

2018 
EUROPEAN YEAR
OF CULTURAL
HERITAGE
#EuropeForCulture

Europeanana



Data quality in Europeanana

Pablo Uceda Gomez | National workshop Bulgaria

30 May 2018



Co-financed by the European Union
Connecting Europe Facility

The Madrid Codices of Leonardo da Vinci (Vol. I), f. 7r
Leonardo da Vinci

c.1493, Biblioteca Nacional de España
Spain, CC BY-SA

Data Quality?

- We are talking about Data Quality **in the context of Europeana**.
- For the moment we have no framework for metadata quality but **we are working on a metadata component for the EPF**.
- In our context, we need data that can be **found, understood** and **used** by users and machines.
- Our definition of quality is closely linked to the **User Scenarios** defined by the **Data Quality Committee**. The basic concept is that the more you can do with the data the better for different kinds of users.

[Our mission](#)[Priorities](#)[Network Association](#)[Services](#)[Resources](#)[Join us](#)[Blog](#)[Europeana Collections](#)[Get an API key](#)[Jobs](#)[Press](#)[Contact us](#)

Data Quality Committee

Quality is a key priority for our whole community! The Data Quality Committee works to address key data quality issues over time.

Formally defined as a [Europeana Network](#) and [EuropeanaTech](#) Working Group, the Data Quality Committee is a standing committee that works on the various facets of the data quality challenge over time with a particular focus on reuse and discovery of cultural heritage scenarios. We believe it is crucial to tackle data quality issues at every level of the data exchange chain from its creation to its publication. We have gathered together experts from various backgrounds (metadata experts, software developers, search and retrieval experts) to help us capture all the issues.

Data Quality Mission Statement

Work areas as well as specific tasks will be defined and prioritized as the Committee sees fits and regularly reported and submitted for ratification to the community, notably the Europeana Aggregator Forum. Items

Discovery - User scenarios and their metadata requirements

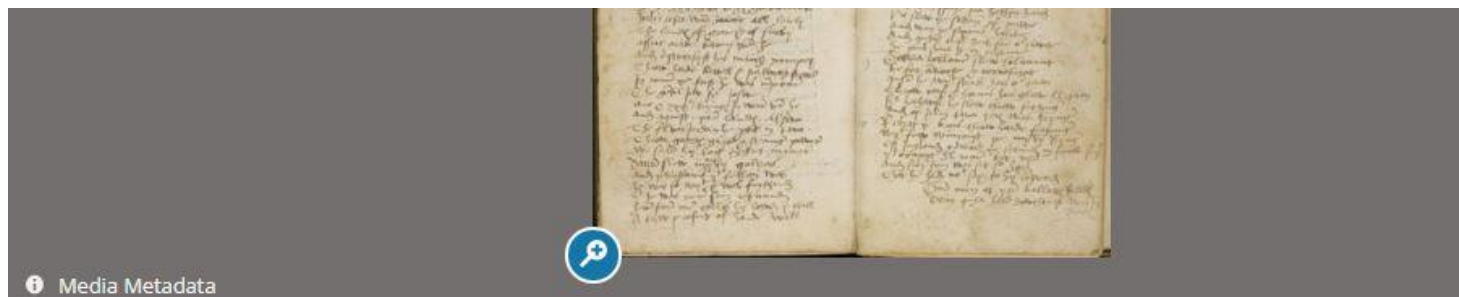
- **1. Basic Retrieval with high precision and recall**
- **2. Cross-language recall**
- **3. Improved language based facets**
- 3a. Improved facets based on language of metadata
- 3b. Improved facets based on language of content
- **4. Entity-based facets**
- 5. Browse by date or timespan based facets
- 6. Browse by subjects and types
- 7. Browse by Agents
- 8. Browse by Places
- 9. Browse by Events
- 10. Entity-based knowledge cards and pages
- 11. Entity-based autocompletion
- 12. Categorised similar items
- 13. Diversification of results

Each scenario has a number of EDM elements that are required to activate it...

- 6. Browse by subjects and types

edm:ProvidedCHO	dc:subject (with a concept), dc:format, dc:type, dcterms:medium
skos:Concept	skos:prefLabel, skos:broader, skos:narrower, skos:exactMatch, skos:closeMatch, skos:related

Basic retrieval needs descriptive metadata...



Title

Unknown and Unknown | Unknown

FIND OUT MORE

[View at University Of t](#)

Description

Unknown

Call Number - 0013012

Shelf Mark - Dc.7.63

CAN I USE IT?

Free Re-use

 CC BY 

EXPLORE BY COLOUR



People

Creator: Unknown

SHARE

In an international and multilingual platform language description is essential....

- All EDM elements supporting literals SHOULD be provided with language tags. We recommend the use of two letter ISO 639-2

`<dc:subject xml:lang="es">Geografía</dc:subject>`

`<dc:subject xml:lang="en">Geography</dc:subject>`

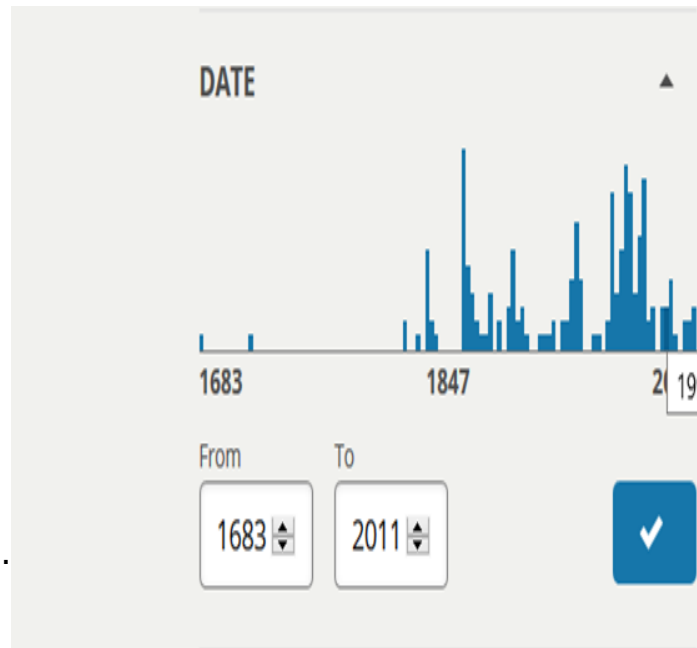
- Using EDM elements in combination with (i.e. which link to) a contextual entity with multilingual features is **HIGHLY RECOMMENDED**.

`<dc:subject rdf:about="http://vocab.getty.edu/aat/300054388" />`

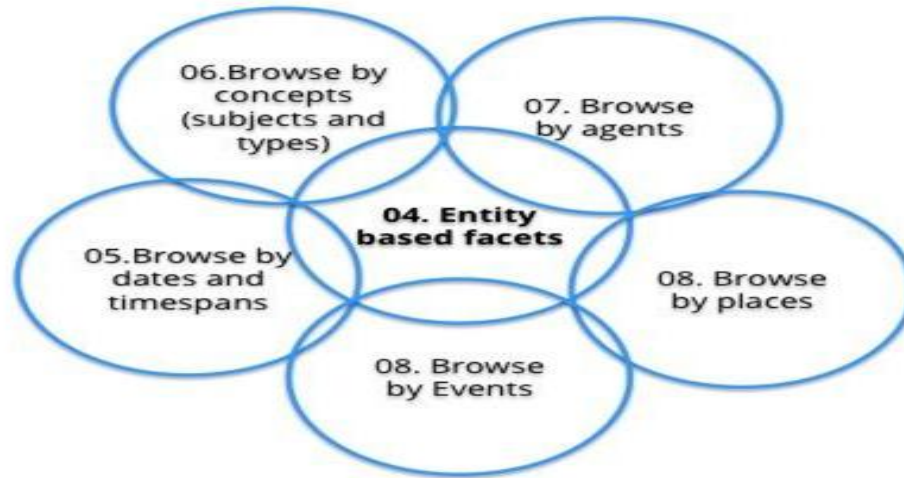
```
<rdf:RDF>
  <gvp:Subject rdf:about="http://vocab.getty.edu/aat/300054388">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <rdf:type rdf:resource="http://vocab.getty.edu/ontology#Concept"/>
    <rdfs:label xml:lang="en">geography</rdfs:label>
    <rdfs:label xml:lang="en">geographical</rdfs:label>
    <rdfs:label xml:lang="zh-hant">地理學</rdfs:label>
    <rdfs:label xml:lang="zh-latn-pinyin-x-hanyu">di li xue</rdfs:label>
    <rdfs:label xml:lang="zh-latn-pinyin-x-notone">di li xue</rdfs:label>
    <rdfs:label xml:lang="zh-latn-wadegile">ti li hsüeh</rdfs:label>
    <rdfs:label xml:lang="nl">geografie</rdfs:label>
    <rdfs:label xml:lang="es">geografia</rdfs:label>
    <rdfs:seeAlso rdf:resource="http://www.getty.edu/vow/AATFullDisplay?find=&logic=AND&note=&subjectid=300054388"/>
    <dc:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1988-01-01T00:00:00</dc:created>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1991-02-01T00:00:00</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">1993-04-13T00:00:00</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2001-07-26T22:00:17</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-03-29T10:09:02</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2010-04-23T02:25:22</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2011-01-07T16:20:45</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-03-29T10:09:03</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2011-01-07T16:20:47</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2007-05-23T12:26:34</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2001-07-26T22:15:09</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-03-16T14:09:07</dc:modified>
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-03-17T21:45:43</dc:modified>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5000017187"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5000017189"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5001074044"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5001074045"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5002343120"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5002750371"/>
    <skos:changeNote rdf:resource="http://vocab.getty.edu/aat/rev/5003446859"/>
```


Dates are important for heritage...

- **Normalisation:** Dates should be fully and consistently normalised to follow standards such as the XSD date-time data types. For instance, dates expressed in styles like “**490 avant J.C**” **that are inherently language dependent should be avoided**, as these are very difficult to normalise (e.g. this should be represented as “-0490”^^xsd:gYear)
- **Specialisation:** Use specialisations (e.g. dcterms:created) of dc:date whenever possible.



Entity based facets could allow us to increase the search options...



.... but for that we need structured data in the form of entities...

- You can map the data from your vocabularies or authority files to one of our edm classes.

```
<skos:Concept rdf:about="http://hispana.mcu.es/lod/oai:biblioteca.ranf.com:10002009511#ent3">  
  <skos:prefLabel xml:lang="es">Farmacia</skos:prefLabel>  
  <skos:prefLabel xml:lang="en">Pharmacy</skos:prefLabel>  
  <skos:closeMatch rdf:resource="http://dbpedia.org/resource/Pharmacy"/>  
  <skos:closeMatch rdf:resource="http://bibliotecavirtual.ranf.com/aut/RANFA20090064324"/>  
</skos:Concept>
```

You can provide a link to a LOD vocabulary like GETTY AAT, FIAF , GEONAMES....

```
<dc:type @rdf:resource='http://vocab.getty.edu/aat/300028702' >
```

Classifications	Type: inscriptions , rubbing (transferring techniques) Subject: Monumental writing , Epigraphy Medium: Squeeze paper
Entities	SUBJECTS, RESOURCE TYPES, GENRES AND FORMS (CONCEPTS) ▾ Concept Term http://vocab.getty.edu/aat/300028702 Concept Label ["Inscript"] (de); ["inscriptions"] (en); ["inscripciones"] (es); ["inscripties"] (nl); Broader Concept http://vocab.getty.edu/aat/300026031 Concept Term http://vocab.getty.edu/aat/300178924 Concept Label ["rubbing (transferring techniques)"] (en); ["frotamiento"] (es); ["een rubbing maken"] (nl); Broader Concept http://vocab.getty.edu/aat/300155207

If the above is not possible we can enrich your strings with the European Semantic Enrichment

- Named Entity Recognition during the enrichment process complements the source data with information about concepts and agents from DBpedia
- Allows for automatic translation terms when entities have multiple language labels
- Connecting artworks from institutions all over Europe

An object with linked entities



An object purely with metadata and no linked entities



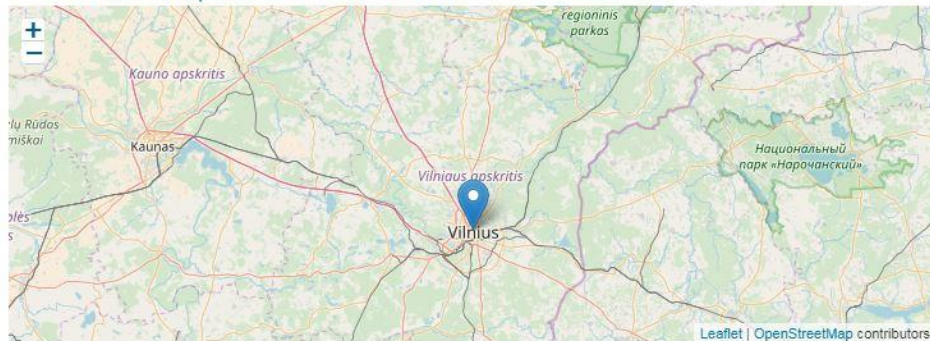
Enriching location information

If a name of a place is matched to the Geonames database, coordinates and a mini-map are added to the data

Location

Location: Vilnius, Vılna, Βίλνιους

Current Location: <http://lod.b3kat.de/bib/DE-12>



VILNIUS 54.68916° North, 25.2798° East

Actions from DPS

- Advise on how to map your data to EDM
- Data Quality Planning: year objectives for metadata improvement
- Analysis on how to improve the legacy
- Working on a metadata component for the EPF



Woman Reading a Letter | Johannes Vermeer
1663, Rijksmuseum
Netherlands, Public Domain

What can you do?

- Provide data in EDM
- Provide as much **meaningful metadata** as possible
- Use **language attributes** for literal values
- Provide entities or **links to LODs** (f.e. Getty AAT) instead of strings for concepts, agents and places
- Normalise your dates and avoid language dependant expressions
- Use the **edm:webresource class** to describe your digital resources

Benefits on Europeana

- Increased **metadata quality** improves the **discoverability** of objects.
- Increased discoverability of objects leads to increase of users **finding** the object, and increased **reuse** of the material (*provided the rights statements allow this*):
 - Europeana Exhibitions
 - Use in third party apps
 - Use in Europeana **Research, Education** and **Creative Industries**
 - Social media
 - Use on Wikimedia
 - Etc.



Lithuanian Girl with Palm Sunday Fronds

Kanutas Ruseckas

1844, Lithuanian Art Museum, Public Domain

2018 
EUROPEAN YEAR
OF CULTURAL
HERITAGE
#EuropeForCulture



europæana

10th ANNIVERSARY

The Music Lesson
Louis Moritz
1808, Rijksmuseum
Netherlands, Public Domain