Entitifying Europeana: Building an ecosystem of networked references for Cultural Objects

Hugo Manguinhas, Valentine Charles, Antoine Isaac, Tim Hill
Europeana Foundation
What is Europeana?

The Platform for Europe’s Digital Cultural Heritage

We aggregate metadata:

- From all EU countries
- ~3,500 galleries, libraries, archives and museums
- More than 53M objects
- In about 50 languages
- Huge amount of references to places, agents, concepts, time
Europeana Linked Data Strategy

Our efforts and lines of work

- Europeana Data Model (EDM) offers a base for linking data
- We apply automatic enrichment to link source data to reference data
- We encourage data providers to contribute their own vocabularies so that we can benefit from data links made at data providers’ level
- We encourage alignment activities between domain vocabularies

Significant progress have been made, most of it presented in past SWIB!
As a cornerstone for our strategy we are building an "Entity Collection"

- A service that acts as a centralized point of reference and access to data about contextual entities
- Caching and curating data from the wider Linked Open Data cloud
- A sort of Europeana "knowledge graph"
Europeana Linked Data Strategy

Motivation

• Improve user experience
  • Support better ways of searching and navigating through the collections, eliminating ambiguity and clarifying the meaning of descriptions
  • Adapt better to the language of the user

• by improving the interlinking of data
  • Brings more context to the objects
  • Alleviates polysemy issues
  • Expands language coverage
  • Contributes to build a web of data ('knowledge graph') that third parties can use to improve their users' experience
### The Entity Collection

**Use Cases**

<table>
<thead>
<tr>
<th>Enrichment of Provider’s Data</th>
<th>Crowdsourcing</th>
</tr>
</thead>
<tbody>
<tr>
<td>● A controlled vocabulary to help identify named references to entities</td>
<td>● Objects can be annotated with references to entities</td>
</tr>
<tr>
<td></td>
<td>● A controlled vocabulary for client applications</td>
</tr>
</tbody>
</table>

**Europeana Collections Portal**

- **Findability**: users can look for entities, not only records (Entity-Based Search)
- **Understandability**: Entity Pages group and present all assertions about an entity
- **Exploration**: Navigation along relationships becomes possible

**Republication for Re-use**

- Entities can be republished as an open source to the community
The Entity Collection

What can it enable?

Semantic auto-completion

Entity based facets

Semantic and Metadata annotations

Entity Pages

Food and Drink Project

Pundit Annotation Client

Google Knowledge Card

SWIB16 - Entitifying Europeana: Building an ecosystem of networked references for Cultural Objects

CC BY-SA
The Entity Collection

*How do we choose our target vocabularies?*

As defined in the recent *Europeana Tech Task Force on enrichment and evaluation (presented last year)*, we consider the following criteria when selecting a vocabulary:

- Properly documented and supported by a community
- Technically available on the web according to the Linked Data best practices and recipes
- Available under an open licence
- Multilingual
- Abide to a minimal ontological commitment principle
- Apply the best practices and standards for the representation, structure and description of vocabularies
- Well-connected internally and externally to other vocabularies (preferably spine vocabularies)
The Entity Collection

Which target vocabularies are we using?

For historical reasons, the target vocabularies correspond to the ones being used for Semantic Enrichment (as of November 2016):

- **Places**
  a subset of **Geonames**, corresponding to places which are part of European countries and of some specific feature classes.

- **Agents**
  a subset of **DBpedia** corresponding to most of the instances of dbp:Artist with some exceptions, and integrated from **49 DBpedia language editions**.

- **Concepts**
  a subset of **DBpedia** corresponding to a handful of concepts matching the needs from Europeana Collections.

- **Time Spans**
  The chronological periods from **Semium Time**.
The Entity Collection

Contribution to multilingual coverage

Entities effectively used to enrich Europeana Objects

Entities present in the Entity Collection
The Entity Collection

Are these target vocabularies enough?

- Not enough coreferencing information to other vocabularies
  - particularly to the ones we receive from data providers (e.g. musical instruments, MIMO)
- Labels and values are not always accurate and normalized
  - need for better reference data (e.g. VIAF)
- Missing relevant information
  - e.g. roles and professions
- Need to expand coverage to other types of entities
  - namely Works and Events
The Entity Collection

Challenges

Investigate and design strategies for:

- Integrating new vocabularies that can further improve
  - entity descriptions and multilingual coverage (e.g. VIAF)
  - linking between entities (e.g. Wikidata)

- Integrating alignments, in particular:
  - links between local/domain vocabularies to pivot vocabularies

- Supporting manual curation of existing and new entities

- Keeping up-to-date the information collected from external sources
The Entity Collection

Our roadmap for the next years

✔ • Mint Europeana URLs for Entities and update internal references
✔ • Make entity services and data available via an API
  • Make use of the API in the Collections Portal
  • Implement support for new vocabularies and entity types
## Entity Collection - REST API

This Swagger API console provides an overview of and interface to the Entity Collection REST API. For more help and information, head to our comprehensive [online documentation](#).

**Entity retrieval**

<table>
<thead>
<tr>
<th>Method</th>
<th>URI</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>/entity/{type}/{namespace}/{identifier}</td>
<td>Retrieve a known entity</td>
</tr>
<tr>
<td>GET</td>
<td>/entity/{type}/{namespace}/{identifier}.jsonld</td>
<td>Retrieve a known entity</td>
</tr>
</tbody>
</table>

**Discovery API Search Controller**

<table>
<thead>
<tr>
<th>Method</th>
<th>URI</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GET</td>
<td>/entity/suggest</td>
<td>Suggest entities for the given text query. Supported values for type: Agent, Place, Concept, Timespan, All</td>
</tr>
<tr>
<td>GET</td>
<td>/entity/suggest.jsonld</td>
<td>Suggest entities for the given text query. Supported values for type: Agent, Place, Concept, Timespan, All</td>
</tr>
</tbody>
</table>

[BASE URL: /, API VERSION: 0.0.1]  

More methods will come, for:  

*Creation, Update and Delete; URI resolution to Europeana Entities*
The Entity Collection

DBpedia resource for “Mozart” in our data

{  
  "@context": http://data.europaea.eu/entity/context.json,
  "id": "http://data.europaea.eu/agent/base/146911",
  "type": "Agent",
  "prefLabel": {  
    "ar": "فولتيمانج آجادم موتسارت",
    "az": "Volfang Amadey Mozart",
    "be": "Вольфганг Адэмус Mozart",
    "bg": "Волфганг Амануzas Мозарт",
    "bs": "Wolfgang Amadeus Mozart",
    "ca": "Wolfgang Amadeus Mozart",
    "cs": "Wolfgang Amadeus Mozart",
    "cy": "Wolfgang Amadeus Mozart",
    "da": "Wolfgang Amadeus Mozart",
    "de": "Wolfgang Amadeus Mozart",
    "el": "Βέλγικος Αναπτυκτικός Μόσχαρτ",
    "en": "Wolfgang Amadeus Mozart",
    "es": "Wolfgang Amadeus Mozart",
    "et": "Wolfgang Amadeus Mozart",
    "eu": "Wolfgang Amadeus Mozart",
    "fi": "Wolfgang Amadeus Mozart",
    "fr": "Wolfgang Amadeus Mozart",
    "ga": "Wolfgang Amadeus Mozart",
    "gd": "Wolfgang Amadeus Mozart",
    "gl": "Wolfgang Amadeus Mozart",
    "he": "יונהן אמסדוס מוסקארט",
    "hi": "जोस्फोक्स अमाडुस मॉस्कार्ट",
    "hr": "Wolfgang Amadeus Mozart",
    "hu": "Wolfgang Amadeus Mozart",
    "hy": "հույսի մարակերտ Մաձատ",
    "is": "Wolfgang Amadeus Mozart",
    "it": "Wolfgang Amadeus Mozart",
    "ja": "ヴォルフガング・アマデウス・モーツァルト",
    "ka": "ვოლფანგ ამადეუხს მოზარტ",
    "ko": "볼프강 아마데우스 모차르트",
    "lt": "Wolfgangas Amadėjus Mozartas",
    "lv": "Wolfganga Amadejs Mootarts",
    "mk": "Бонфран Мамуец Мураап",
    "nl": "Wolfgang Amadeus Mozart",
    "no": "Wolfgang Amadeus Mozart",
    "pl": "Wolfgang Amadeusz Mozart",
    "pt": "Wolfgang Amadeus Mozart",
    "ro": "Wolfgang Amadeus Mozart",
    "ru": "Моцарт, Вольфганг Адам",
    "sk": "Wolfgang Amadeus Mozart",
  
  }  
  "altLabel": {  
    "en": "Mozart, Johann Chrysostom Wolfgang Amadeus",
    "en.en": "Mozart, Wolfgang Amadeus"
  }  
  ,  
  "biographicalInformation": [  
    ,  
    "dateOfBirth": [  
      "1756-01-27"
    ],
    "dateOfBirth": [  
      "1791-12-05"
    ],
    "placeOfBirth": [  
      "Salzburg"
    ],
    "placeOfBirth": [  
      "Austria"
    ]
  }  
  ,  
  "sameAs": [  
    "http://sw.dbpedia.org/resource/Wolfgang_Amadeus_Mozart",
    "http://ko.dbpedia.org/resource/Wolfgang_Amadeus_Mozart",
    "http://bg.dbpedia.org/resource/Волфганг_Амануzas_Мозарт",
    "http://nda.dbpedia.org/resource/Wolfgang_Amadeus_Mozart",
    "http://ja.dbpedia.org/resource/ヴォルフガング・アマデウス・モーツァルト",
    "http://ca.dbpedia.org/resource/Wolfgang_Amadeus_Mozart"
  ]
}
The Entity Collection

Entity API - suggest method

/show/suggest.json?text=neo&lang=en&rows=6
Conclusion

• A Strategy for Entities is a “must” for Europeana

• There is no “one fits all” vocabulary

• We have a long way to go…

  ...but we are making progress
Thank you!

hugo.manguinhas@europeana.eu