



EPOS Metadata Training within Geo-INQUIRE

Rossana Paciello and Daniele Bailo

Geo-INQUIRE is funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



Training Content

- **Part I (Daniele Bailo)**
 - Introduction
 - Achieving interoperability across systems
 - Approaches for achieving interoperability through Data, Metadata, Semantics, WebServices
 - EPOS: Making Data Interoperable
 - EPOS Technical Architecture
- **Part II (Rossana Paciello)**
 - INTEROPERABILITY LAYER
 - EPOS-DCAT-AP
 - EPOS-DCAT-AP Metadata Validation
 - EPOS Metadata Quality check
 - SHAP*eness* Metadata Editor



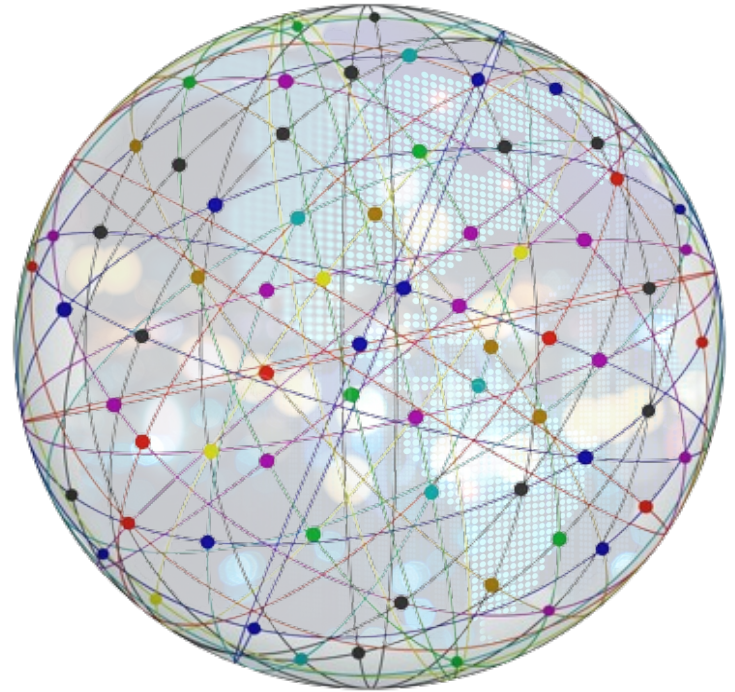
Introduction

- Technological progress and **open data** in **environmental and life sciences** are fostering **research opportunities**.
- Such progress offers a wealth of data for **studying complex phenomena**.
- Challenges in **data management** emerge, necessitating solutions for **effective data sharing** and reuse (e.g., **FAIR** principles).
- **Interdisciplinary research** amplifies these challenges because of the diverse practices, formats, and metadata standards in different disciplines.



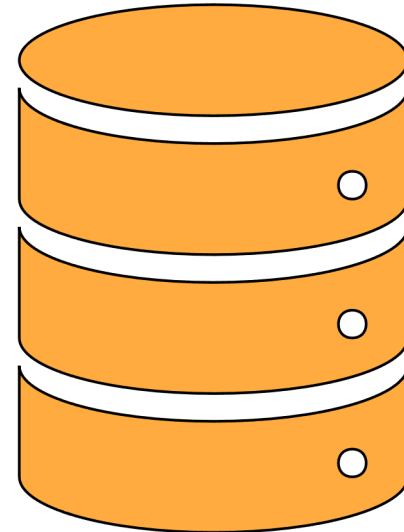
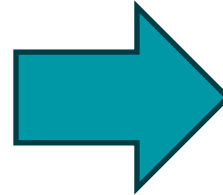
Achieving interoperability across systems

- The effective use of interdisciplinary data, requires implementing interoperability across different systems.
- The interoperability can be addressed at multiple levels:
 - Data
 - Metadata
 - Semantics
 - Services



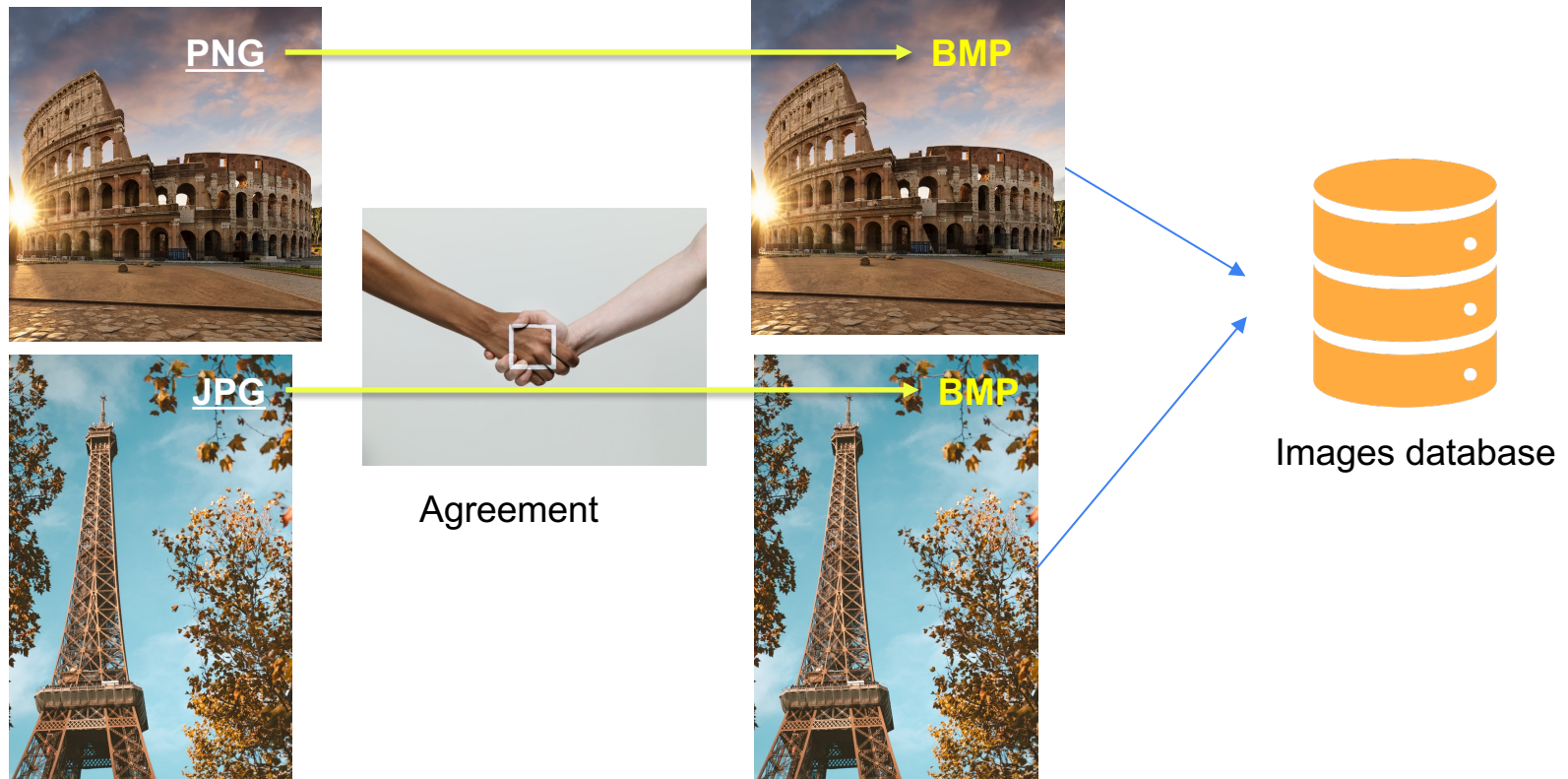
Interoperability

Example: *creating a database of European Cities images in BMP format
FROM heterogeneous sources*

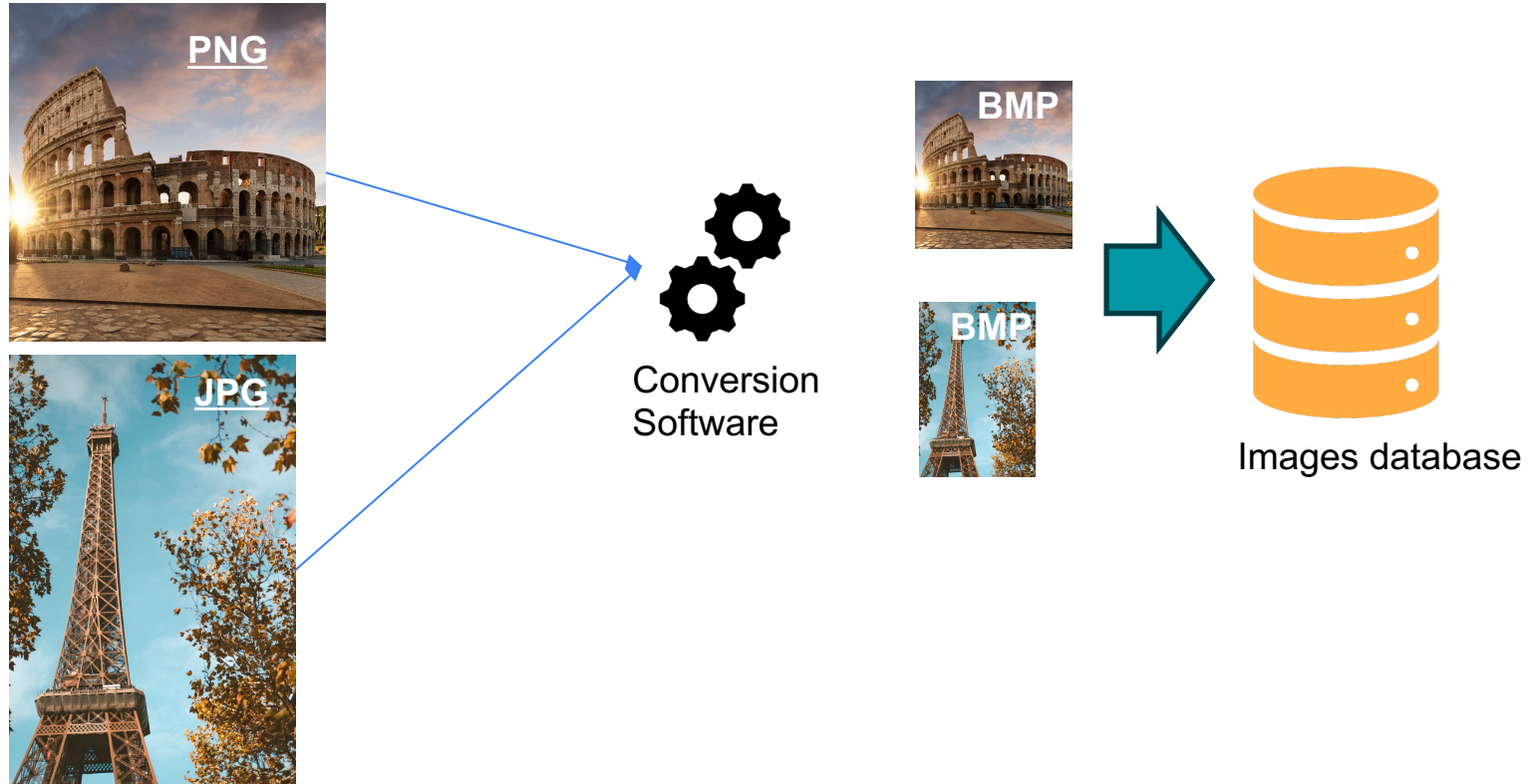


BMP Images database

Approach 1. Achieving interoperability through Data (1)



Approach 1. Achieving interoperability through Data (2)



Approach 1. Achieving interoperability through Data

- ***DATA is Physical representation*** of facts, atomic events, objective phenomena, information suitable for communication, interpretation and processing by human beings or automatic means
- The goal in an interdisciplinary landscape is to **harmonize** the plethora of different data formats and representations.
- Data harmonization refers to the process of **aligning and integrating data from various sources** to create a unified and standardized representation.
- The main tasks to achieve interoperability are:
 - **Convergence of “similar” data towards same (standard) formats (e.g., fresh start)**
 - **Conversion or mapping across different formats (e.g., existing data)**
- ***!! Requires coordination and consensus among different stakeholders***
- ***!! Requires conversion of datasets towards one common format***



Approach 2. Achieving interoperability through Metadata

Metadata = data about data



Image
(e.g., .jpeg format)

```
<?xpacket begin=" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmp:tk="Adobe XMP Core 5.6-c015 81.1
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/">
        <dc:format>image/jpeg</dc:format>
        <dc:title>
          <rdf:Alt>
            <rdf:li xml:lang="x-default">Colosseo</rdf:li>
          </rdf:Alt>
        </dc:title>
        <dc:creator>
          <rdf:Seq>
            <rdf:li>John Doe</rdf:li>
          </rdf:Seq>
        </dc:creator>
      </rdf:Description>
      <rdf:Description rdf:about=""
        xmlns:xmp="http://ns.adobe.com/xap/1.0/">
        <xmp:CreateDate>2023-10-01T10:00:00Z</xmp:CreateDate>
      </rdf:Description>
    </rdf:RDF>
  </x:xmpmeta>
<?xpacket end="w"?>
```

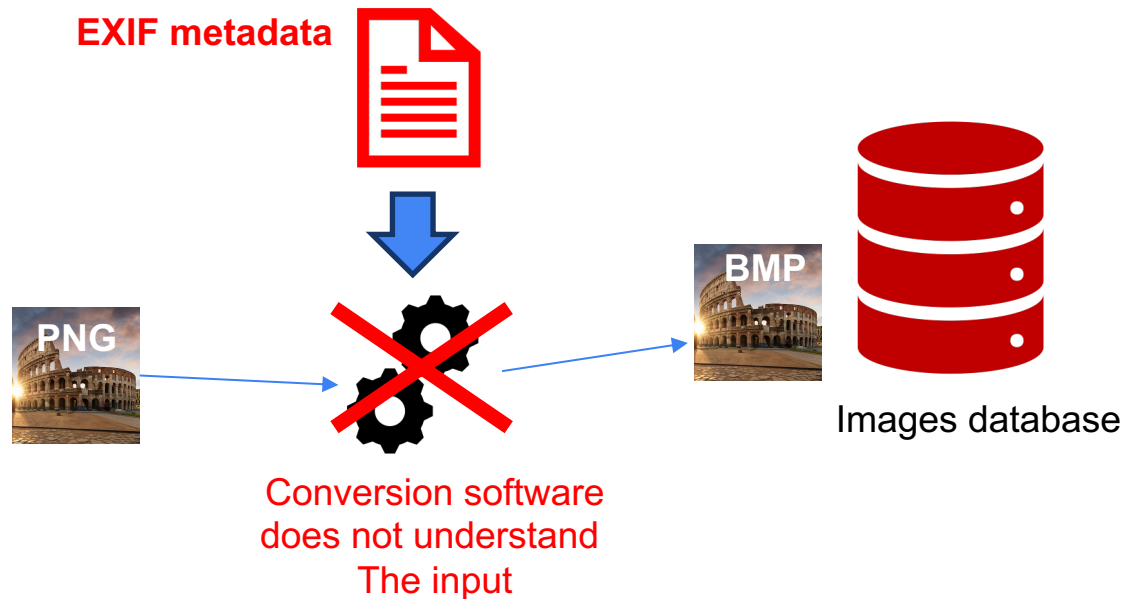
XMP Metadata Description



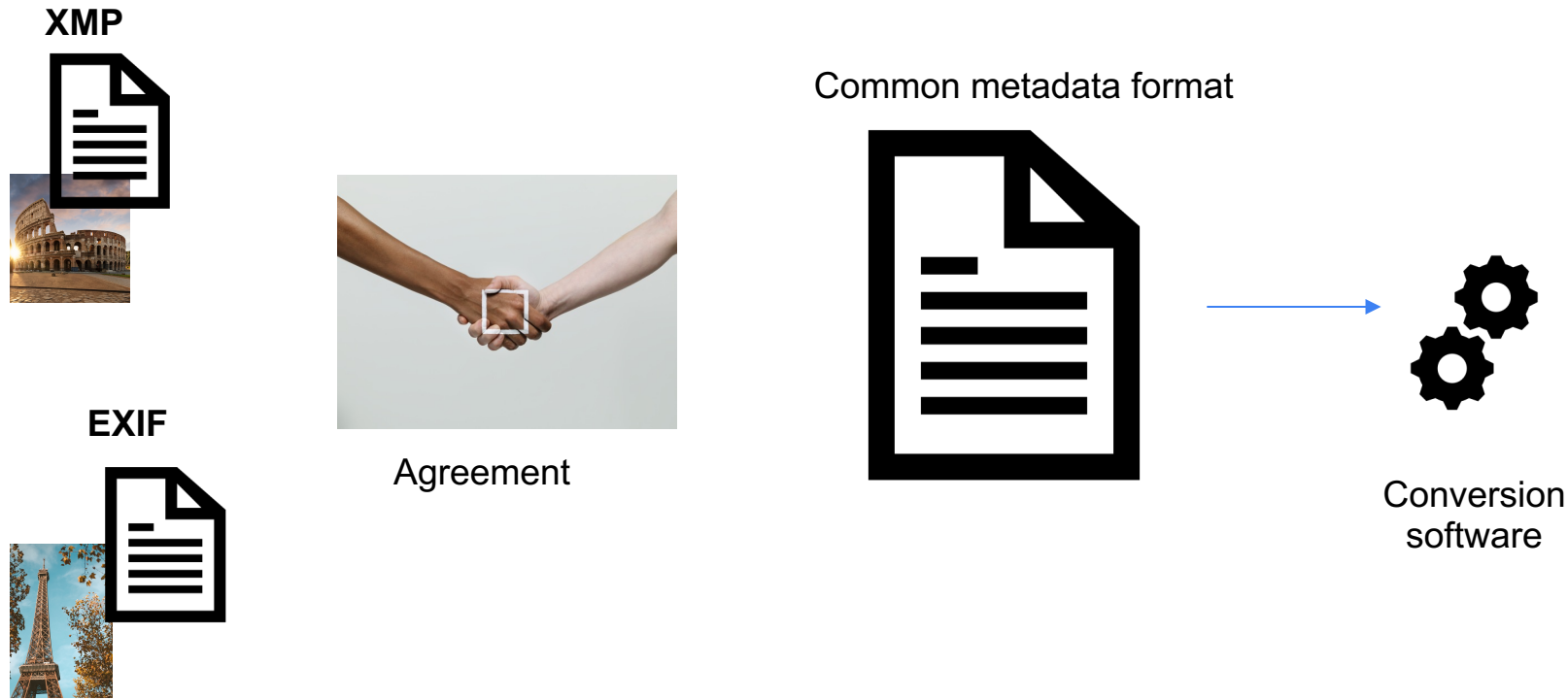
Approach 2. Achieving interoperability through Metadata

What if metadata has a different format?

EXIF metadata



Approach 2. Achieving interoperability through Metadata



Approach 2. Achieving interoperability through Metadata

- **METADATA is data that defines other data.** It provides essential contextual information, such as data source, format, authorship, and quality, making data more understandable and manageable.
- **Metadata improves data management by making data more discoverable** and understandable and providing insights into data source, quality, and usage.
- Metadata **specifies data standards and formats**, ensuring data adheres to common conventions, thus enabling machine readability.
- Metadata helps achieving Interoperability by offering context and details about the data **without directly altering the data itself.**
- Interoperability through metadata requires:
 - **Adoption of commonly accepted metadata standards (e.g., Dublin Core) or development of custom standards that can be interpreted by different machines**
 - **Define mappings between different metadata schemas or standards to ensure compatibility among sources**
- *!! Also requires coordination and consensus among different stakeholders*
- *!! Requires efforts for metadata creation and continuous curation to keep metadata up-to-date*

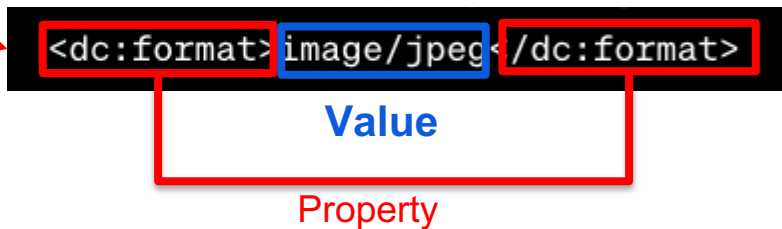


Approach 3. Achieving interoperability through Semantics

```
<?xpacket begin="" id="W5MOMpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmp:tk="Adobe XMP Core 5.6-c015 81.1
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      <dc:format>image/jpeg</dc:format>
      <dc:title>
        <rdf:Alt>
          <rdf:li xml:lang="x-default">Colosseo</rdf:li>
        </rdf:Alt>
      </dc:title>
      <dc:creator>
        <rdf:Seq>
          <rdf:li>John Doe</rdf:li>
        </rdf:Seq>
      </dc:creator>
    </rdf:Description>
    <rdf:Description rdf:about=""
      xmlns:xmp="http://ns.adobe.com/xap/1.0/"
      <xmp:CreateDate>2023-10-01T10:00:00Z</xmp:CreateDate>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?>
```

- **Semantics = meaning of data**

- **Semantics is embedded into the metadata**



Approach 3. Achieving interoperability through Semantics

- All properties (and classes) needs to be described (e.g., online vocabularies),
- so that machines can “understand” what they are about.

```
<?xpacket begin=" id="W5MOMpCehiHzreSzNTczkc9d"?">
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 5.6-c015 81.1.1"
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    <rdf:Description rdf:about=""
      xmlns:dc="http://purl.org/dc/elements/1.1/" >
      <dc:format>image/jpeg</dc:format>
      <dc:title>
        <rdf:Alt>
          <rdf:li xml:lang="x-default">Colosseo</rdf:li>
        </rdf:Alt>
      </dc:title>
      <dc:creator>
        <rdf:Seq>
          <rdf:li>John Doe</rdf:li>
        </rdf:Seq>
      </dc:creator>
    </rdf:Description>
    <rdf:Description rdf:about=""
      xmlns:xmp="http://ns.adobe.com/xap/1.0/" >
      <xmp:CreateDate>2023-10-01T10:00:00Z</xmp:CreateDate>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end="w"?">
```

Term Name: format	Property description in rdf Dublin Core More details
URI	http://purl.org/dc/terms/format
Label	Format
Definition	The file format, physical medium, or dimensions of the resource.
Comment	Recommended practice is to use a controlled vocabulary where available. For example, for file formats one could use the list of Internet Media Types [MIME] . Examples of dimensions include size and duration.
Type of Term	Property Controlled vocabulary
Range Includes	<ul style="list-style-type: none">http://purl.org/dc/terms/MediaTypehttp://purl.org/dc/terms/Extent
Subproperty of	<ul style="list-style-type: none">Format (http://purl.org/dc/elements/1.1/format)

Approach 3. Achieving interoperability through Semantics

Semantic can be expressed in different formats (rdf, JSON-LD etc.)

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "ImageObject",
  "description": "Colosseo",
  "encodingFormat": "image/jpeg",
}
</script>
```

JSON-LD description in schema.org

Schema.org

Docs

Schemas

Validate

About

encodingFormat

A Schema.org Property

Thing > Property :: encodingFormat

Media type typically expressed using a MIME format (see IANA site and MDN reference), e.g. application/z for .mp3 etc.

In cases where a **CreativeWork** has several media type representations, **encoding** can be used to indicate **encodingFormat** information.

Unregistered or niche encoding and file formats can be indicated instead via the most appropriate URL, e.g entry.

Approach 3. Achieving interoperability through Semantics

- **Semantics is used for defining the meaning** of (meta)data elements, relationships between data, and the intended use of data.
- Semantic interoperability involves **encoding knowledge in a machine-readable format (vocabularies, ontologies)**, making it possible to understand data beyond syntactical interpretation. Usually done within the metadata.
- Systems that achieve semantic interoperability can perform **sophisticated semantic queries that consider the meaning and context of data**, making data retrieval more precise and relevant.
- Semantics interoperability requires:
 - **Adopt semantic standards such as RDF (Resource Description Framework) and OWL (Web Ontology Language) to represent data in a machine-understandable format.**
 - **Create semantic mappings between ontologies or data models used by different organizations or systems, enabling data translation between different semantic schemas.**
- *!! Creating or adopting ontologies is quite effort consuming (is it worth it?)*
- *!! Ontologies and semantic mappings require continuous maintenance to ensure they are up to date and accurately reflect the evolving real world.*

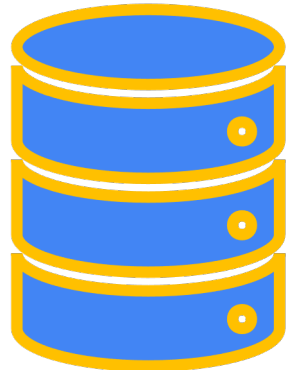
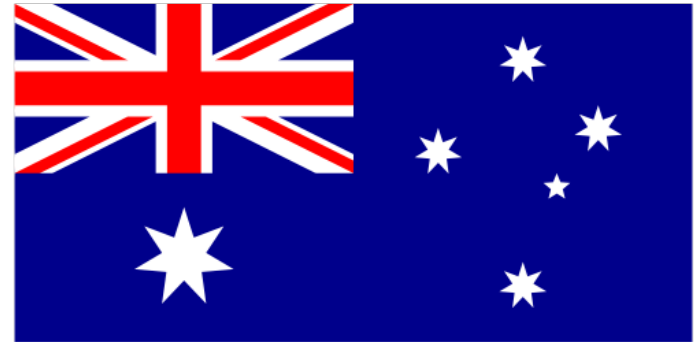


Approach 3. Achieving interoperability through Semantics

Mapping between two standards: DCAT-AP → schema.org

DCAT-AP Specification	Obl.	Label	QName	Schema.org	Comments
[DCAT-AP]	M	access URL	dcat:accessURL	schema:contentURL	
[DCAT-AP]	<i>R</i>	description	dct:description	schema:description	
[DCAT-AP]	<i>R</i>	availability	dcatap:availability	schema:??	TBD
[DCAT-AP]	<i>R</i>	format	dct:format →	schema:encodingFormat	

Approach 4. Achieving interoperability through WebServices



**EU Images
database**

- ▲ Root
 - ▲ Folder #1
 - ▷ Folder #1
 - ▷ Folder #2
 - ▲ Folder #2
 - ▷ Folder #1
 - ▷ Folder #2

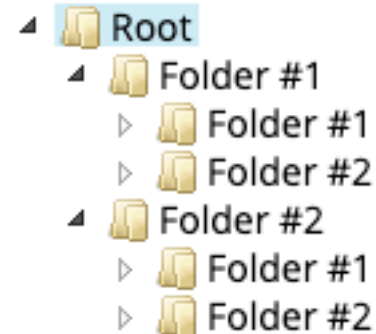
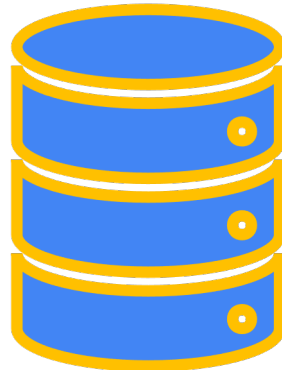
**AU Images
database**



Approach 4. Achieving interoperability through WebServices



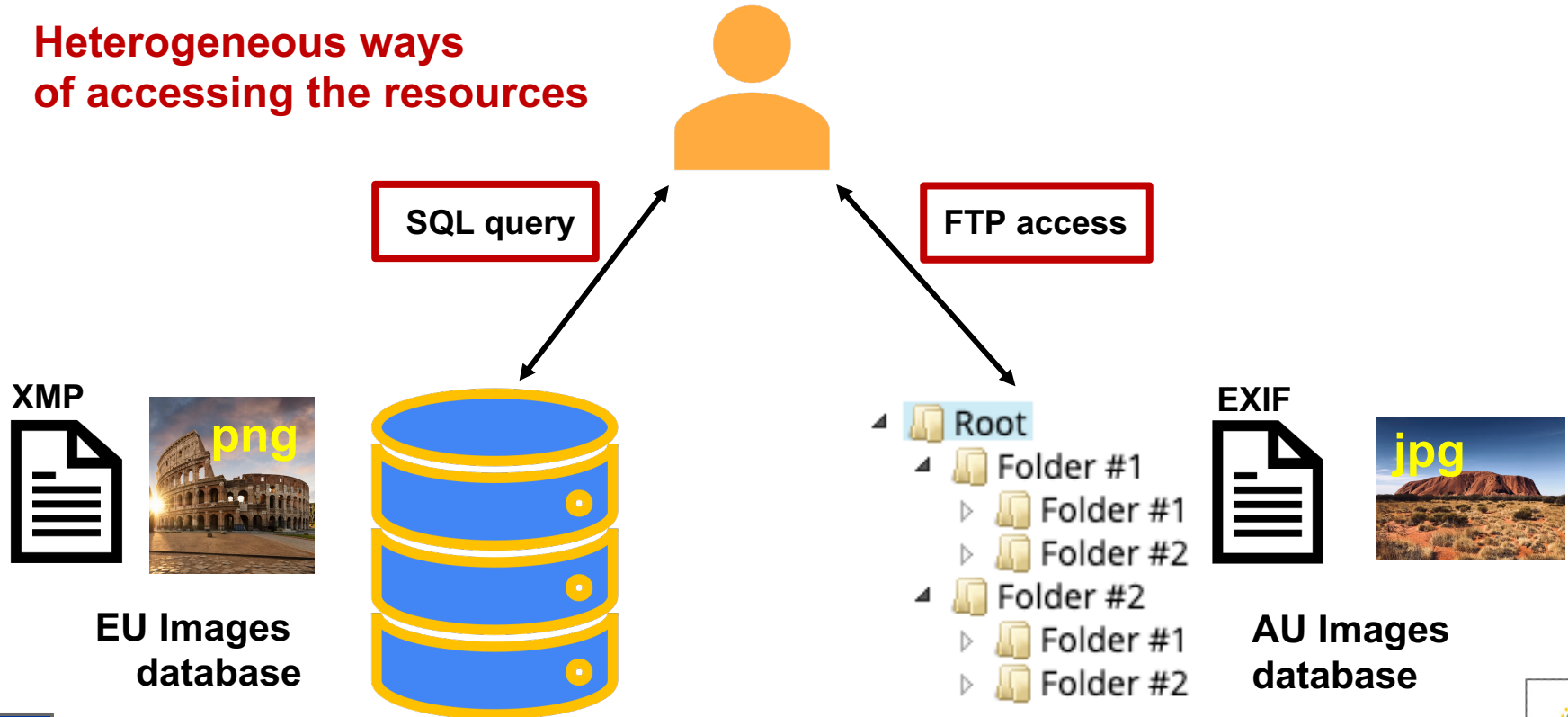
EU Images
database



AU Images
database

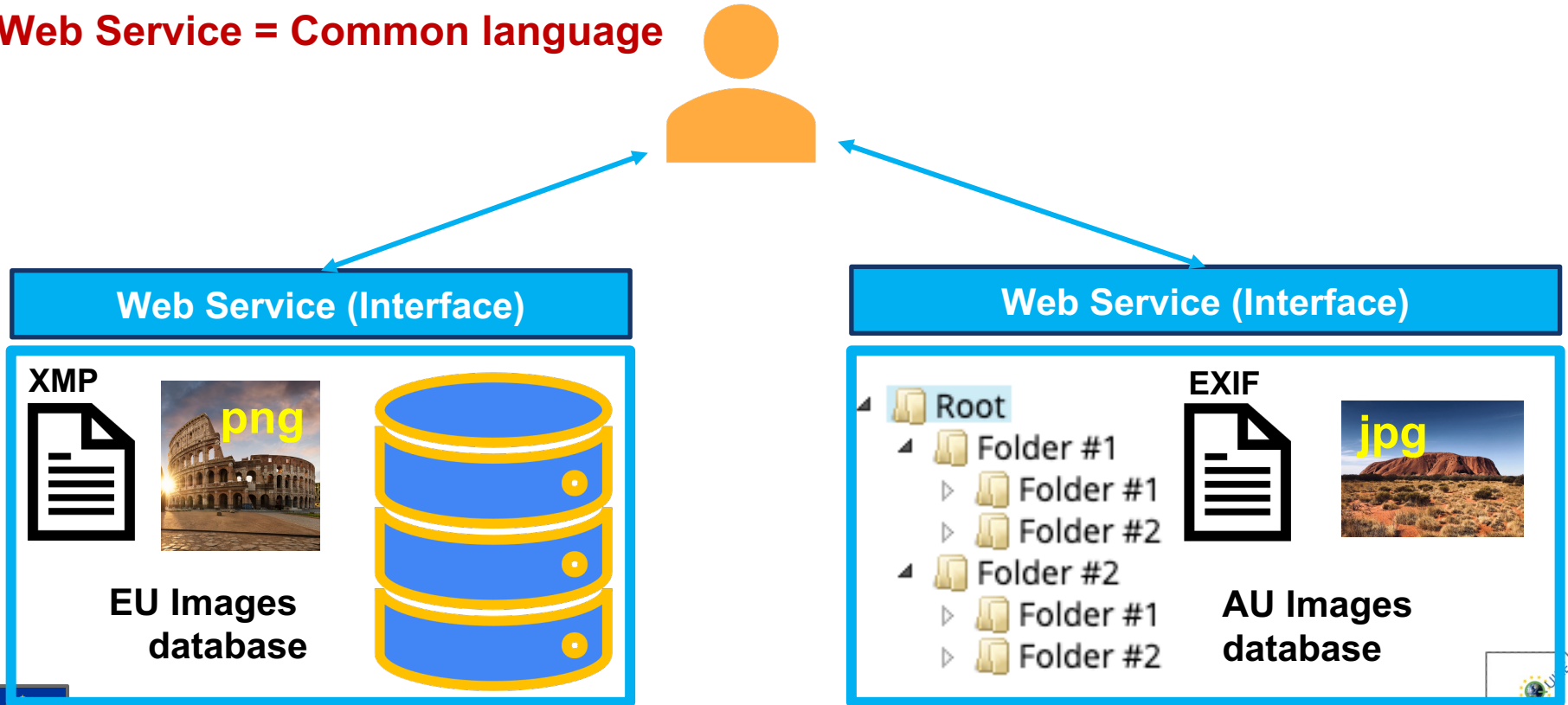
Approach 4. Achieving interoperability through WebServices

**Heterogeneous ways
of accessing the resources**



Approach 4. Achieving interoperability through WebServices

Web Service = Common language



Approach 4. Achieving interoperability through WebServices

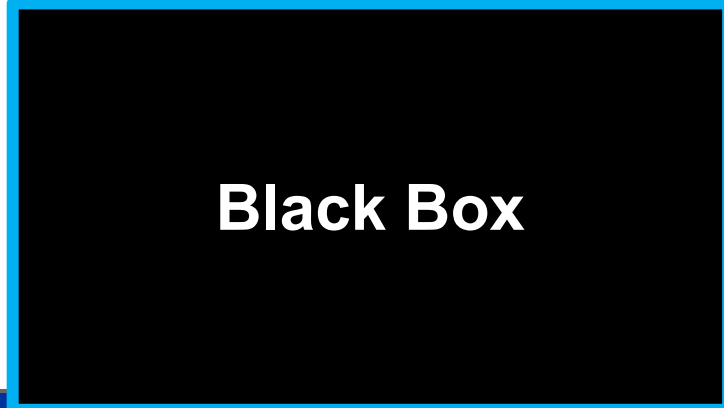
**Web Service = Common language
for resources access**



Uniquely interface definition:

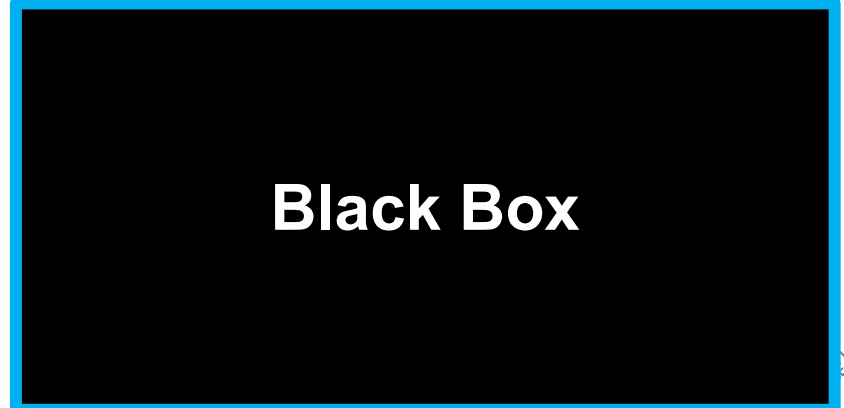
- communication Protocols
- input requests (paramters)
- output responses

Web Service (Interface)



Black Box

Web Service (Interface)



Black Box



Approach 4. Achieving interoperability through WebServices

- **A web service is a generic term for an interoperable machine-to-machine software** that is hosted at a network addressable location.
- A web service is an **interface, which hides the implementation details** so that it can be used independently of the hardware, software programming language by which it is implemented
- Interoperability through web services requires:
 - Decide on a communication protocol, commonly SOAP (Simple Object Access Protocol) or REST (Representational State Transfer).
 - Select Request/Response Format. Typically, XML or JSON (JavaScript Object Notation) is used to structure the data.
- *!! May require agreements about using same web service (same interface). Hard to achieve in multi disciplinary context*
- *!! Web services can be described through semantic rich metadata information. Homogenous access can then be achieved through advanced software that adopts semantic crosswalks for parameters mappings*



Approach 4. Achieving interoperability through WebServices

Rich metadata and semantic description of web services parameter(s)

```
hydra:mapping[ a hydra:TriTemplateMapping:  
  hydra:variable "minlatitude"^^xsd:string;  
  hydra:property "epos:southernmostLatitude" ;  
  rdfs:range "xsd:float";  
  rdfs:label "Minimum Latitude";  
  schema:minValue "-90.0";  
  schema:maxValue "90.0";  
  hydra:required "true"^^xsd:boolean;  
  schema:defaultValue "36.61" ;  
];
```

Web Service A (Interface)

Black Box A

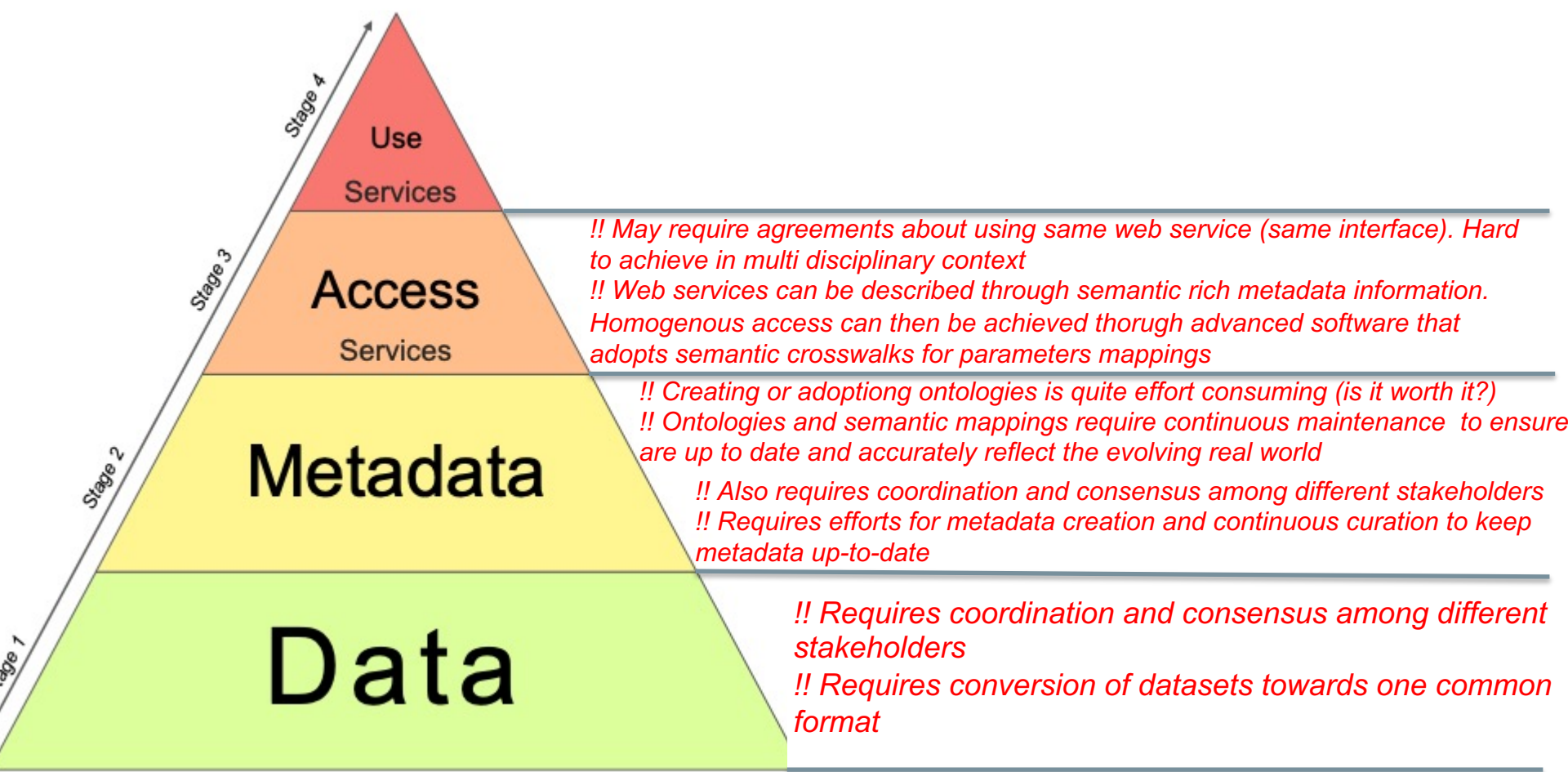
mapping

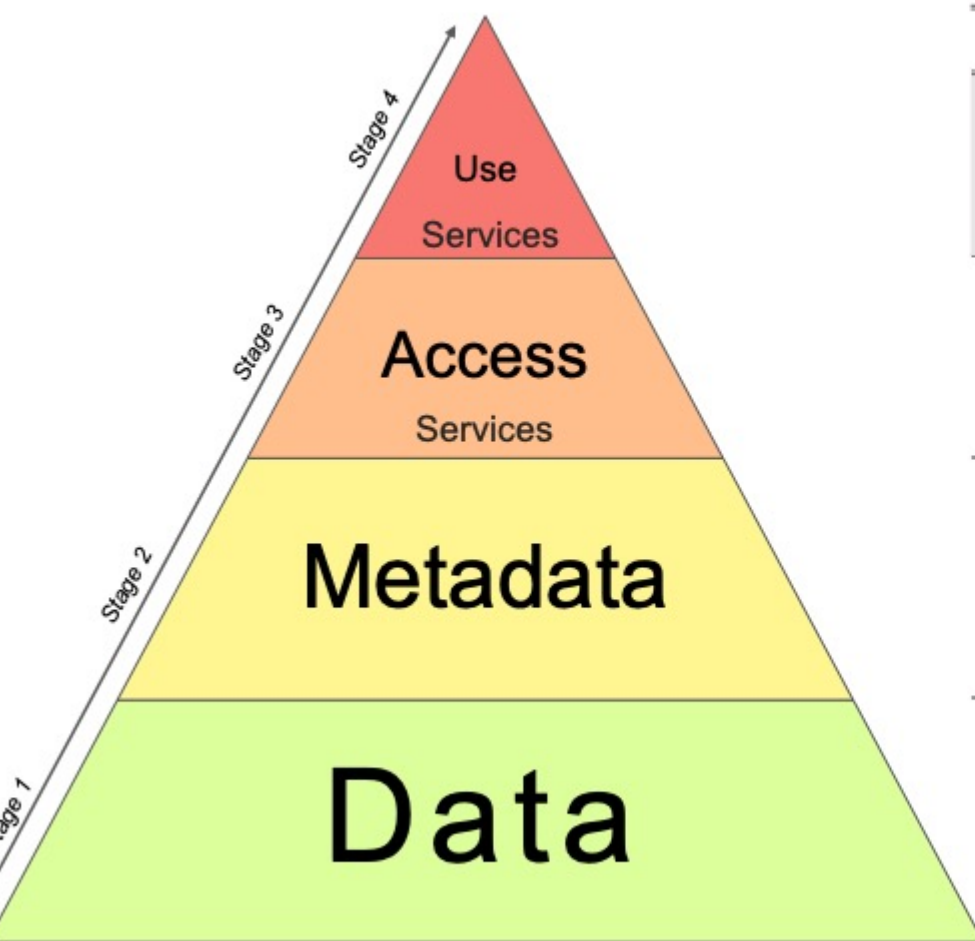
Semantic information

```
hydra:mapping[ a hydra:TriTemplateMapping:  
  hydra:variable "minlt"^^xsd:string;  
  hydra:property "epos:southernmostLatitude" ;  
  rdfs:range "xsd:float";  
  rdfs:label "Minimum Latitude";  
  schema:minValue "-90.0";  
  schema:maxValue "90.0";  
  hydra:required "true"^^xsd:boolean;  
  schema:defaultValue "34.61" ;  
];
```

Web Service B (Interface)

Black Box B





FAIR PRINCIPLES

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1.** the protocol is open, free, and universally implementable.
- A1.2.** the protocol allows for an authentication and authorization procedure, where necessary.
- F4.** (meta)data are registered or indexed in a searchable resource.

- F1.** Metadata are assigned a globally unique and eternally persistent identifier.
- F2.** data are described with rich metadata.
- F3.** metadata specify the data identifier.
- F4.** metadata are registered or indexed in a searchable resource.
- A2.** metadata are accessible, even when the data are no longer available.
- I1.** metadata use a formal, accessible, shared, and broadly applicable.
- I2.** metadata use vocabularies that follow FAIR principles.
- I3.** metadata include qualified references to other metadata.
- R1 (R1.1 - R1.2 - R1.3)** Metadata are richly described with a plurality of accurate attributes.

- F1.** Data are assigned a globally unique and eternally persistent identifier.
- F4.** Data are registered or indexed in a searchable resource.
- I1.** Data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** Data use vocabularies that follow FAIR principles.
- I3.** Data include qualified references to other (meta)data.
- R1.1.** Data are released with a clear and accessible data usage license.
- R1.3.** Data meet domain-relevant community standards.

Bailo D, Paciello R, Sbarra M, Rabissoni R, Vinciarelli V and Cocco M (2020) Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures. *Front. Earth Sci.* 8:3. doi: 10.3389/feart.2020.00003

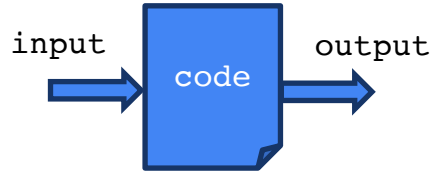


From local Data Repositories to web services

Your Laptop



Local Data Storage



Local Software
Processing

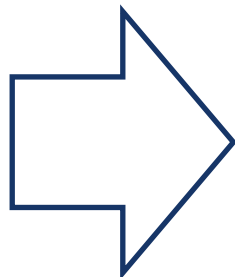


From local Data Repositories to web services

Your Laptop



Local Data Storage



STANDARD WEB SERVICE

ACCESS

Metadata catalogue
Or equivalent



META
DATA

input

code

output

Local Software
Processing



input

code

output

Software Processing

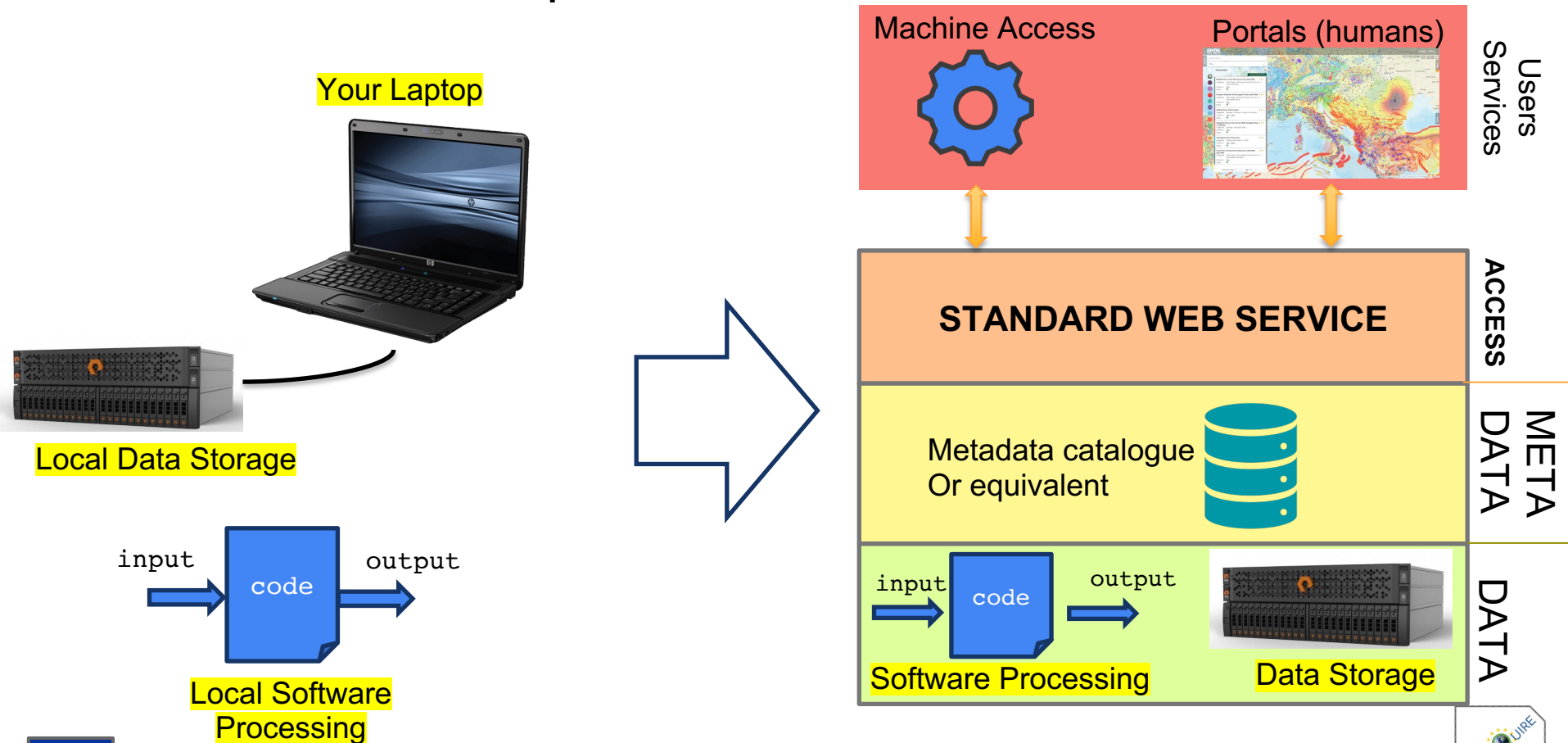
Data Storage



DATA



From local Data Repositories to web services



EPOS: Making Data Interoperable

- EPOS adopts the **combined** approach to provides true integration, offering a single data portal for accessing multidisciplinary data resources.
- **Semantic** descriptions in **metadata** enable adaptation to different Web Services APIs and data source complexities.
- **Data sources not forced to adopt one common standard** but only to implement basic recommendations based on the **FAIR** principles and providing metadata descriptions of the data sources.
- **Data consumers can use Web Services APIs to consume heterogeneous resources as if they were homogeneous** in terms of communication protocols, metadata description and data formats.



Free text search

Filters

46.69467 - 8.21777

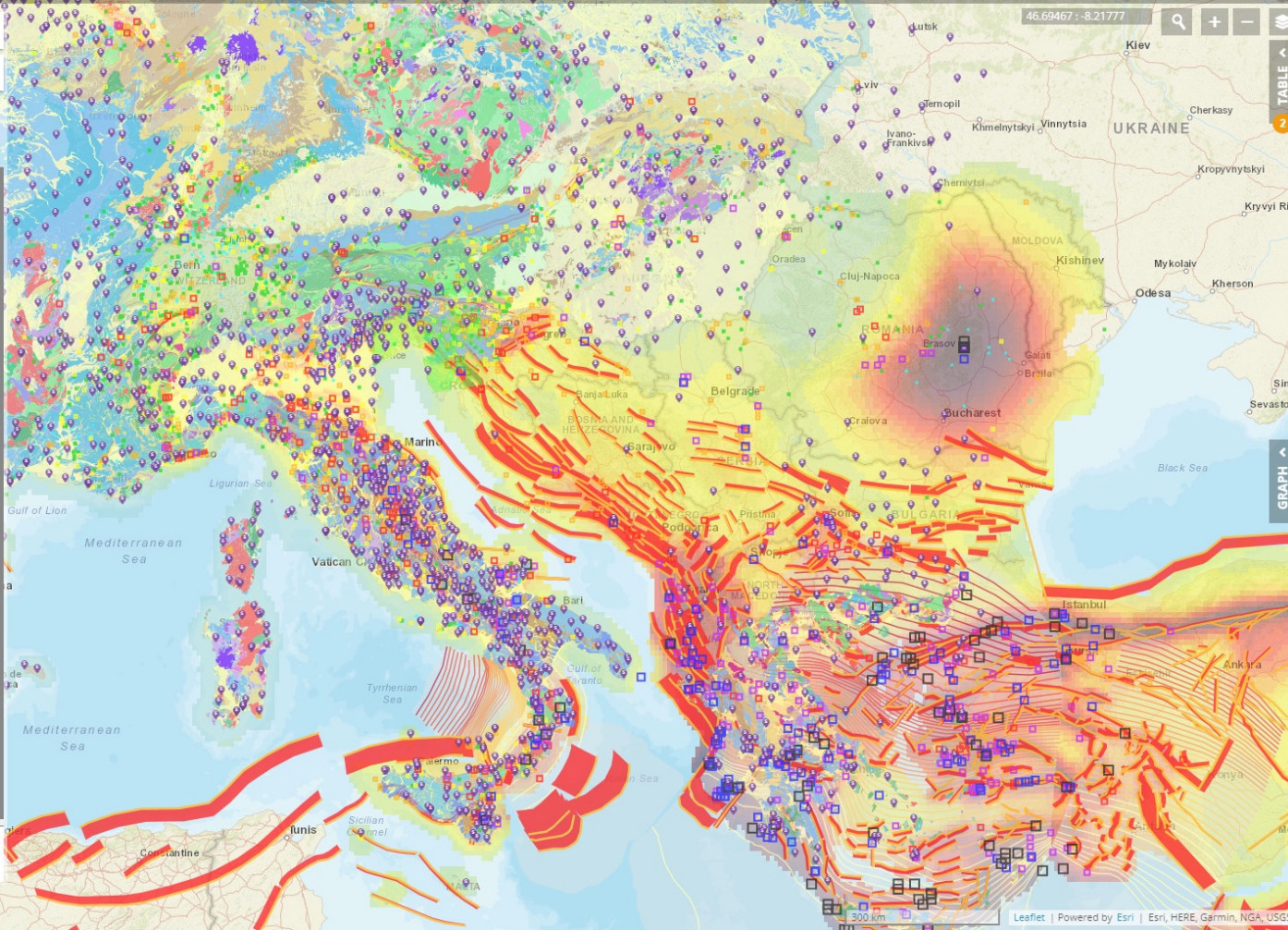


SEARCH

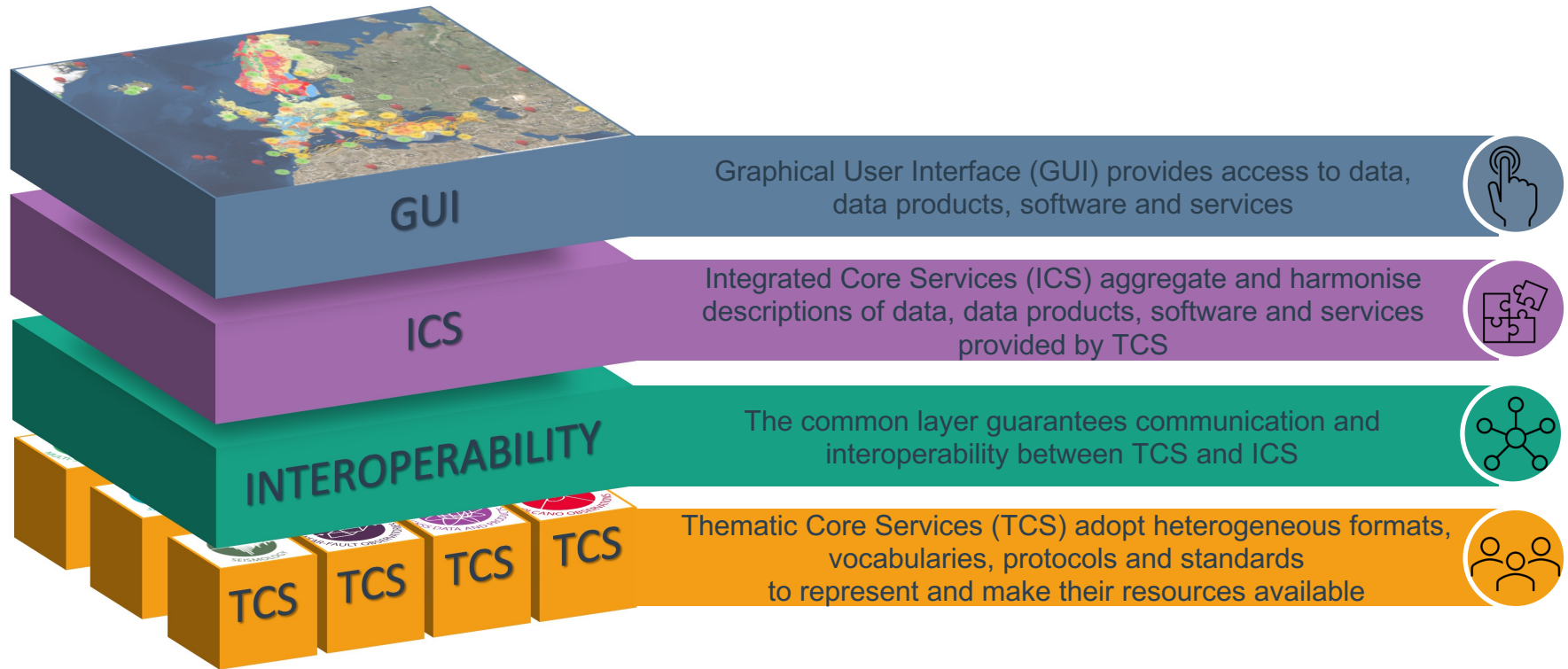
Favourites 6

Clear all favourites ✕

- ESHM20 475 yr mean PGA hazard map (OGC WMS)** i ★ ▼
Categories: Seismology > Earthquake hazard and risk ser... > Hazard products
Visible on: [Map](#)
Status: ✔
- European Database of Seismogenic Faults (OGC WMS)** i ★ ▼
Categories: Seismology > Earthquake hazard and risk ser... > Seismogenic faults
Visible on: [Map](#)
Status: ✔
- GNSS Stations with Products** i ★ ▼
Categories: Geodesy > Products > Station Information
Visible on: [Map](#) [Table](#)
Status: ✔
- Geological Feature View Service (EGDI Geological Map 1:1,000,000)** i ★ ▼
Categories: Geology > Geological Maps
Visible on: [Map](#)
Status: ✔
- LOS Displacement Time Series** i ★ ▼
Categories: Satellite Observations > InSAR
Visible on: [Map](#) [Table](#)
Status: ✔
- Parameters of historical earthquakes (1000-1899) - OGC WMS** i ★ ▼
Categories: Seismology > Seismological products service... > Earthquake parameters
Visible on: [Map](#)
Status: ✔



EPOS Technical Architecture

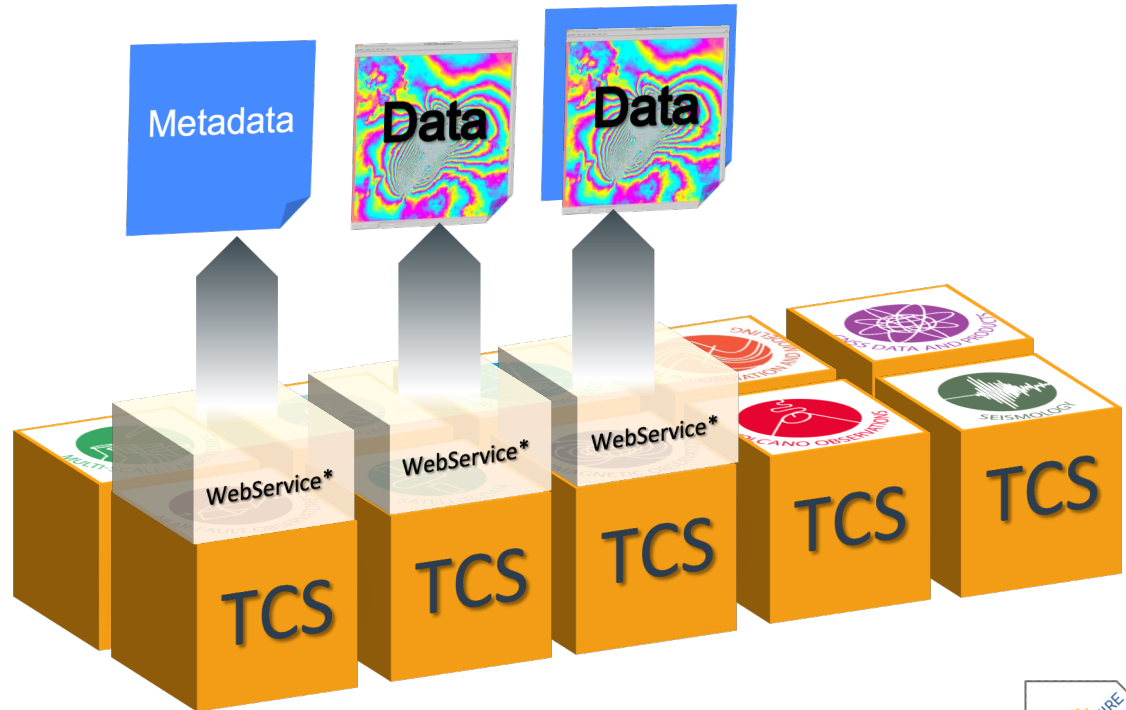


D. Bailo, K.G. Jeffery, K. Atakan, L. Trani, R. Paciello, V. Vinciarelli, J. Michalek, A. Spinuso, **Data integration and FAIR data management in Solid Earth Science**, in *Annals of Geophysics*, vol. 65 No. 2 (2022), Special issue: EPOS a Research Infrastructure in solid Earth: open science and innovation. Doi: <https://doi.org/10.4401/ag-8742>.



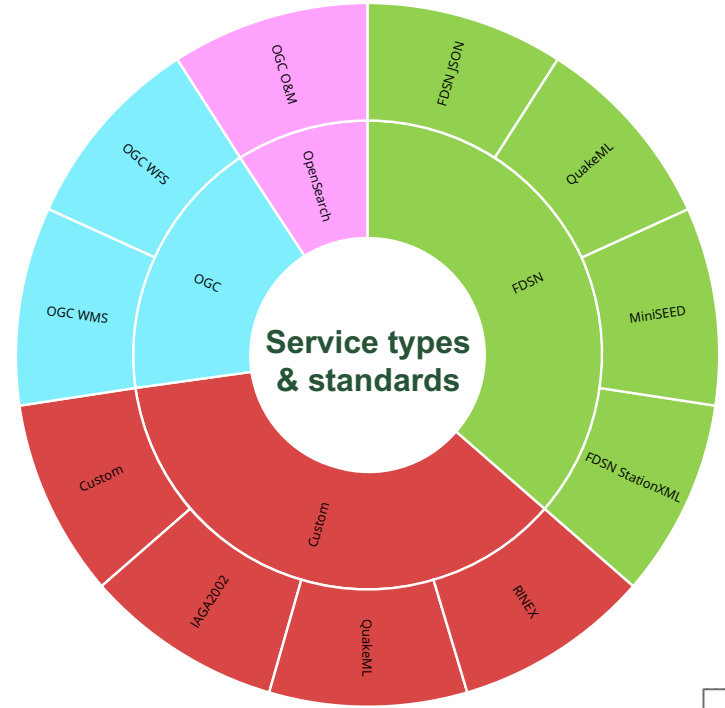
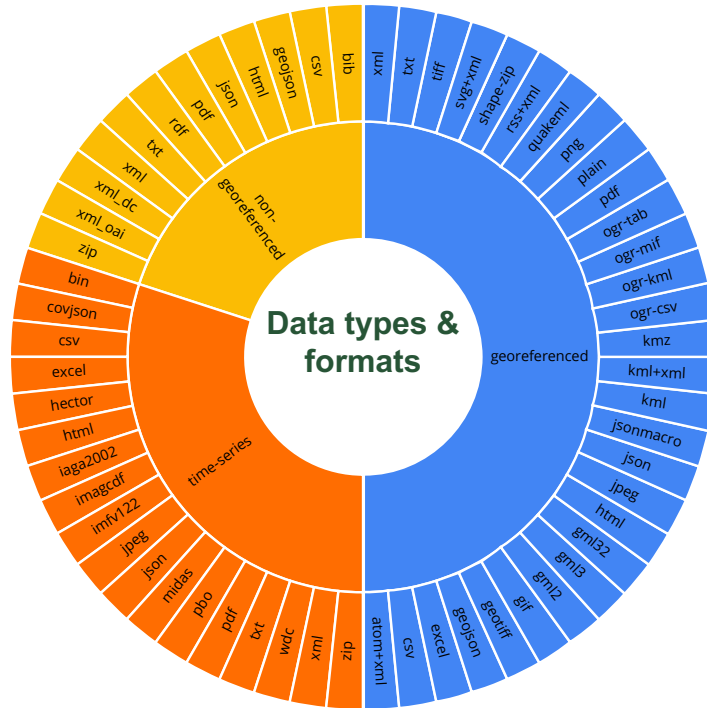
TCS= Data provider

- TCS represent datasets and services provided by domain specific communities.
- Datasets are produced by a heterogeneous set of data providers.
- Datasets are accessible by means of web services*.





Heterogeneity of TCS Data and Services

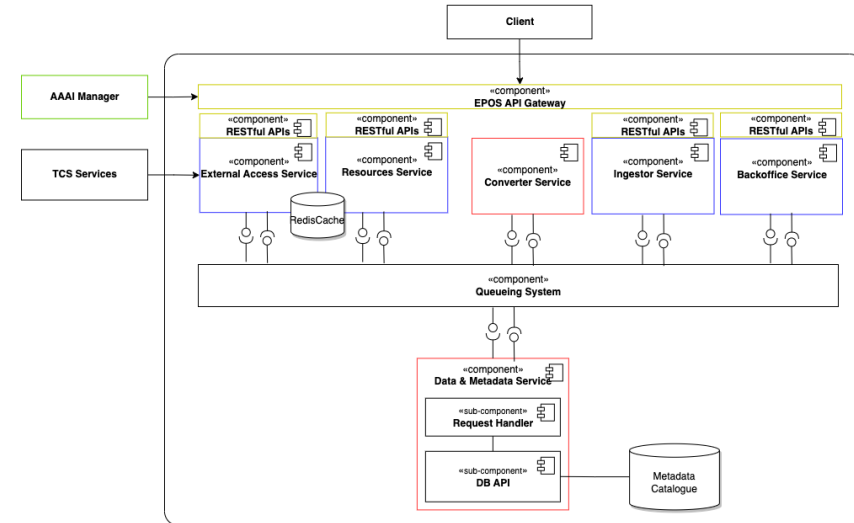




ICS= Central Integration HUB

It consists of three sub tiers:

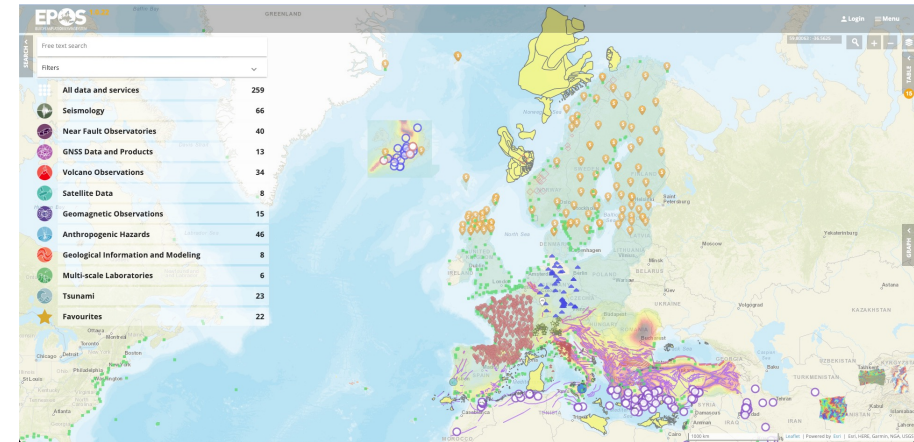
- **WebAPIs**, which provide a set of RESTful endpoints to enable the communication with GUI tier.
- **Micro-services**, which enable ICS to run in a distributed environment, properly scale specific system functionalities and enhance reliability.
- **Metadata Catalogue**, which is used to store information about data, data products, software, services, and other information associated with them.





Main features:

- **Search**, by specifying a set of search criteria (e.g., free-text search, spatio-temporal range).
- **Results Exploration**, browsing the retrieved results and previewing them in different ways (Map, Table, Graph).
- **Results Refinement**, results can be further refined through service-specific options.
- **Final Selection**, downloading the selected results or putting them into a list of favourites.



What is the **LAYER** where heterogeneous resources are harmonized to provide users with an integrated access?



Q&A

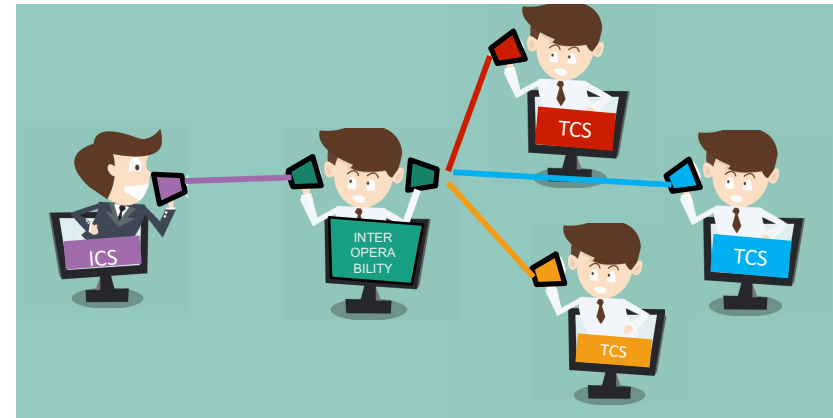


INTEROPERABILITY LAYER





- It was developed to enable the interaction between **ICS** and **TCS**.
- It allows **TCS** to provide descriptions of their assets.
- It allows **ICS** to capture, organise and harmonise information from different sources.
- It adopts a common knowledge representation language defined in the form of application profile.



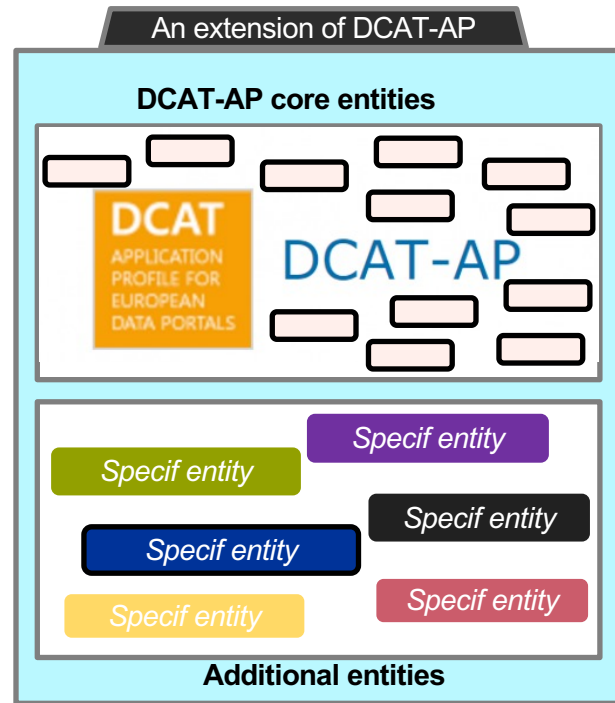
An Application Profile to exchange knowledge

- A profile supports a community interested in **interoperability** and **data exchange** within of specific context.
- It defines what **entities** are described, their **relationships**, the use of controlled **vocabularies**, cardinality of fields/properties and data types.
- The DCAT Application Profile (**DCAT-AP**) is a specification developed by W3C to enable the exchange of dataset descriptions between data portals in Europe.
- Several extensions of DCAT-AP have been created to address domain-specific requirements.



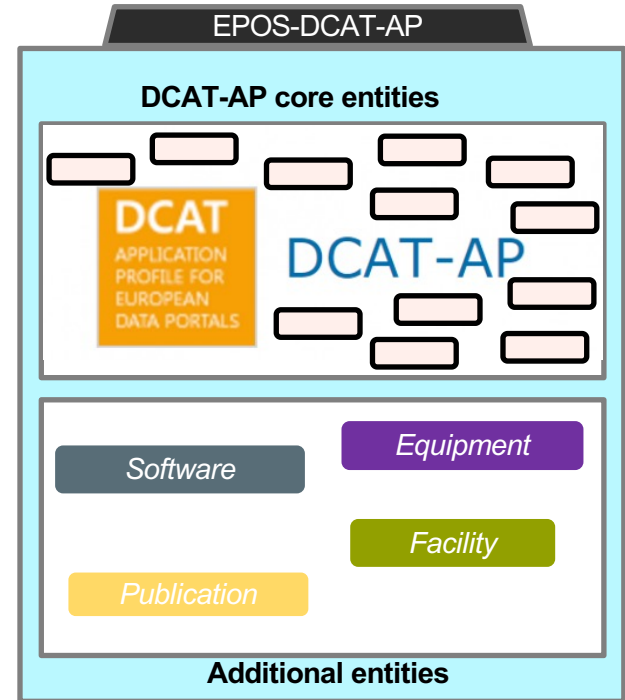
What is an extension of DCAT-AP?

- DCAT-AP defines a set of **classes and properties**.
- An extension of DCAT-AP address **domain-specific requirements** by adding classes and properties.
- An extension of DCAT-AP needs to meet the **minimum requirements** described in DCAT-AP specification.



EPOS-DCAT-AP: an extension of DCAT-AP for Research Infrastructures in the solid-Earth domain

- **EPOS-DCAT-AP extends DCAT-AP** to address the specific requirements of the EPOS communities.
- It **adds a set of classes and properties** that covers the main assets and resources useful in the EPOS context.
- It introduces **new relationships and roles**.
- It **reuses elements** of popular vocabularies (e.g., Schema.org and the Hydra Core Vocabulary).
- It enables integrity checks and validation (via **SHACL**).



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:
(i) Identified the essential elements and attributes of EPOS-DCAT-AP in the EPOS context;
(ii) Identified the controlled vocabularies to be used in the EPOS context; (iii) Identified the strict minimum description metadata to be exchanged between TCS and ICS.

Version 0.12
Revision of previous version taking into account the following criteria:
(i) Ensure compliance with DCAT-AP v 1.1; (ii) update EPOS namespace.

Version 0.14
Revision of previous version taking into account the following criteria:
(i) Ensure compliance with DCAT-AP v 1.1; (ii) In order to steer towards quality metadata descriptions, the implementation of a number of properties are recommended or made mandatory.

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.

Version 1.0
This version addresses bugs identified by the previous versions and covers changes related to the Equipment and Facility entities.

Version 2.0
This version is based on DCAT-AP v 2.1.1 because some of the gaps identified by EPOS-DCAT-AP v 1.0 were filled by DCAT-AP v2 (e.g., introducing DataService class).

Version 3.0
This version is based on DCAT-AP v 3.0 introducing Dataset Series and DatasetInSeries classes.

**February
2018**

**July
2018**

**February
2019**

**June
2019**

**September
2020**

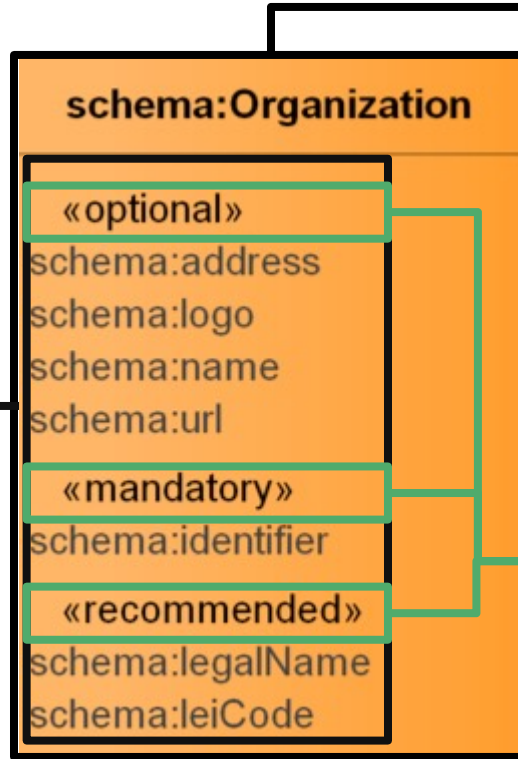
May 2022

**September
2023**



Class and Properties definition

PROPERTIES: describe the structure and value of an instance of a class



CLASS (or Entity): describes a set of objects that share the same features, constraints, semantics.

Mandatory: information for class or property **must** be provided.

Recommended: information for class or property **should** be provided.

Optional: information for class or property **may** be provided.



Classes characterizing EPOS-DCAT-AP

- **Operation**: a description of a web service operation.
- **IriTemplate**: it consists of a template literal and a set of mappings.
- **IriTemplateMapping**: a variable (parameter) used in the template.
- **Service**: a generic service.
- **Equipment**: a generic equipment (e.g., GPS sensor, seismic station).
- **Facility**: facility with resources and related services used by the scientific community to conduct top-level research in their respective fields, e.g., Seismic Network.
- **SoftwareApplication**: software packages, applications and programs.
- **SoftwareSourceCode**: computer programming source code.



EPOS-DCAT-AP: Syntactic and Semantic aspects

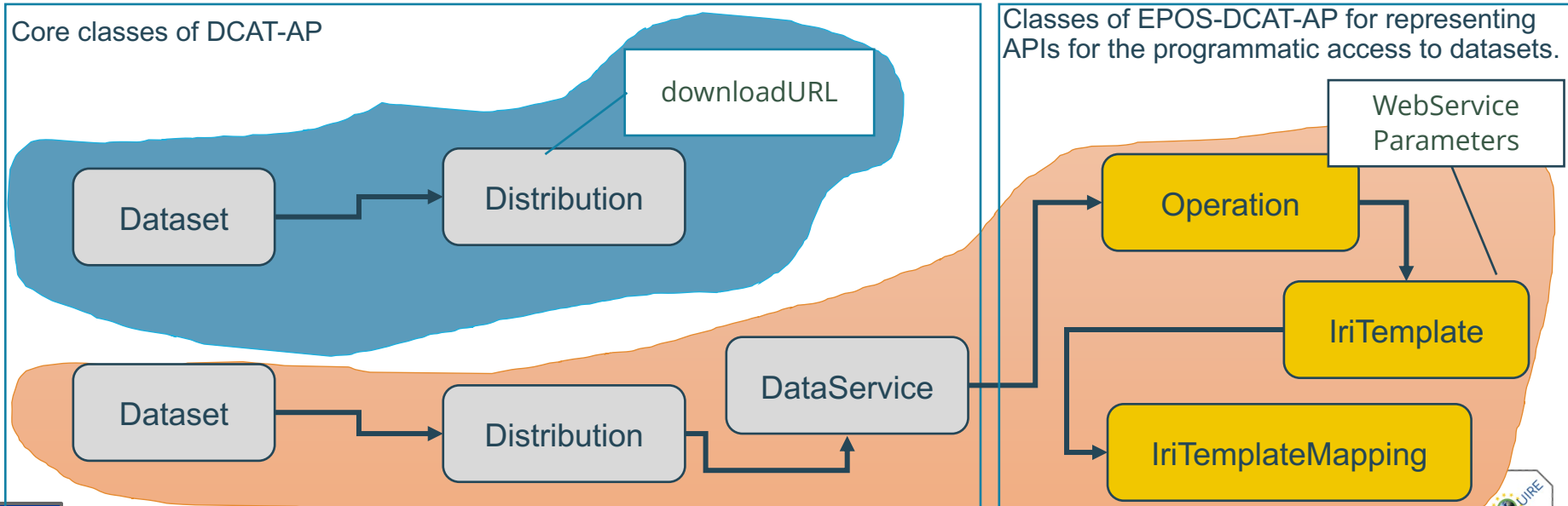
- EPOS-DCAT-AP allows to implement syntactic interoperability between ICS and TCS:
 - RDF/Turtle format is the encoding format used by TCS to provide metadata descriptions to ICS.
- EPOS-DCAT-AP allows to implement semantic interoperability between ICS and TCS:
 - TCS describe their main resources by using standard vocabularies (e.g., DCAT, Schema, Hydra) which allow ICS, and any machine, to capture the meaning associated with the metadata content.



Datasets and Web Services description through EPOS-DCAT-AP

The profile allows to describe Datasets distributed through:

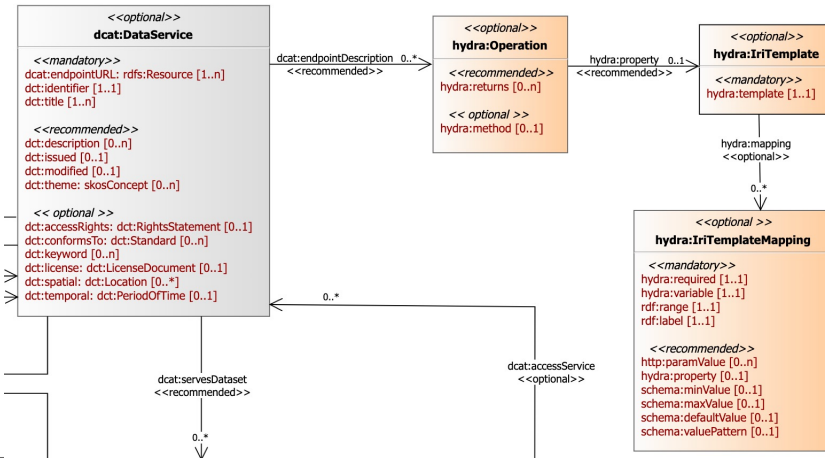
- **downloadable** files available from direct URLs;
- **web services** implemented according to global, well-established and community specific standards (e.g., OGC, RESTful, FDSN);



EPOS-DCAT-AP – Web Service description with semantics information

EPOS-DCAT-AP uses classes of **Hydra vocabulary** to describe access-specific properties for Application Programming Interface (API):

- **Operation**: IRI template
- **IriTemplate**: list of parameters that enable data access;
- **IriTemplateMapping**, each parameter of the IRI template.



```

<https://www.epos-eu.org/epos-dcat-ap/example/Operation/001> a hydra:Operation;
  hydra:method "GET";
  hydra:returns "application/xml" ;

hydra:property[ a hydra:IriTemplate ;
  hydra:template "http://www.test.org/{?param1,param2}^^xsd:string" ;
  hydra:mapping[ a hydra:IriTemplateMapping ;
    hydra:variable "param1"^^xsd:string ;
    rdfs:label "Start of the timespan" ;
    schema:valuePattern "YYYY-MM-DDThh:mm:ss" ;
    hydra:property "schema:startDate";
    hydra:required "true"^^xsd:boolean ;
    rdfs:range "xsd:dateTime" ;
    schema:minValue "2012-01-01T00:00:00" ;
    schema:maxValue "2017-12-01T00:00:00" ;

];
hydra:mapping[ a hydra:IriTemplateMapping ;
  hydra:variable "param2"^^xsd:string ;
  rdfs:label "Output format" ;
  hydra:required "false"^^xsd:boolean ;
  rdfs:range "xsd:string" ;
  hydra:property "schema:encodingFormat";
  http:paramValue "xml";
  http:paramValue "text";
  http:paramValue "json";

];
  
```

Semantic tags



EPOS-DCAT-AP RDF/Turtle serialization

EXAMPLE

DATASET

```
<https://www.epos-eu.org/Seismology/Dataset/001> a
dcats: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 ;
  dcat:contactPoint <http://orcid.org/0000-0001-0002-
0003/contactPoint>;
  dct:publisher <PIC:000518944> .
  dcat:distribution <https://www.epos-
eu.org/Seismology/Distribution/001> ;
  ...
.
```

DISTRIBUTION

```
<https://www.epos-eu.org/Seismology/Distribution/001> a
dcats:Distribution ;
  dct:title "Title of Seismic Waveform Distribution" ;
  dct:description "Description of Seismic Waveform
Distribution";
  dct:format
"http://publications.europa.eu/resource/authority/file-
type/BIN"^^xsd:anyURI ;
  dcat:accessService<https://www.epos-
eu.org/Seismology/Operation/001> ;
  ...
.
```

DATASERVICE

```
<https://www.epos-eu.org/Seismology/WebService/001> a
dcats:DataService ;
  dct:title "FDSN Dataselect - ORFEUS Data Center" ;
  dct:description "FDSN Standard webservice to download
waveform data" ;
  dcat:contactPoint <http://orcid.org/0000-0001-0002-
0003/contactPoint>;
  dcat:endpointDescription <https://www.epos-
eu.org/Seismology/Operation/001>
  ...
.
```

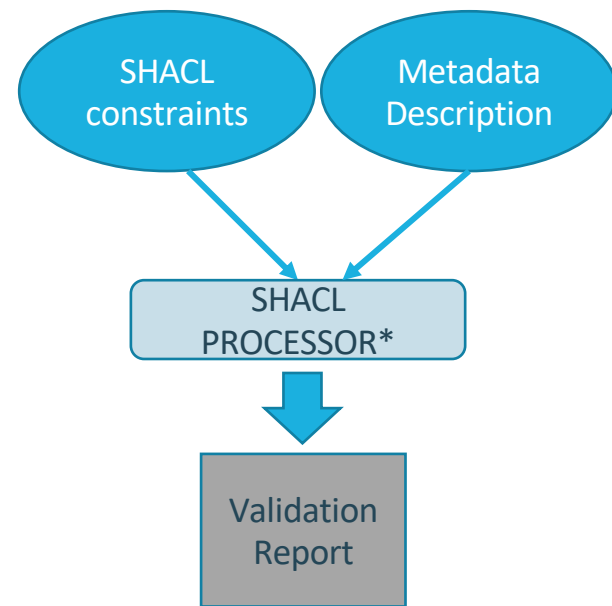
OPERATION

```
<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/dataselect/1/query/{?starttime,endtime}"^^xs
d: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 Metadata Validation

- EPOS-DCAT-AP enables integrity checks and validation of metadata content by using **SHACL constraints** (https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/epos-dcat-ap_shapes.ttl);
- SHACL is 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.



* SHACL Processors available online:
<https://shacl.org/playground/>
<https://shacl-play.sparna.fr/play/validate>



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

METADATA SERIALIZED IN TURTLE FORMAT

```

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

```

METADATA IN TURTLE FORMAT

```

<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" ; ✗
.

```



Metadata Quality check



- Filling in only the mandatory or recommended fields is not sufficient to ensure high-quality metadata.
- It is necessary to define a checklist of fields for verifying the accuracy, completeness, consistency, and relevance of metadata in a specific context.
- In addition, a set of fields are typically defined to assess compliance with the FAIR principles
(i.e., **Findability, Accessibility, Interoperability, Reusability**)



EPOS Metadata Quality check



Categories

- Dataset entities need to be linked to one of the 'dcat:theme' defined for the EPOS context. These categories are used to categorize TCS assets, ensuring that they are presented as organized results in the Portal.

Spatio-temporal coverage

- These are used to enable ICS to query TCS web services in an interoperable way. Additionally, they play a key role in discovering data and products

Institutions/organizations

- Datasets and Web services need to be linked to the Organization entities in order to define data and service providers, respectively.

Keywords

- Keywords need to be associated with Dataset and Web service entities. Their proper definition is crucial for the correct functioning of the discovery features.



How Metadata contributes to data discovery features



Data Search



Search Refinement



Data Visualization



Data Download

Data Search

The screenshot displays the EPOS 1.0.23 data search interface. On the left, a sidebar contains search filters and a list of providers. The main search area includes a search bar with 'earthquakes' entered, filters for Providers and Spatial, and options for Geolocation (Italy), date ranges (Last Month, Last Week, Last Day), and Data and Service Providers (1 item(s) selected). Below the search area, a map shows the location of the search results. The details panel on the right provides information about the selected event: Parameters of historical earthquakes (1000-1899) - FDSN event. The details panel includes sections for Name, Domain, Categories, Description, Spatial Coverage, Temporal Coverage, Persistent Identifier(s), License, Keywords, Update Frequency, Quality Assurance, Data Provider(s), Further information, and Service Details. A blue box highlights the 'dcat:contactPoint' property in the details panel, which is linked to a 'Contact Us' button. The interface also features a search bar, filters, and a sidebar with a list of providers.

Search Filters:

- Free text search: earthquakes
- Filters: Providers, Spatial
- Geolocation: Italy
- Date range: YYYY-MM-DD HH:mm:ss (Last Month, Last Week, Last Day)
- Data and Service Providers: 1 item(s) selected
- Data Visualization: Clear All

Search Results:

- Seismology
- Parameters of historical earthquakes (1000-1899) - FDSN event
- Categories: Seismological products service... > Earthquake parameters
- Visible on: Map Table
- Status:
- Advanced search filters (4 of 12): Coordinates: 47,12 36,62 6,75 18,48; Date range: 1000-01-01 00:00:00 to 1899-12-31 23:59:59; Limit: 300; Circular search center latitu...; Circular search center longi...; Circular search radius (in km); Event ID; Include all magnitudes

Details Panel:

- dcat:contactPoint** (highlighted) - Contact Us
- Name:** Parameters of historical earthquakes (1000-1899) - FDSN event
- Domain:** Seismology
- Categories:** Seismological products services > Earthquake parameters
- Description:** The distribution of event parameters via the FDSN-event web service is the main and preferred way to access historical earthquake data archived in AHEAD.
- Spatial Coverage:** [Map]
- Temporal Coverage:** 1000-01-01 00:00:00 - 1899-12-31 23:59:59
- Persistent Identifier(s):** 10.6092/INGV.IT-AHEAD
- License:** https://www.emidius.eu/AHEAD/description.php#copyright
- Keywords:** seismology; magnitude; earthquake; event; catalogue; seismicity
- Update Frequency:** http://purl.org/cld/freq/irregular
- Quality Assurance:** https://www.emidius.eu/AHEAD/data_quality_assurance.php
- Data Provider(s):** INGV - Istituto Nazionale di Geofisica e Vulcanologia
- Further information:** https://www.epos-eu.org/tcs/seismology
- Service Details:** [Up arrow]



Search Refinement

The screenshot displays the EPOS (European Plate Observing System) search interface. The top left shows the EPOS logo and version 1.0.23. The search bar contains the text "earthquakes". Below the search bar, there are filter tabs for "Providers" and "Spatial". The geolocation is set to "Italy". The search criteria include a date range from "1000-01-01 00:00:00" to "1899-12-31 23:59:00" and a circular search radius of 300 km. The search results are displayed as a map of Italy with numerous yellow star markers representing earthquake events. A blue box highlights the map area. The interface also includes a sidebar with various icons and a bottom navigation bar with a star icon and a page indicator.

`dcat:Distribution`

`dcat:DataService`
`hydra:Operation`
`hydra:IriTemplate`
`hydra:IriTemplate Mapping`



Data Visualization

The screenshot displays the EPOS web application interface. On the left, there is a search sidebar with filters for 'earthquakes', 'Providers', 'Spatial', and 'Geolocation' (set to Italy). Below the filters, there are options for time ranges (Last Month, Last Week, Last Day) and a 'Data and Service Providers' section showing 1 item selected. The main content area is titled 'Parameters of historical earthquakes (1000-1899) - FDSN event'. It features a table with columns: Origin time, Epicentral area, Epicentre latitude, Epicentre longitude, Evaluation mode, Magnitude value, Magnitude uncertainty, and Magnitude type. The table shows five rows of earthquake data. Below the table, there is a 'Trace Selector' for 'TABOO Vp/Vs' with options for 'error_max', 'error_min', and 'vpvs_value'. A graph shows the trace data over time from May 11, 2014, to June 8, 2014.

Origin time	Epicentral area	Epicentre latitude	Epicentre longitude	Evaluation mode	Magnitude value	Magnitude uncertainty	Magnitude type
1899-12-26T10:...	Northern Italy	44.475	7.517	manual	4.22	±0.38	Mw
1899-12-19T21:...	Slovenia	45.500	15.250	manual	3.87	±0.50	Mw
1899-11-15T23:...	Northern Italy	45.404	11.440	manual	4.31	±0.30	Mw
1899-11-13T21:...	Slovenia	45.830	14.250	manual	3.87	±0.50	Mw
1899-11-07T12:...	Switzerland	46.063	7.181	manual	2.60	±0.67	Mw

Several software converters are described by using `schema:SoftwareApplication` and `schema:SoftwareSourceCode` classes



Data Download

The screenshot shows the EPOS 1.0.12 web interface. On the left, a search sidebar is visible with filters for 'earthquakes', 'Providers', and 'Spatial'. The 'Geolocation' is set to 'Italy'. Below this, there are date range options: 'Last Month', 'Last Week', and 'Last Day'. A 'Data and Service Providers' section shows '1 item(s) selected'. The 'Data Visualization' dropdown is set to 'Map'. The main content area is titled 'Seismology' and shows 'Categories: Seismological products service... -> Earthquake parameters'. The 'Visible on' is set to 'Map Table' and the 'Status' is 'On'. Under 'Advanced search filters (4 of 12)', the coordinates are '47,12 36,62 6,75 18,48' and the date range is '1000-01-01 00:00: 1899-12-31 23:59:'. There are also options for 'Limit the no. of output entries' (set to 300), 'Circular search center latitu...', 'Circular search center longi...', 'Circular search radius (in km)', 'Event ID', 'Include all magnitudes', 'Include all origins', 'Maximum magnitude (Mw)', 'Minimum magnitude (Mw)', and 'Output ordering time'. At the bottom of the sidebar are 'Set to defaults' and 'Apply' buttons. The main map area shows a map of the Mediterranean region with a blue box highlighting a specific area around Sicily and Malta. A modal window titled 'Parameters of historical earthquakes (1000-1899) - FDSN event Files available for download' is open over the map. The modal contains a table with columns for 'File Name', 'File Format', 'Download', and 'Copy URL'. The table lists four entries, all labeled 'raw service response' with different file formats: 'text', 'json', 'xml', and 'application/epos.geo+json'. Each entry has a download icon and a copy URL icon. Below the table, there is a '0 selected' indicator, a 'Filter' input field, 'Results per page: 10', 'Page 1 of 1', and 'Download files' and 'Generate URLs' buttons.

Distribution >
dct:format
Hydra:Operation >
schema:encodingFormat



SHAPEness: a metadata management tool

- It is a Java desktop application conceived to help users creating and updating EPOS-DCAT-AP metadata descriptions.
- It provides a graph-based interface which allows users to:
 - Create or update metadata as data graphs;
 - validate metadata against SHACL constraints;
 - serialize metadata in RDF/Turtle format;
 - deposit metadata turtle file to a GitLab repository.
- The latest version of SHAPEness is available on GitHub at <https://epos-eu.github.io/SHAPEness-Metadate-Editor/gitpage/index.html>

SHAPEness Metadata Editor

Graph View

Palette View

RDF/TURTLE View

Outline View

Properties View

The screenshot displays the SHAPEness Metadata Editor interface, which is divided into several panels:

- Graph View (Top Center):** A central graph showing relationships between various metadata classes. The 'Dataset' class is the central node, connected to 'ContactPoint', 'Location', 'Concept', 'Distribution', and 'PeriodOfTime'. 'ContactPoint' is further connected to 'Person' and 'Organization'. 'Location' is connected to 'Concept'. 'Concept' is connected to 'ConceptScheme'. 'Distribution' is connected to 'Concept'. 'PeriodOfTime' is connected to 'Concept'. The graph also shows relationships between 'Concept' and 'ConceptScheme' via 'skos:inScheme'.
- Palette View (Top Left):** A vertical list of metadata classes with corresponding icons, including Address, ApiDocumentation, Catalog, CatalogRecord, Concept, ConceptScheme, ContactPoint, ControlAction, CreativeWork, Dataset, Distribution, Equipment, Facility, Identifier, InTemplate, InTemplateMapping, Location, Operation, Organization, and PeriodOfTime.
- Outline View (Bottom Left):** A hierarchical tree view of the metadata instance. It shows the 'Dataset' node expanded to show its children: 'Distribution', 'Location', 'Organization', 'PeriodOfTime', and 'Person'. Each node is accompanied by its URI.
- Properties View (Bottom Center):** A form for editing the properties of the selected 'Dataset' instance. It is divided into 'Mandatory properties' and 'Recommended properties'.
 - Mandatory properties:**
 - dct:title:** European Archive of Historical Earthquake Data (AHEAD)@en
 - dct:description:** AHEAD, the European Archive of Historical Earthquake Data (AHEAD)-1899, is a distributed archive aiming at preserving, inventorying and making available, to investigators a the earthquake history of Europe, such as papers, reports, Macroseismic Data Points (MDPs) in different scales (EMS-98, MSK, or MCS), and parametric catalogues atThe / regionally managed data archives. At present, eight regional, online macroseismic archives team up in AHEAD covering Italy, France, Switzerland, Spain, Catalunya, Belgium covered by a regionally managed archive is retrieved by checking out reliable and published scientific literature. AHEAD inventory seismological data organized either in re covered by a regionally managed archive is retrieved by checking out reliable and published scientific literature. AHEAD inventory seismological data organized either in re
 - dct:identifier:** https://doi.org/10.6092/INGV.IT-AHEAD
 - Recommended properties:**
 - dc:theme:** Concept-9e52db6a762e345acb348469a51d744; Concept-5a0956e817991cb827318f48f13f; Concept-540e099ec1a8415d70cb3152b365a808
 - publisher:** Organization-35bf3f8256f22c4c408c7a5ee61734ea
 - dc:at:keyword:** Historical seismology; Earth sciences and geology
- RDF/Turtle View (Top Right):** A text area showing the RDF/Turtle representation of the metadata instance. It starts with a prefix declaration for 'https://www.epos-eu.org' and contains various triples describing the dataset's title, description, identifier, theme, publisher, and keywords.



Q&A



References

- **EPOS-DCAT-AP:** <https://epos-eu.github.io/EPOS-DCAT-AP/v3/>
- **SHAPEness Metadata Editor:** <https://epos-eu.github.io/SHAPEness-Metadata-Editor/gitpage/index.html>
- **DCAT-AP v3:** <https://www.w3.org/TR/vocab-dcat-3/>
- **RDF:** <https://www.w3.org/RDF/>
- **TURTLE:** <https://www.w3.org/TR/turtle/>
- **SHACL:** <https://www.w3.org/TR/shacl/>
- Bailo, D. et al. (2023). **Integrated Access to Multidisciplinary Data Through Semantically Interoperable Services in a Metadata-Driven Platform for Solid Earth Science.** In: Garoufallou, E., Vlachidis, A. (eds) Metadata and Semantic Research. MTSR 2022. Communications in Computer and Information Science, vol 1789. Springer, Cham. https://doi.org/10.1007/978-3-031-39141-5_20
- Paciello, R., Trani, L., Bailo, D., Sbarra, M. (2023). **EPOS-DCAT-AP 2.0 – State of Play on the Application Profile for Metadata Exchange in the EPOS RI.** In: Garoufallou, E., Vlachidis, A. (eds) Metadata and Semantic Research. MTSR 2022. Communications in Computer and Information Science, vol 1789. Springer, Cham. https://doi.org/10.1007/978-3-031-39141-5_21
- Paciello, R., Trani, L., Bailo, D., Vinciarelli, V., Sbarra, M. (2023). **SHAPEness: A SHACL-Driven Metadata Editor.** In: Garoufallou, E., Vlachidis, A. (eds) Metadata and Semantic Research. MTSR 2022. Communications in Computer and Information Science, vol 1789. Springer, Cham. https://doi.org/10.1007/978-3-031-39141-5_23.
- Bailo, D. et al. **Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures.** *Front. Earth Sci.* **8**, (2020).
- D. Bailo, K.G. Jeffery, K. Atakan, L. Trani, R. Paciello, V. Vinciarelli, J. Michalek, A. Spinuso, **Data integration and FAIR data management in Solid Earth Science**, in Annals of Geophysics, vol. 65 No. 2 (2022), Special issue: EPOS a Research Infrastructure in solid Earth: open science and innovation. Doi: <https://doi.org/10.4401/ag-8742>.



Thank you for your attention!

Geo-INQUIRE is a joint effort of 51 institutions



Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call.