



EDLproject: challenges of multilingual access to multilingual European content

Maja Žumer, University of Ljubljana, Slovenia

DELOS-MultiMatch workshop "Ontology-driven Interoperability"

Pisa, 15 February 2007





EDLproject

One more step towards the European digital library





Background

- ❑ A **Targeted Project** within the area of cultural and scientific/scholarly content of the *econtentplus* Programme
- ❑ A building block of The European Library and a continuation of TEL-ME-MOR
- ❑ Started in Sept 2006. Will last 18 months.
- ❑ Budget € 2.114 m, 1m contribution by the EC



Partners

- German National Library - Co-ordinator
- National Library of Austria
- Royal Library of Belgium
- Eremo Srl, Italy
- National Library of Greece
- National and University Library of Iceland
- National Library of Ireland
- Liechtenstein National Library
- National Library of Luxembourg
- National Library of the Netherlands
- National Library of Norway
- National and University Library of Slovenia
- National Library of Spain
- Royal Library, National Library of Sweden
- National Library of Switzerland



Objectives

- ❑ To **include remaining 9 EU and EFTA countries:** Belgium, Greece, Iceland, Ireland, Liechtenstein, Luxembourg, Norway, Spain, Sweden.
- ❑ To **further enhance access** by continuing to develop the multi-lingual capacity of the portal
- ❑ To **promote and increase the use and awareness of the extended The European Library service** in the target countries and worldwide
- ❑ To **leverage the influence of CENL** as a key player in the content field to work towards a consensual resolution of issues raised by the “i2010: Digital Libraries” Communication



WP2 Extending the multilingual capacity of the network

- ❑ **Task 1 Unicode compliance**
 - Extending the Unicode compliance analysis carried out in TEL-ME-MOR to EDL partners.
- ❑ **Task 2 Subject access tools**
 - Extending the initial analysis in TEL-ME-MOR, resulted in an overview of classification systems, subject headings and thesauri in NMS national libraries, to EDL partners
- ❑ **Task 3 Cross-language access to bibliographic subject data**
 - Testing of a selection of cross-language approaches to subject data, including MACS, MSAC and CrissCross, building on results of task 2 and TEL-ME-MOR deliverables on cross-language options and Implementation proposal for multilingual access
- ❑ **Task 4 Cross-language access to full text**
 - Identifying one or more general (not subject specific) data dictionaries and test their applicability for cross-language access within the European Library (in cooperation with DELOS); identifying available automatic translation tools, stemmers etc.
- ❑ **Task 5 Name authority control**
 - Extending the work of TEL-ME-MOR and building on projects such as LEAF, VIAF and ONESAC, design and test the feasibility of a name authority control tool
- ❑ **Task 6 Interoperability with other communities**
 - Exploring options of interoperability with museums; testing the use of new conceptual models of bibliographic information.



Outcomes of TEL-ME-MOR

- Reports on UNICODE requirements, cross language & authorities options
- Analysis of subject access tools
- Reports on Implementation proposal & scenario for multilingual access



Layers of multilinguality

- **Language of the interface**
- **Language of bibliographic records/cataloguing language/language of metadata**
- **Language of the resources**



The user should be able to:

- **Use his/her own language when communicating with the system**
 - solved
- **Use his/her own language to formulate a query and, as the result, retrieve all relevant (digital) objects in any language**
 - remains a problem: too many language combinations; all-to-all mapping may not be feasible



Investigation of smaller steps:

- Test a selection of cross-language approaches to subject data, including MACS, MSAC and CrissCross
- Test MACS system using existing links
- Test MACS with the Luxembourg subject headings (Laval English / French) and the Spanish subject headings
- Provide an updated overview of recent European projects and initiatives that may have relevance to cross-language access to various types of collections.
- Use the MACS Link Management system (replicated as required) central source (clearinghouse) of mapping results between subject headings used in European national libraries (classifications and subject headings)



Testing automated mapping

- Investigate HILT, TermSciences and WebDewey as tools to speed up link creation
- [In TELplus, plans to test Stitch]
- Contact OCLC to investigate their dynamic mapping between subject authority lists
- Multi-Lingual Information Retrieval (MULIR) project: maps from English Library of Congress Subject Headings (LCSH) to words and phrases in over 100 languages and vice versa



Improving interoperability across (name) authorities

- **Building on projects such as LEAF, VIAF and ONESAC:**
 - automatic clustering of existing variant headings



Full-text searching

- **An overview of existing tools is necessary:**
 - **language recognition tools**
 - **stemmers**
 - **automated translation etc.**



At this point we are not aiming at a fully functional cross-language IR for all languages, but at least a prototype of a limited application.

Therefore cooperation with any relevant research is needed.