes Constraint Language

SHACL

- A W3C recommendation
- Language for describing and validating RDF data graphs against a set of constraints
- These constraints are provided as **shapes**
- Shapes are used to define **classes** together with constraints on their **properties** such as:
 - cardinality (minCount and maxCount);
 - value type and allowed values;
 - more complex kinds of constraints.



Shapes Constraint Language

EXAMPLE OF SHACL VALIDATION

SHAPE

```
:DatasetShape a sh:NodeShape ;  
  sh:targetClass dcat:Dataset ;  
  sh:property [  
    sh:path dct:title ;  
    sh:datatype xsd:string ;  
    sh:minCount 1 ;  
  ] ;  
  sh:property [  
    sh:path dct:description ;  
    sh:datatype xsd:string ;  
    sh:minCount 1 ;  
    sh:maxCount 1 ;  
  ]  
.
```

PROPERTIES

DATA GRAPH

```
<https://www.epos-eu.org/Seismology/Dataset> a  
dcat:Dataset ;  
  dct:title 123 ; ✗  
  dct:description "Continuous seismic waveforms" ; ✓  
.
```

DATA GRAPH

```
<https://www.epos-eu.org/Seismology/Dataset> a  
dcat:Dataset ;  
  dct:title "Seismic Waveform Data" ; ✓  
  dct:description "Continuous seismic waveforms" ; ✓  
  dct:description "Description of Dataset" ; ✗  
.
```

DCAT-AP

Versions History

Version 1.0

(1) Identified the essential elements and attributes of DCAT in the European context; (2) Identified the controlled vocabularies to be used in the European context; (3) Identified the strict minimum description metadata to be exchanged between data portals in Europe.

Version 1.1

Revision of version 1.0 taking into account the following criteria: (1) Improve the discovery of datasets; (2) Ensure compliance with DCAT; (3) Ensure simplicity: keep a narrow scope; (4) Ensure application domain neutrality.

Version 1.2

Update of version 1.1 based on requests for change coming from real-world implementations of the specification.

Version 1.2.1

Update of version 1.2 based on requests for change coming from real-world implementations of the specification.

Version 2.0

The changes implemented in this version are related to the following issues: (1) Alignment with W3C DCAT 2.0 which introduces the ability to share data services; (2) In order to steer towards quality metadata descriptions, the implementation of a number of properties are recommended or made mandatory; (3) Removal of non-existing controlled vocabularies; (4) The alignment with W3C DCAT 2.0 resulted in the replacement of the URIs with respect to the last DCAT-AP release 1.2.1.

Version 2.1

This version implements bug fix changes to the version 2.0, addressing the comments received via GitHub. The main updates cover minor editorial changes.

September
2013

September
2017

November
2018

May
2019

November
2019

June
2020

DCAT-AP

Extensions

- ISA Programme has published guidelines on how to create DCAT-AP extensions.
- An extension needs to meet the minimum requirements described in DCAT-AP specification.
- Several extensions of DCAT-AP have been produced to address different requirements.

Application Profiles

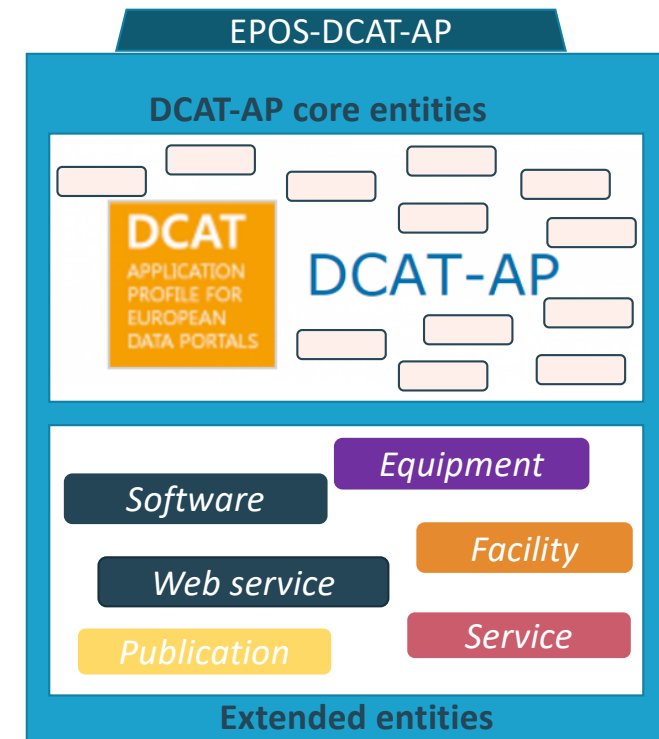
source: <https://www.w3.org/2017/dxwg/>

for DCAT

- [DCAT-AP](#): DCAT application profile for data portals in Europe
- [DCAT-AP SHACL](#): DCAT-AP SHACL expression
- [GeoDCAT-AP](#): Geospatial extension to DCAT-AP
- [StatDCAT-AP](#): Statistical extension to DCAT-AP
- [DCAT-BE](#): Belgian extension to DCAT-AP
- [DCAT-AP.de](#): German extension to DCAT-AP
- [DCAT-AP \(Ireland\)](#): Irish extension to DCAT-AP
- [DCAT-AP_IT](#): Italian extension to DCAT-AP
- [GeoDCAT-AP_IT](#): Italian extension to GeoDCAT-AP
- [DCAT-AP-NL](#): Dutch extension to DCAT-AP
- [DCAT-AP-DONL](#): Dutch extension to DCAT-AP used in <https://data.overheid.nl/>
- [DCAT-AP-NO](#): Norwegian extension to DCAT-AP
- [DCAT-AP-SE](#): Swedish extension to DCAT-AP
- [DCAT-AP-DK](#): Danish extension to DCAT-AP
- [Czech DCAT-AP 2.0.1 implementation](#)
- [CH-DCAT-AP](#): Swiss extension to DCAT-AP
- [European TransportDCAT-AP](#): Extension to DCAT-AP concerning the transport domain
- [Report on DCAT-AP use](#)
- [Summary about the Report on DCAT-AP use](#)
- [Analysis of the DCAT-AP Extensions](#)
- [Another analysis of DCAT-AP variants](#)
- [EPOS-DCAT-AP](#): European Plate Observing System DCAT-AP
- [DCAT-AP-JRC](#): Extension to DCAT-AP for multi-disciplinary research data

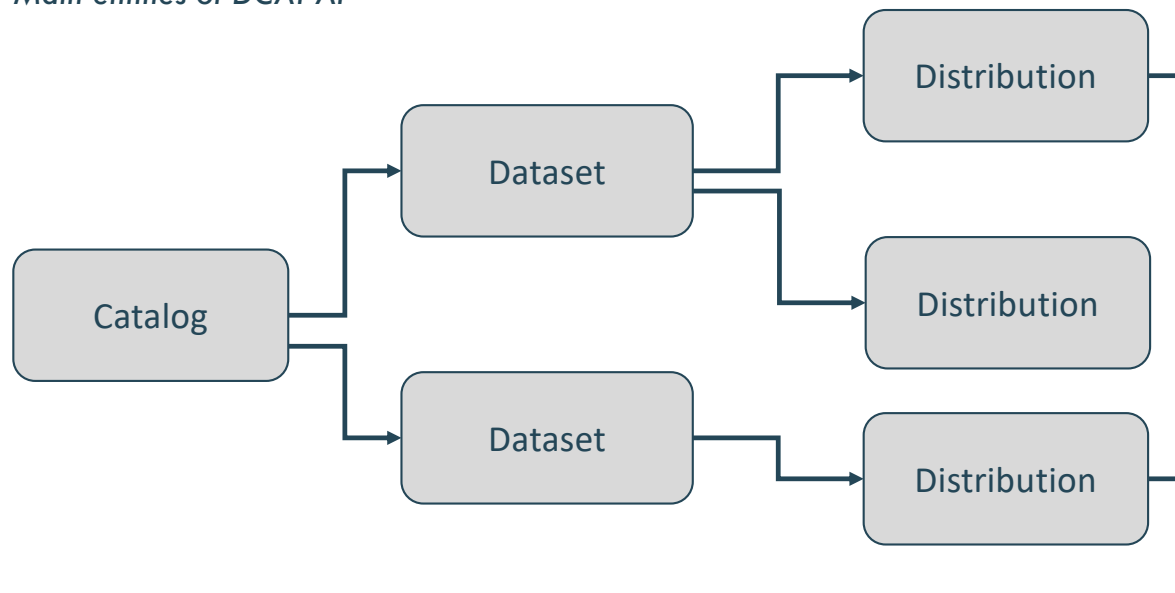
EPOS-DCAT-AP

- An extension of the DCAT-AP for Research Infrastructures in the solid-Earth domain.
- It extends DCAT-AP by adding a set of high level Core Concepts that covers the main assets and resources of the EPOS communities.
- The extension is the result of a collaborative work that brought together diverse expertise within multidisciplinary context in order to create a shared information space and enable semantic interoperability.

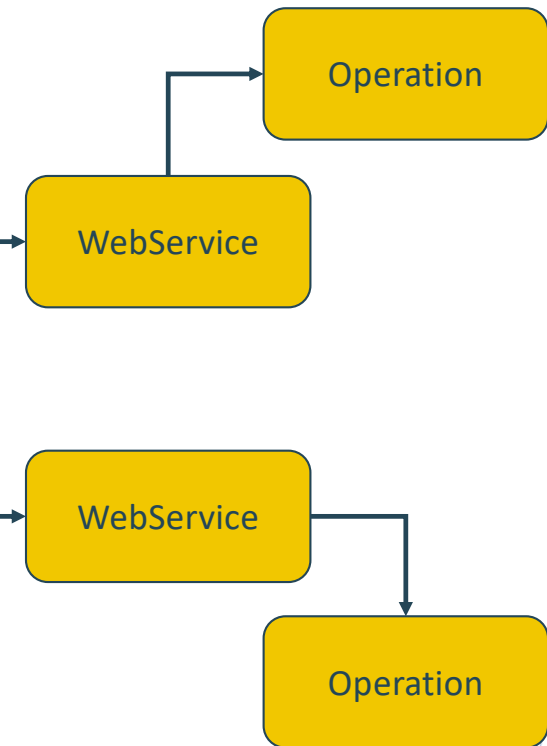


EPOS-DCAT-AP

Main entities of DCAT-AP



Main entities of EPOS-DCAT-AP for representing APIs for the programmatic access to datasets.



EPOS-DCAT-AP

Example of RDF/Turtle representation

DATASET

```
<https://www.epos-eu.org/Seismology/Dataset/001> a
dcat:Dataset ;
  dct:title "Primary Seismic Waveform Data" ;
  dct:created "2016-01-01T00:00:00Z"^^xsd:dateTime ;
  dct:type
"http://purl.org/dc/dcmitype/Collection"^^xsd:anyURI ;
  dct:publisher <PIC:000518944> ;
  dcat:distribution <https://www.epos-
eu.org/Seismology/Distribution/001> ;
  ...
.
```

DISTRIBUTION

```
<https://www.epos-eu.org/Seismology/Distribution/001> a
dcat:Distribution ;
  dct:title "Title of Seismic Waveform Distribution" ;
  dct:description "Description of Seismic Waveform Distribution" ;
  dct:type
"http://publications.europa.eu/resource/authority/distribution-
type/WEB_SERVICE"^^xsd:anyURI ;
  dct:conformsTo <https://www.epos-
eu.org/Seismology/WebService/001> ;
  dcat:accessURL <https://www.epos-eu.org/Seismology/Operation/001> ;
  ...
.
```

WEBSERVICE






```
<https://www.epos-eu.org/Seismology/WebService/001> a
epos:WebService ;
  schema:name "FDSN Dataselct - ORFEUS Data Center" ;
  schema:description "FDSN Standard webservice to download
waveform data" ;
  dcat:contactPoint <http://orcid.org/0000-0002-6250-0000/> ;
  schema:datePublished "2016-01-01T00:00:00Z"^^xsd:dateTime ;
  hydra:entrypoint "http://www.orfeus-
eu.org/fdsnws/dataselct/1/application.wadl"^^xsd:anyURI ;
  ...
.
```

OPERATION

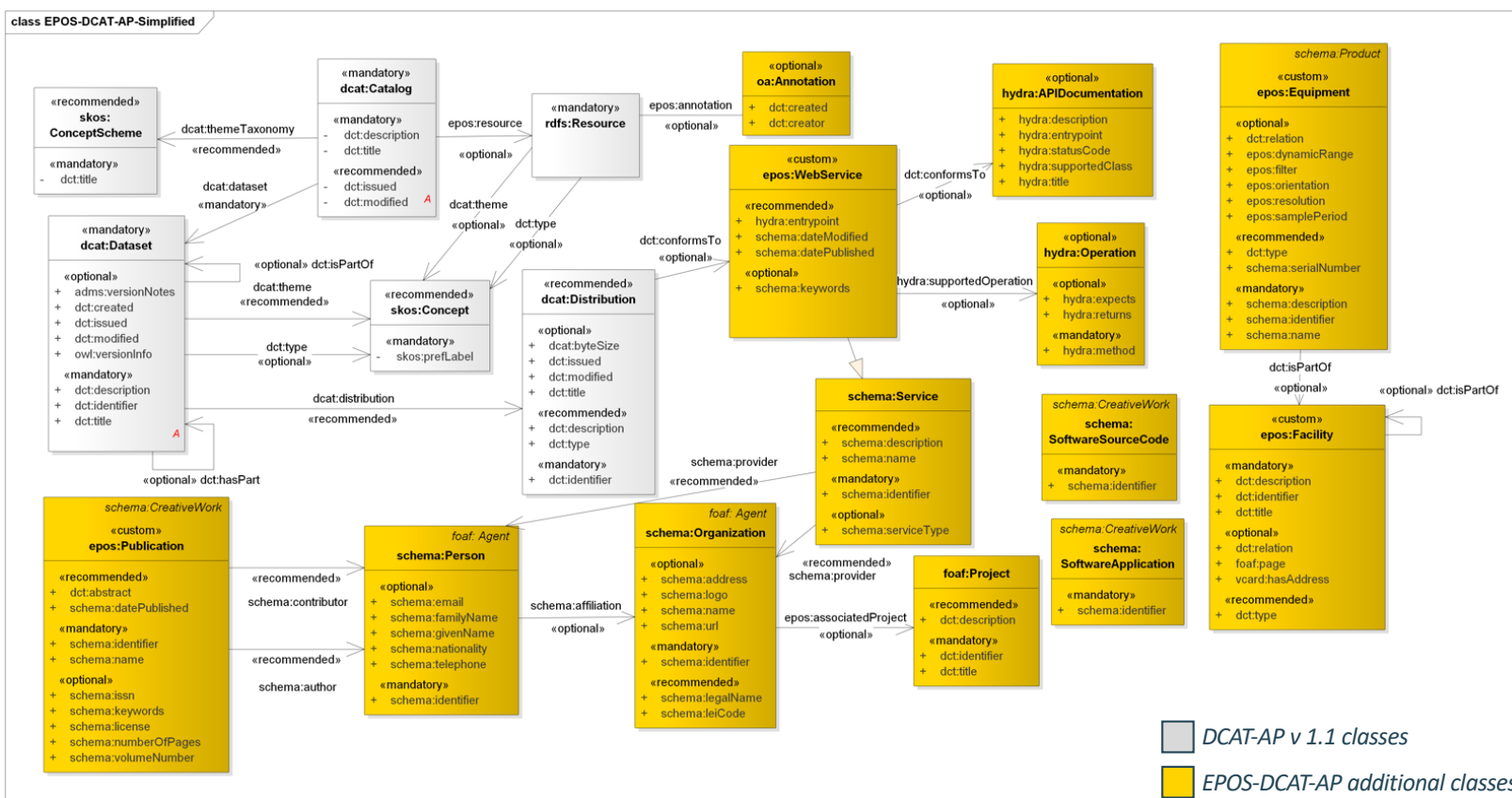
```
<https://www.epos-eu.org/Seismology/Operation/001> a
hydra:Operation ;
  hydra:method "GET" ;
  hydra:returns "application/vnd.fdsn.mseed" ;
  hydra:property[ a hydra:IriTemplate ;
    hydra:template "http://www.orfeus-
eu.org/fdsnws/dataselct/1/query/{?starttime,endtime,network,quality}"^^xsd:string ;
    hydra:mapping[ a hydra:IriTemplateMapping ;
      hydra:variable "starttime" ;
      rdfs:label "Start of the timespan" ;
      schema:valuePattern "YYYY-MM-DDThh:mm:ss" ;
      hydra:required "true"^^xsd:boolean ;
      rdfs:range "xsd:dateTime" ;
      schema:minValue "2012-01-01T00:00:00" ;
      schema:maxValue "2017-12-01T00:00:00" ;
    ] ;
  ...
.
```

EPOS-DCAT-AP

Versions History

Version 0.9 Extension of the DCAT-AP v 1.1 for EPOS context taking into account the following criteria: (1) Identified the essential elements and attributes of EPOS-DCAT-AP in the EPOS context; (2) Identified the controlled vocabularies to be used in the EPOS context; (3) Identified the strict minimum description metadata to be exchanged between EPOS communities.	 February 2018
Version 0.12 Revision of previous version taking into account the following criteria: (1) Ensure compliance with DCAT-AP v 1.1; (2) update EPOS namespace.	 July 2018
Version 0.14 Revision of previous version taking into account the following criteria: (1) Ensure compliance with DCAT-AP v 1.1; (2) In order to steer towards quality metadata descriptions, the implementation of a number of properties are recommended or made mandatory.	 February 2019
Version 0.15 The changes implemented in this version are related to the Software entities as well as requests coming from real-world implementations of the specification.	 June 2019
Version 1.0 This version implements bug fix changes to the previous version 2.0 and cover changes related to the Equipment and Facility entities.	 September 2020

EPOS-DCAT-AP UML CLASS DIAGRAM



EPOS-DCAT-AP

- EPOS-DCAT-AP follows the recommendation on DCAT-AP extensions and addresses the following concerns:
 - Extending the data model with additional concepts
 - *E.g. WebService, Equipment, Facility, Software*
 - Introducing new relationships and roles
 - Reusing elements of popular vocabularies
 - *Schema.org and the Hydra Core Vocabulary to represent evolvable Web APIs*
 - Describing APIs for the programmatic access to datasets

dcat:Distribution $\xrightarrow{dct:conformsTo}$ *epos:WebService*
dcat:Distribution $\xrightarrow{dcat:accessURL}$ *hydra:Operation*

- Enabling integrity checks and validation (via SHACL)

EPOS-DCAT-AP

- Most of the gaps identified by EPOS-DCAT-AP extension have been addressed in DCAT-AP version 2 such as introducing Data Services entity in order to represent API that provide access to one or more datasets.
- Currently EPOS-DCAT-AP achieved a level of maturity which makes it suitable to be considered by other Research Infrastructures and to be applied in different contexts.
- The latest version of the EPOS-DCAT-AP is available on GitHub and it includes a UML diagram, ontology definition, examples and more details. (<https://github.com/epos-eu/EPOS-DCAT-AP>)

Metadata Provision

RI Services

Metadata
description

ENVI-FAIR
catalogue

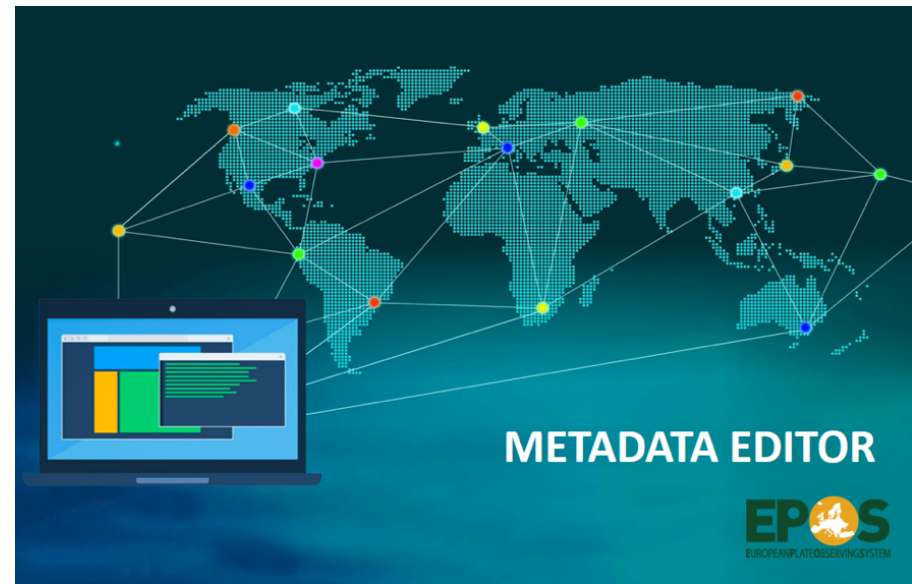
Quite demanding manual process!



DCAT Metadata Editor

LIVE DEMO >> 

- A Java Desktop Application released in **Beta version** on September 21, 2020.
- It provides a user-friendly Interface based on a graph representation for populating metadata files.
- It is a SHACL driven Metadata Editor which allows users to populate RDF data graphs by importing the shapes (classes with constraints) defined in the SHACL file of a domain-specific DCAT-AP extension.
- It automatically validates RDF data graphs against SHACL constraints.
- RDF data graphs are serialized and exported in RDF/Turtle format.



<https://epos-eu.github.io/DCAT-Metadata-Editor>

DCAT Metadata Editor

Graph View

Palette View

RDF/TURTLE View

Outline View

Properties View

The screenshot displays the EPOS Metadata Editor interface with the following components:

- Palette View:** A sidebar on the left containing a list of metadata classes such as Address, ApiDocumentation, Catalog, CatalogRecord, Concept, ConceptScheme, ContactPoint, ControlAction, CreativeWork, Dataset, Distribution, Equipment, Facility, Identifier, InTemplate, InTemplateMapping, Location, Operation, Organization, and PeriodOfTime.
- Graph View:** A central area showing a network graph of relationships between classes like Person, ContactPoint, Location, Concept, ConceptScheme, Dataset, Distribution, Organization, and PeriodOfTime.
- Outline View:** A sidebar on the bottom left showing a hierarchical tree of nodes, including Concept, ConceptScheme, ContactPoint, Dataset, Distribution, Location, Organization, PeriodOfTime, and Person.
- Properties View:** A central area on the bottom right showing the properties of a selected 'Dataset' instance, including mandatory properties like dct:title, dct:description, and dct:identifier, and recommended properties like dcat:theme, publisher, and dcat:keyword.
- RDF/TURTLE View:** A sidebar on the right displaying the RDF/Turtle representation of the metadata, showing prefixes and triples.



DCAT Metadata Editor

GUI Overview

- **Graph View** – a graph representation of the metadata elements connected as a graph. The Graph is interactive, so the user can manage nodes and relationships.
- **Palette View** – user can create nodes by using this view (e.g. Drag & Drop, double click, right click) which contains all shapes defined in the SHACL constraints.
- **Properties View** – this view allows the user to fill in metadata properties through table representation, where the properties are grouped by mandatory, recommended and optional as they are defined in the SHACL constraints.
- **Outline View** – This view shows all nodes and relationships presents on graph. The nodes are grouped according to the shape type and are displayed in a tree-like structure.
- **RDF/Turtle View** – This view shows the data graph serialised in RDF/Turtle format.

DCAT Metadata Editor

Main features

- Create or import DCAT RDF/Turtle metadata files
- Import a SHACL file
- Validate RDF data graphs by using the SHACL constraints
- Metadata population through a graph manipulation (nodes and relationships)
- Apply filters on the graph in order to visualize or hide nodes
- Node type classification through colours
- Specify spatial information (e.g. POLYGON, POINT) by using an interactive GeoMap
- Import metadata information from Datacite
- Push DCAT RDF/Turtle metadata files to a GitLab repository



ENVRI
FAIR



envri.eu/envri-fair



[@envri_fair](https://twitter.com/envri_fair)



[company/envri-fair](https://www.linkedin.com/company/envri-fair)



facebook.com/ENVRIcomm