Data Aggregation in the Data Space for Cultural Heritage

JOCHEN VERMEULEN - 21 JUNE 2023

jochen.vermeulen@europeana.eu

Funded by the European Union

Poznán (1924) - Bibliothèque Nationale de France - Public Domain

DATA SPACE FOR CULTURAL HERITAGE

The Europeana Initiative is at the heart of the *Common European Data Space for Cultural Heritage*.

At the Europeana Foundation we share and promote Europe's digital cultural heritage to be used and enjoyed by everyone for learning, for work, or just for fun.

europeana

CONSORTIUM

Capgemini
datoptron
inception
A START-UP COMPANY
PHOTOCONSORTIUM
AIT
AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY
Time Machine
europeana foundation
Facts & Files
Think History!
EUROPEAN FASHION HERITAGE ASSOCIATION
Michael Culture Association
europeana
PSNC
ARCHIVES PORTAL EUROPE
europeana network association
europeana aggregators forum
DFF
DEUTSCHES FILMINSTITUT FILMMUSEUM
EuroClio
Inspiring History and Citizenship Educators
SOUND AND VISION
carare
European Schoolnet
DARIAH-EU
europeana

THE AGGREGATION PIPELINE

europeana

ANYONE CAN ACCESS

64+ million

CULTURAL HERITAGE ITEMS THROUGH EUROPEANA

europeana

VISITED MORE THAN

6 MILLION

TIMES EVERY YEAR

europeana

WHAT CAN YOU FIND?

36.8M	**IMAGES**	Photos, Maps, Paintings, Comics, Diagrams, Designs, Fashion, Drawings, Architecture, Artifacts, Icons, Symbols, Portraits, etc.
26.1M	**TEXTS**	Documents, Newspapers, Stories, Blogs, Manuscripts, Histories, Sheet music, etc.
721K	**AUDIO CLIPS**	Music, Speeches, Recordings, Radio clips, etc.
532K	**VIDEO CLIPS**	Movies, Trailers, News clips, Broadcasts, etc.
11K	**3D RECORDS**	Artifacts, Architecture, Archaeology, etc.

europeana

DATA PROCESSING AT EUROPEANA

europeana

METIS/ECLOUD STACK

- ➔ All code is Java, open-source, available on Github.
- ➔ All components are (to be) containerized and scalable (Docker/Kubernetes).

europeana

USER EXPERIENCE

Hispana_BibliotecaRioja

Provider	Hispana
Number of items published	Not published yet
Last published	07/07/2022 - 01:53
Number of items depublished	Not depublished
Last depublished	Not depublished

View in Preview | View in Collections

PROCESS MEDIA Running 914 / 4659 20% 20/06/2023-14:53 STARTED BY Adina Ciocoiu Cancel Log Report

WORKFLOW STEP	PROCESSED / RECORDS	START TIME	END TIME	ACTIONS
Pr Preview	0 / 0	—	—	
M Process Media	914 / 4659	19/06/2023 14:43	—	Report
E Enrich	4659 / 4659	19/06/2023 14:32	19/06/2023 14:43	
N Normalise	4659 / 4659	19/06/2023 14:28	19/06/2023 14:32	
Vi Validate (EDM internal)	4659 / 4659	19/06/2023 14:27	19/06/2023 14:28	
T Transform	4659 / 4659	19/06/2023 14:25	19/06/2023 14:27	
Ve Validate (EDM external)	4659 / 4659	19/06/2023 14:23	19/06/2023 14:25	
Ho Import OAI-PMH Full Processing	4659 / 4659	19/06/2023 14:13	19/06/2023 14:23	

SEE FULL PROCESSING HISTORY

Dataset Information
Last date of harvest
19/06/2023 - 14:23

Workflow | Depublication | Mapping | Raw XML

Processing history
Last updated
20/06/2023 - 14:53

Workflow started on: 07/11/2022 - 14:30 | Plugin: Preview | Compare with

Search for record by eCloud or europeana id

22PGWSCNS2NBTNIHMSSC4ZHPQVMWGNUD7UPNKZVLZ4WVRLFSIICQ

Preview

```
1 <?xml version="1.0" encoding="UTF-8"?>
```

europeana

EUROPEANA DATA MODEL (EDM)

- A linked open data model equipped with an XML schema.
- Retains provenance and distinguishes between source data and data we're adding ourselves.
- Is built on W3C standards for the Semantic Web (RDF, SKOS, OWL), Dublin Core, among others.

europeana

ENRICHMENT

Adding **context** to the record. This:

- Improves findability
- Encourages browsing
- Adds language support

```
<dc:subject>music</dc:subject>
```

```
<dc:subject rdf:resource="http://data.europeana.eu/12345"/>

<skos:Concept rdf:about="http://data.europeana.eu/12345">
  <skos:prefLabel xml:lang="en">Music</skos:prefLabel>
  <skos:prefLabel xml:lang="fr">Musique</skos:prefLabel>

  [ OTHER PROPERTIES OF CONTEXTUAL OBJECT ]

</skos:Concept>
```

europeana

NORMALIZATION

```
<dc:date>January, 1765</dc:date>
```

```
<dc:date rdf:resource="#1765-01"/>

<edm:TimeSpan rdf:about="#1765-01">
  <skos:prefLabel xml:lang="zxx">1765-01</skos:prefLabel>
  <dcterms:isPartOf rdf:resource="http://data.europeana.eu/timespan/18"/>
  <edm:begin>1765-01-01</edm:begin>
  <edm:end>1765-01-31</edm:end>
  <skos:notation rdf:datatype=
    "http://id.loc.gov/datatypes/edtf/EDTF-level1">1765-01</skos:notation>
</edm:TimeSpan>
```

Normalizing the record:

- Converts dates and language codes to ISO and EDTF standards.
- Also performs various clean-ups on values and removes duplicates.
- Improves findability.

europeana

MEDIA PROCESSING

```
<edm:hasView rdf:resource="http://.../12345.mpg"/>
```

```
<edm:WebResource rdf:about="http://.../12345.mpg">
  <edm:codecName>mpeg2video</edm:codecName>
  <ebucore:hasMimeType>video/mpeg</ebucore:hasMimeType>
  <ebucore:fileByteSize>19480576</ebucore:fileByteSize>
  <ebucore:duration>40400</ebucore:duration>
  <ebucore:width>720</ebucore:width>
  <ebucore:height>576</ebucore:height>
  <ebucore:bitRate>3853958</ebucore:bitRate>
  <ebucore:frameRate>25.0</ebucore:frameRate>
</edm:WebResource>
```

Adding **technical metadata** to the record. This:

- Creates site thumbnails.
- Allows searching and filtering.
- Powers our media viewer.

europeana

BUILDING THE DATA SPACE

europeana

Wierzbak
Stacja telegrafu iskrowego
Dworzec
Stacja komp.
Watu Leszczyńskiego
Plac
Wielka
Nowa
Wrzesińska
Cment.
Targi Poznańskie
Cment. żyd.
Dworzec zachodni
Dworzec główny
Cment. ew.
Watu Jadwigi
Watu Kroł. Jadwigi
Przystań
Błonia Wildeckie
Watu Kazimierza W.
Mostowa
Skarbowa
Ogrodowa
Watu Łarenczuka
Ul. Poznańska
Ul. Sienkiewicza
Ul. Zwierzyńiec
Ul. Samarzewskiego
Ul. Grudnia Rotnosi
Ul. Chwaliszewska
Ul. Raczyńska
Ul. Kościelna
Ul. Wrzesińska
Ul. Przemysłowa
Ul. Kazimierzewska
Ul. Łarenczuka
Ul. Jeckiego
Ul. Wypka
Ul. Kupiek
Ul. Kupiek
Ul. Kupiek