



europaena

connect

WP2 – Multilingual Access to Content

Vivien Petras
Berlin, 10 May 2010



coordinated by the [Austrian National Library](#)  Österreichische
Nationalbibliothek

WP2

- **2.1** User preferences for multilingual access (D2.1.1, Dec. 2009)
- **2.2** Repository of language resources (M2.2.1, April 2010)
- **2.3** Multilingual mapping of controlled vocabularies (M21)
- **2.4** Translation services for querying (M24)
- **2.5** User behavior & retrieval evaluation of the services (M26)
- **2.6** Integration with Europeana retrieval framework (M30)

WP2 – Multilingual Access to Content

- Make Europeana talk European...
 - Multilingual search (query translation)
 - Multilingual browsing (mapping multilingual KOS)

- Core language set

French	English	Spanish
German	Italian	Polish

- Secondary language set

Dutch	Portuguese
Hungarian	Swedish

2.1: User Preferences for Multilingual Access

- What do users want?
 - **Workshop MLIA4DL** (September 2009): European projects on multilingual digital libraries / resources
 - **Survey** on multilingual information access
- **D2.1.1: Report on User Preferences and Information Retrieval Scenarios for Multilingual Access in Europeana** (December 2009)
 - User studies on MLIA
 - Survey results
 - Suggested usage scenarios

2.2 Repository of Language Resources

- Aggregate resources
- Which resources are out there?
 - Partner survey on language resources
 - Memo: Necessary Language Resources for a Multilingual Query Translation System in Europeana
- Licensed resources \leftarrow \rightarrow open source resources
- Europeana aim: open source
 - Memo: Workflow for the integration of open source language resources

2.2 Repository of Language Resources

- Infrastructure for Language Resource repository is set up (M2.2.1)
- <https://europeanalabs.eu/core/svn/contrib/lrr/trunk>
- Resources which will be stored in the repository:
 - Stop word lists
 - Language guessers
 - Linguistic analysis components
 - Translation dictionaries
- Data will be provided via standard API

2.2 Evaluation of language resources

- **Memo:** Language Resources - Evaluation Criteria (Draft)
- Languages
 - Quantitative evaluation criteria (Size, Speed, Coverage)
 - Qualitative evaluation criteria (Accuracy, Completeness)
 - Other (Compatibility, Availability, Maintainability)
- **Summer 2010:**
- Catalogue of evaluation / prioritization criteria
 - Suggestions for open source language resources from Europeana community

2.3 Multilingual Mapping of Controlled Vocabularies

- Alignments bridge collections
- Multilingual browsing (People, places, topics, object types...)
- How will this look like?
 - Europeana ThoughtLab
 - Use cases & requirements for semantic functions
(D1.4.1)

WP2.3 Languages and Vocabularies for Selection

- Which vocabularies in which languages?
 - **Memo:** Languages and Vocabularies for Selection
 - Surveys (WP1, EuropeanaLocal)
 - Prioritizing vocabularies (relevance, language)
 - Pivot vocabularies (multilingual, wide-coverage, widely adopted: UDC, DDC, wordnets, TGN, VIAF)
- Mapped vocabularies (prototype):
 - AAT (English, Dutch)
 - Cornetto (Dutch)
 - Iconclass (English, Dutch)
 - OSZK tezaurusz (Hungarian)
 - Rameau (French)
 - SCRAN (English)
 - Systematik Österreichische Mediathek (German)

WP2.4 Translation modules for Europeana

Challenges:

- query language unknown
 - desired target languages unknown
 - search sequence
 - ambiguous translation candidates
-
- Dynamic at query-time
 - Disambiguation of translation candidates (CACAO Approach)

WP2.4 Translation modules for Europeana

→ **Memo:** Necessary resources and steps for a query translation module

→ **2.2** Aggregation of languages resources

→ **Demo:** Language identifier for queries

Next Steps

2.2 Language resources repository:

- Populating the repository
- Evaluation of open-sources resources (suggestions from community)

2.3 Mapping multilingual controlled vocabularies:

- Prioritization, Vocabulary aggregation
- Alignments (with WP1.2)

Next Steps

2.4 Query translation:

- Language identification
- Specification for translation module

2.5 Sandbox integration, texting & evaluation:

- Functionalities in accordance with user requirements