The European Library

- Collects metadata for digital and non-digital materials from 48 National Libraries and more than 50 University/Research libraries
  - http://www.theeuropeanlibrary.org/
- A very large scholarly dataset
  - 200 million bibliographic records
  - 600 collections and their descriptions
  - 19 million digitised objects
  - Large corpus of full-text pages
- Data distributed as Linked Data
  - 61 datasets available as RDF dumps at http://www.theeuropeanlibrary.org/tel4/access/data/opendata

Europeana

- Aggregates metadata from the cultural heritage sector in Europe
  - Libraries, museums, archives and audio-visual archives
- Provides a portal for users to access data and objects
  - http://www.europeana.eu
- Metadata under Creative Commons Zero - public domain
- Previews and links to source
- Data distributed via
  - API: http://labs.europeana.eu/api/
  - Linked Data (currently being updated): http://data.europeana.eu/

Facing similar challenges

- Data aggregation issues
  - Too many interpretations of cataloguing practices and bibliographic data models
  - Legacy data affected by diverse data migration processes
  - Language diversity issues
    - Provide access to multilingual resources
    - Allow the searching in various languages
    - Make sure users can understand the metadata

Integrating Knowledge Organization Systems in Cultural Heritage Data

Data frameworks

- Europeana Data Model (EDM)
  - Re-uses several existing Semantic Web-based models: Dublin Core, OAI-ORE, SKOS, CIDOC-CRM...
  - Supports granular metadata
  - Relationships between cultural heritage objects
  - Multilingual & semantic links to contextual resources (e.g., concepts, person, places)

- MACS (Multilingual Access to Subjects)
  - Multilingual subject data enrichment through MACS (Multilingual Access to Subjects)
  - MACS is integrated in The European Library as links included in bibliographic data.
    - English - Library of Congress Subject Heading (LCSH)
    - French - RAMEAU
    - German - Schlagwortnormdatei (SWD)

Encouraging providers to contribute their own vocabularies

- Benefit from data links made at data providers’ level
- Ingestion of vocabularies is made possible if the vocabularies used the data structures EDM expects
  - For instance SKOS for concepts

Rely on knowledge organisation systems

- Create a “semantic layer” on top of cultural heritage objects
  - Include multilingual “value vocabularies”
- From Europeana’s providers or from third-party data sources

Automatic enrichment

- Objects are connected to third-party sources
  - Places from GeoNames, Agents from DBpedia, Concepts from GEMET and DDBpedia

Linking library data with the wider cultural heritage domain – WWI example

- Integration of library data with other cultural heritage data from specific communities
- Europeana 1914-1918 (http://www.europeana1914-1918.eu/) developed its own vocabulary based on a subset of LCSH
  - Terms translated in 10 languages and linked to id.loc.gov
  - Published in SKOS via the OpenSkos vocabulary service
- Future work
  - Re-use of the vocabulary by WWI communities
  - Alignment with other vocabs to make our collections more interoperable
  - Full integration of SKOS data in the portal

Challenges ahead

Continue populating the “semantic layer”….

- Creating synergies between domains to create LOD at low cost
- Focus on cross-domain multilingual vocabulary alignment and publish the results as Linked Data

...at The European Library

Subject browsing is based on normalization of bibliographic data regarding the subject of the resources

- Subject normalization centered in the Common European Research Classification Scheme
  - A research and development oriented knowledge organization system
  - It is part of the Common European Research project Information Format - CERIF

Ongoing and future work:

- Mining the aggregated dataset for subject vocabularies in use throughout Europe
- Improving the subject data delivered to Europeana, by adding support for vocabularies supported by Europeana:
  - Art & Architecture Thesaurus (AAT), Getty Thesaurus of Geographic Names (TGN), Iconclass

...at Europeana

- Two aggregation services
  - Data aggregation issues
    - Too many interpretations of cataloguing practices and bibliographic data models
    - Legacy data affected by diverse data migration processes
    - Language diversity issues
      - Provide access to multilingual resources
      - Allow the searching in various languages
      - Make sure users can understand the metadata

The subject enriched record in EDM for delivery to Europeana

Enrichments added through MACS

Multilingual subject data enrichment through MACS (Multilingual Access to Subjects)