



europaena

connect

## WP2 – Multilingual Access to Content

Vivien Petras

Den Haag, 14 September 2009



# 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	

- Secondary language set

Dutch	Polish	Swedish
Hungarian	Portugese	

# WP2: Multilingual infrastructure and tools

---

- **2.1 user study** for multilingual access
- **2.2 repository of language resources**
- **2.3 multilingual mapping** of controlled vocabularies for precise translations
- **2.4 translation services** for querying and browsing
- **2.5 User behavior & retrieval evaluation** of the services
  
- Make sure that multilingual issues are not forgotten in data modelling and user interface design!

# Clarification issues

---

- Data model
  - Language identification in every “textual field”
  - Parallel translations of concepts → SKOS labels
  - Parallel translations of objects → how are they represented?
- Language resources
  - 2 languages = 1 dictionary
  - 5 languages = 10 dictionaries...
  - Some open source, some licensed
  - Who has resources?

# Clarification issues

---

- Multilingual interface requirements
  - Language selection / detection for:
    - interface / querying / browsing / result representation
  - Automatic / manual
  - Multiple languages for results? Only as facet?
- Language selection for results
  - Metadata vs. Full-text
  - what about non-textual documents: images, video, audio?

# Clarification issues

---

- Browsing interface
  - Browse by person, place, time → different language representations?
  - Browse by topic → how? Which language representations?
  - Visualize KOS information? & Mapping information? → How to visualize in different languages?

# Clarification issues

---

- User interaction for querying
  - Automatic / recommended translations?
  - User-suggested translations → archive? / Data model?
- Named entities
  - Proper names need special treatment:
  - Ex. Jill Cousins = Jill Cousinen?
  - Berlin – Berlin / Wien – Vienna / New York – Neues York?
  - Spelling variants → more than 1 translation

# Clarification issues

---

- Query expansion through semantic information
  - Automatic / recommended
  - Which KOS are being mapped?
  - Semantic mapping  $\leftarrow \rightarrow$  translation
  - How to add multilinguality? Just another mapping? Additional switch / interaction?
- Performance issues
  - Adding processing in the query process



# Use case

---

- User: Polish → Europeana interface = Polish ??
- Query: xxxxxxx
  - (assume Polish? / ask for language? / detect language? / search string)
- Language selection for results: Polish / French / English
  - provide selection? Provide facets after result set has been shown
  - No language selection: Polish / now in list of importance / frequency?
- Query expansion: → let's assume query was in English ←
  - Query phrase = “cats in England”
  - Which comes first mapping or translation? Automatic / user-directed
- Browsing...