



europeana

connect

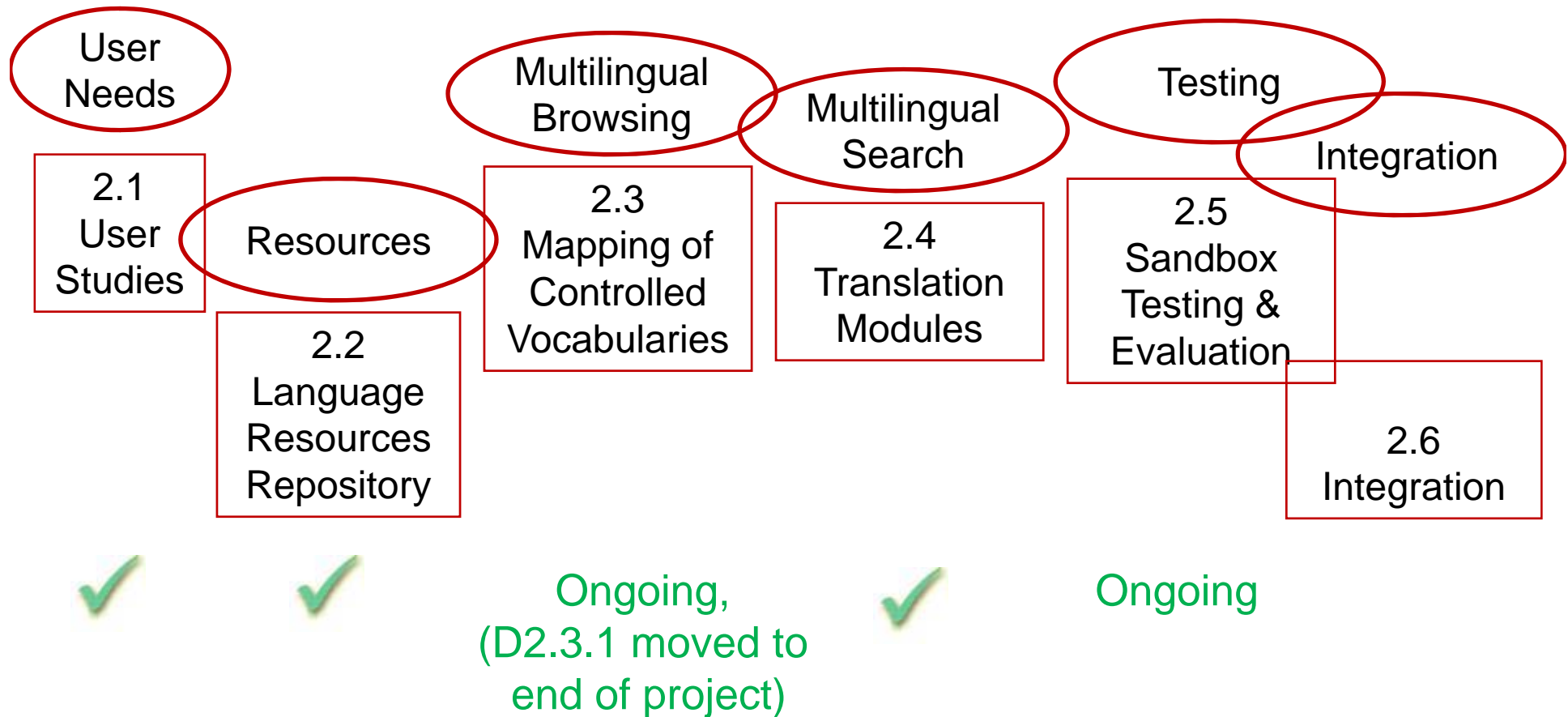
WP2: Multilingual Access to Content

EuropeanaConnect

Vienna, 23 May 2011

Vivien Petras (HUB)

State of the Art WP2



Make Europeana talk European...

- 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	Portugese
Hungarian	Swedish

Task 2.1: User Preferences for Multilingual Access

- Report on User Preferences and Information Retrieval Scenarios for Multilingual Access in Europeana (**D2.1.1, Oct 2009**)
 - Interface
 - Users are more likely to visit a collection if the site was translated in their preferred language.
 - Search (Query Translation)
 - Users routinely search and browse in languages other than their native language, especially if the native language is not English.
 - Users often search for place names and subject keywords → special treatment of named entities necessary
 - Result Representation
 - Subject translation seems to be useful, translation of whole documents seems not be desired or sensible.
 - Users routinely refine results by language.
 - Browsing

Task 2.2: Repository of Language Resources

- M2.1.1 Europeana Language Repository online (April 2010)
- D2.1.1 Europeana Language Resources Repository (July 2010)
- Language Resources Repository: software + linguistic data
 - **Repository Code is now free available**
 - Online in EuropeanaLabs
 - Open-source (anon. download) & password-protected sections (proprietary resources)
- Publicly-visible **register** of available resources
(<http://europeanalabs.eu/wiki/LinguisticResourceRegister>)
- Next step:
 - Extract open source language resources and move to public accessible server.



Task 2.3: Multilingual Mapping of Controlled Vocs

- augmenting the Europeana Semantic layer with semantic resources and alignment between these, which give the layer a multilingual dimension
- M3.2.1 Mapped vocabularies and metadata schemes (Feb. 2011)
- **D3.2.1 Multilingual mapping of schemes and vocabularies was moved to October 2011**
- Memo: Multilingual Mapping of Controlled Vocabularies – Languages and Vocabularies for Selection (April 2010)
 - **Listing:** Languages and Vocabularies for Selection
 - Surveys (Europeana data, WP1, EuropeanaLocal)
 - Identified several vocabularies in all 10 languages
 - **Prioritizing** vocabularies
 - By access, relevance, language

Task 2.3: Multilingual Mapping of Controlled Vocabs

- Methodology:
 - Priority on mapping pivot vocabularies (AAT, Wordnet), then more specific vocabularies (domain, institution, language) → flexible pattern
- Development of **tools for alignment**:
 - AMALGAME tool. see WP1, Task 1.2 Aligning the Semantic Data
- Vocabularies / alignments:
 - Skosified: en, fr, de, nl, hu, es, it, se
 - Mappings: en, fr, nl, de, se, it
 - Polish and Portuguese not enough vocabularies available
- <http://europeanalabs.eu/wiki/WP12Vocabularies>
- Demo presentation of ML mapping tomorrow

2.4: Translation Modules for Europeana

- D2.4.1 Modules for partial query result translation / enrichment (Mai 2011)
 - software module: MultiLingual Information Access (MLIA) based on bilingual dictionaries
- Challenges for query translation:
 - Short & Query language unknown
 - Named entities
 - Ambiguous translation candidates
- Language identifier: **software**
 - Online in EuropeanaLabs (also as web service: <http://www.cross-library.com/FAUSS/>)
 - For all 10 languages + Greek
 - Externally evaluated

2.4 Translation Modules for Europeana

<http://demo.celi.it:8096/portal>



Refine your search:

- By provider ▼
- By language ▼
- By country ▼
- By date ▼
- By type ▼

Actions:

[Save this search](#)

[Login](#) | [Register](#)

Choose a language ▼

did you mean: [null](#)

vienna

Search

[Advanced search](#)

Multilingual search

[Translation Details](#)

Query Translation Details

vienna [query language: EN]

- nl = Wenen
- pl = wiedeń Wiedeń Wieden
- it = Vienna
- fr = Vienne
- hu = Bécs
- de = Wien
- es = Viena
- sv = Wien
- pt = Viena

Matches for: vienna

Results 1 - 12 of 13 All [Texts](#) [Images](#) [Videos](#) [Sounds](#)



EuropeanaConnect WP2
Vivien Petras, Humboldt-Universität zu Berlin
EuropeanaConnect All Staff Meeting, 23-05-2011, Vienna

Coordinated by the [Austrian National Library](#)

Österreichische
Nationalbibliothek

2.5 Sandbox Integration, Testing & Evaluation

- Report on Europeana Language Resources - Evaluation Criteria (March 2010)
 - Definition of criteria to evaluate the usability of linguistic resources
- Report on Evaluation of Multilingual Information Access to Europeana (Dec. 2010)
 - Roadmap for query translation evaluation
 - Describes the experimental collection, the experimental set-up and methodology.



Information Retrieval Evaluation

- IR evaluation: compare retrieval results for:
 - Open Source LR
 - Licensed LR
 - Google Translate
 - Bing Translator
- For Europeana query translation we evaluate two tasks:
 - **monolingual tasks** where the language of the source query is the same as that of the target collection, for example an English query against an BL collection;
 - **bilingual tasks** where the language of the source query is different from that of the target collection, for example a German query against the BL collection.



Experimental Set-Up

3 source languages

- British Library (BL): 1,000,100 records,
- Bibliothèque Nationale de France (BNF): 1,000,100 records,
- Austrian National Library (ONB): 869,353 records

10 target languages

- 100 topics for each language:
- English, French, German, Italian, Polish, Spanish, Portuguese, Swedish, Dutch and Hungarian



Topic Translation

- Translation of 100 CLEF-topics into 10 Europeana languages (English, French, German, Italian, Polish, Spanish, Portuguese, Swedish, Dutch and Hungarian)
- Topics are structured statements representing information needs, e.g.:

```
<topic>  
<topic lang="en">  
<identifier>451-AH</identifier>  
<title>Roman Military in Britain</title>  
<description>Find books or publications on the Roman invasion or military occupation of Britain.</description>  
<narrative />  
</topic>
```

IR Runs for Query Translation

Source	Target English (BL)						Target French (BnF)						Target German (ONB)					
	en	de	fr	es	po	...	en	de	fr	es	po	...	en	de	fr	es	po	...
Monolingual																		
Google Translate																		
Bing Translator																		
Celi Translator with:																		
OS Language Resources																		
CELI Resources																		
Xerox Resources																		

Ongoing Evaluation Activities

- Preparation of DIRECT (Distributed Information Retrieval Evaluation Campaign Tool)
- The DIRECT @ Europeana instance is available at:
<http://svrims.dei.unipd.it/europeana-mlia-evaluation/>



DIRECT



for Evaluation of Multilingual Information Access to Europeana

Logged as cell @ EC2011 [Logout] - Thursday 19 May 2011

Main

[DIRECT Portal](#) | [Main Page](#) | [User Management](#) | [Translation Manager](#) | [About](#) **DIRECT**

Track Identifier	Track Description	Topic Status	Submission Status	Metrics Status
AH-EC2011	Ad-Hoc Track	Available	Ongoing <i>Deadline: Sep 30, 2011</i>	Available
Task Identifier	Task Description			Download Assessments
AH - BILI - X2DE - EC2011	Ad-Hoc Bilingual German Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +
AH - BILI - X2EN - EC2011	Ad-Hoc Bilingual English Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +
AH - BILI - X2FR - EC2011	Ad-Hoc Bilingual French Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +
AH - MONO - DE - EC2011	Ad-Hoc Monolingual German Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +
AH - MONO - EN - EC2011	Ad-Hoc Monolingual English Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +
AH - MONO - FR - EC2011	Ad-Hoc Monolingual French Task	Download Topics	Submit Experiment View Task Descriptive Statistics	Download Assessments +

EuropeanaConnect is a Best Practice Network funded by the European Commission within the area of Digital Libraries of the eContentplus Programme.
 DIRECT - Copyright © 2005-2011 Information Management Systems (IMS) Research Group, Department of Information Engineering (DEI), University of Padua.
[Credits](#)





EuropeanaConnect WP2
 Vivien Petras, Humboldt-Universität zu Berlin
 EuropeanaConnect All Staff Meeting, 23-05-2011, Vienna

Coordinated by the **Austrian National Library** **Österreichische Nationalbibliothek**

Topic Identifier: [451-AH](#)

Number of Retrieved Documents	1000
Number of Relevant Documents	33
Number of Relevant Retrieved Documents	6
Average Precision	4,51%
R Precision	9,09%
Binary Preference	18,18%
Interpolated Recall vs Average Precision	
0%	66,67%
10%	1,00%
20%	0,00%
30%	0,00%
40%	0,00%
50%	0,00%
60%	0,00%
70%	0,00%
80%	0,00%
90%	0,00%
100%	0,00%
Precision at R Retrieved Documents	
Precision at 5 Retrieved Documents	40,00%
Precision at 10 Retrieved Documents	30,00%
Precision at 15 Retrieved Documents	20,00%
Precision at 20 Retrieved Documents	15,00%
Precision at 30 Retrieved Documents	10,00%
Precision at 100 Retrieved Documents	3,00%
Precision at 200 Retrieved Documents	1,50%
Precision at 500 Retrieved Documents	0,80%
Precision at 1,000 Retrieved Documents	0,60%

 [Top](#)  [Download XML](#)

Topic Identifier: [452-AH](#)

Number of Retrieved Documents	1000
-------------------------------	------



EuropeanaConnect WP2
 Vivien Petras, Humboldt-Universität zu Berlin
 EuropeanaConnect All Staff Meeting, 23-05-2011, Vienna

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

Evaluation Activities

- A total of 228 experiments have been submitted to the evaluation system (76 different runs for each of the 3 collection)
 - Not all resources are available for each language
- Out of the 76 runs for each task:
 - 58 are from different configurations of the MLIA module
 - 18 from Machine Translations systems publicly available on the web (Google Translate and Bing)

2.6 Integration with Europeana Retrieval Framework

- Europeana Office is task leader
- Query translation integration not yet planned for Danube Release
- Currently document translation with Google Translate and Bing Translate available (initiated by the office)



Document Translation with Google/Bing

The screenshot shows the Europeana website interface. At the top right, there are navigation links: "My Europeana", "About us", "Communities", "Partners", "ThoughtLab", and a "Choose a language" dropdown menu. A search bar contains the text "vienna" and a "Search" button. Below the search bar is a link for "Advanced search".

On the left side, there is a "Translate with:" widget. It has two tabs: "Google" and "Bing". A dropdown menu is open, showing a list of languages: Catalan, Chinese, Chinese_simplified, Chinese_traditional, Croatian, Czech, Danish, Dutch, English, Estonian, Filipino, Finnish, French, Galician, German (highlighted), Greek, Haitian_creole, Hebrew, Hindi, and Hungarian.

The main content area shows search results for "vienna". The first result is titled "poltrona" and includes a photograph of a chair. The details for this item are as follows:

- Title:** poltrona
- Date:** DTZ.DTZG=sec. XIX;DTS.DTSI=1875;DTS.DTSF=1899;DTM=analisi stilistica; [Production]
- Description:** schienale curvato in paglia di Vienna con cornice decorata a motivi circolari, braccioli con testata decorata a volute e girali vegetali e parte interna in paglia di Vienna, piedi anteriori a volute e posteriori lisci, fasce di raccordo tra i piedi decorate a girali d'acanto e con un cestino colmo di fiori al centro di quello anteriore
- Format:** MISU=cm;MISA=82;MISL=74;MISP=78; ; legno ; paglia di Vienna
- Source:** Progetto ArtPast- CulturalItalia
- Rights:** proprietà Stato
- Data provider:** Progetto ArtPast- CulturalItalia
- Provider:** Athena ; Italy

At the bottom of the page, there is a footer with navigation links: "Back to top", "Using Europeana", "Accessibility", "Sitemap", "Terms of use", "Privacy", "Language policy", "Contacts", and "Send us feedback". There are also social media icons for Facebook and Twitter, and a note that the site is "co-funded by the European Union".

WP2 Outcomes & Next Steps

- User preferences and retrieval scenarios for multilingual search
- Language resources + API for translation infrastructure for 10 languages
- Process + flexible infrastructure for integrating more language resources
- > 500,000 mappings of controlled vocabulary terms in several languages in the semantic network layer
- Language identification prototype for 11 languages
- Query translation prototype for all 10 languages (**evaluated**)

Next:

- More mapping (more vocabularies + languages)
- Integration with Europeana infrastructure

