Multilingual Information Access in Digital Libraries

Report on Multilingual Access Strategies to Digital Libraries

co-funded by the European Union

The project is co-funded by the European Union, through the **eContent*plus*** programme
http://ec.europa.eu/econtentplus

Österreichische Nationalbibliothek

EuropeanaConnect is coordinated by the Austrian National Library

ECP-2008-DILI-528001

EuropeanaConnect

Report on Multilingual Access Strategies to Digital Libraries

Deliverable number/name	*D2.7.1*
Dissemination level	*Public*
Delivery date	*2011-10-24*
Status	*Final*
Author(s)	*Vivien Petras* *Humboldt-Universität zu Berlin (HUB)*

e*Content*plus

This project is funded under the *e*Content*plus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

Österreichische Nationalbibliothek

EuropeanaConnect is coordinated by the Austrian National Library

Distribution

Version	Date of sending	Name	Role in project
0.1	2011-09-28	Humboldt-Universität zu Berlin	Task Leader
0.2	2011-10-23	ONB	Return of internal reviews
1.0			

Approval

Version	Date of approval	Name	Role in project
1.0			

Revisions

Version	Status	Author	Date	Changes
0.1	D	Humboldt-Universität zu Berlin		First draft submitted to internal revision
1.0	F	Humboldt-Universität zu Berlin		Final version updated according to internal reviewers suggestions

Abstract
This deliverable reports on the efforts of work package WP2 (Multilingual Access to Content) of EuropeanaConnect to increase and enhance the users' possibilities to access the Europeana portal and Europeana content multilingually, that is with their native or preferred language. It highlights the many different multilingual access strategies that are already available to Europeana users today or have been developed as a prototype during the course of the project. The efforts of WP2 partners to facilitate exchange of ideas, develop joint applications and maintain communication with different research and applied communities about multilingual issues (with a particular focus on Europeana) in the areas of language resources, semantic mapping and multilingual information retrieval are highlighted. The document finishes with some open-ended challenges.

Table of Contents

1 Introduction 5

1.1 EuropeanaConnect Work Package 2: Multilingual Access to Content 6

1.2 Aspects of Multilingual Access in Digital Libraries 7

2 Multilingual Access Strategies in Europeana 8

2.1 Interface Language Change 8

2.2 Multilingual Browsing 8

2.3 Query Result Filtering by Language 9

2.4 Multilingual Mapping of Vocabularies – Semantic Search & Enrichment 10

2.5 Query Translation 11

2.6 Document Translation 12

3 Facilitation & Exchange of Multilingual Access Strategies 14

3.1 Principles for Multilingual Access in Cultural Heritage 14

3.2 Multilingual Digital Library Projects & Applications 14

3.3 Language Resources & Language Processing 14

3.4 Multilingual Semantic Web 15

3.5 Multilingual Information Retrieval 15

4 Challenges and Future Work 16

References 18

1 Introduction

"Language is the most direct expression of culture; it is what makes us human and what gives each of us a sense of identity." (Communication from the Commission, 2005)

"EuropeanaConnect will support the creation of a diverse and inclusive Europeana facilitating access to culture by all communities and individuals and representative of various cultures and language-groups." (EuropeanaConnect, 2009)

Already in 2005, the European Commission published a new framework strategy for multilingualism, which sets out three goals:

- "to encourage language learning and promoting linguistic diversity in society,
- to promote a healthy multilingual economy, and
- to give citizens access" (Communication from the Commission, 2005).

Surveys of EU citizens and their language use in 2001 and 2006 showed that at least half of the EU population speaks at least one more language in addition to their native language, with an increasing tendency (European Commission, 2006). Smaller member states show higher numbers of multilingualism among their citizens. The most frequent second languages spoken are English, French, German, Spanish and Russian and the overwhelming majority of Europeans considers multilingualism and language learning important.

Language use in World Wide Web applications is increasingly becoming more varied. Whereas the web was dominated by English-language contents in its beginnings, non-English internet use and content is dramatically increasing (Chung, 2008). With other languages being integrated, multilingual challenges arise, for example encoding formats for different versions of Chinese, other writing styles (e.g. right-to-left script in Arabic) or simple orthography variations in name spelling.

A recent EU survey of 13,752 phone interviews in all 27 member states revealed that at least 80% of internet users used in on a daily basis and more than half of the respondents used at least one additional language (additional to their mother tongue) when consuming content on the web. Even for actively producing content (e.g. writing emails or posting comments), still more than a third of the respondents frequently used a language other than their native language (European Commission, 2011). Even though English was the most frequently mentioned second language, over 80% stated that websites produced in their country should be available in their country's language and other language versions as well. When given a choice, the large majority of internet users said they preferred web sites in their own language and almost half of the respondents claimed they missed interesting information because the content was in a language they didn't understand.

The Europeana portal now provides access to over 20 million digitized cultural heritage objects. An endeavour as ambitious as the Europeana project has to meet several challenges to fulfil its goal of aggregating cultural heritage objects for European citizens to access and study without restrictions. Different heterogeneous media types (images, texts, sound and videos) have to be organized and presented simultaneously and similarly. Both metadata descriptions and objects come in different standards, formats and – particularly – description languages (both in terms of natural language titles and other descriptions and controlled vocabularies. Figure 1 displays the distribution of Europeana content by content provider country showing some of the multilingual dimensions of the content provided. Content provider country, however, is only a first approximation for the language of the content as institutions collect objects that are in different languages or even provide their metadata in different or parallel languages. Europeana users also access the system with different cultural, societal and language backgrounds therefore having different interaction and search requirements.

Europeana's own user studies also showed that their users know on average 1.5 languages additional to their mother tongue (IRN Research, 2011), but language is still perceived as a significant barrier. Users asked for more assistance in translation for search and understanding the results (Dobreva et

al., 2010). In order to meet the goals of the EC's framework strategy for multilingualism, these language issues need to be specifically addressed.

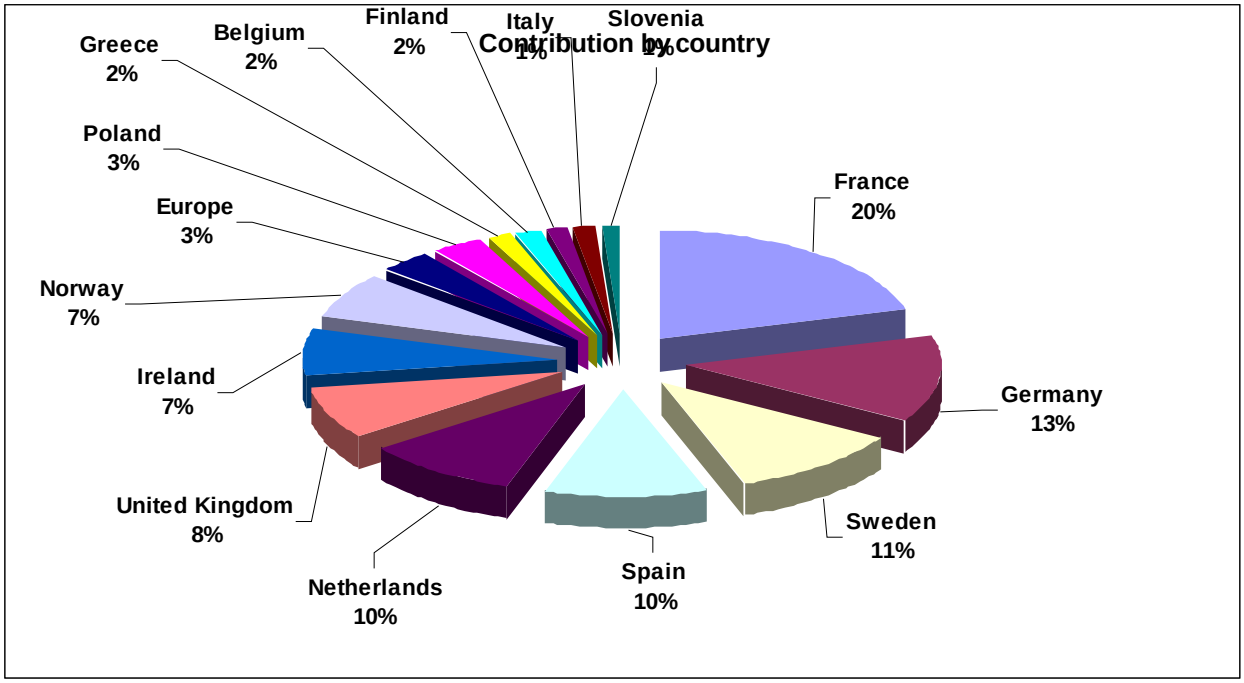

Figure 1. Europeana content contribution by country (Cousins, 2010).

1.1 EuropeanaConnect Work Package 2: Multilingual Access to Content

Within the EuropeanaConnect project (www.europeanaconnect.eu), a separate work package (WP2) was devoted to develop solutions to cope with multilingual access issues for users and objects alike within Europeana. Through the provision of multilingual access capabilities content should be able to be used by all Europeana users equally, regardless of the users' native language or the available native language of resources. The main objective of this work package was to develop a multilingual infrastructure and tools for access within Europeana, including:

- a multilingual user needs assessments,
- the Europeana Language Resources Repository (for translation and mapping),
- multilingual mappings of subject metadata schemas and other controlled vocabularies (for search and metadata enrichment,
- query translation modules or services (for integration with the general search and retrieval infrastructure),
- the evaluation of the query translation prototype, and
- a strategy for the integration of these tools with the Europeana production system.

Since not every Europeana language is equally well developed in terms of multilingual resources for translation, work package 2 proposed to start with a core set of languages for which expanded multilingual capabilities will be implemented. The core set of languages contains: English, German, Spanish, French, Italian, and Polish. A secondary language set, for which fewer resources exist and therefore fewer language capabilities could be developed during the project period, consists of: Dutch, Hungarian, Portuguese and Swedish.

In particular, WP2 built on techniques and experiences from the CACAO (Cross-Language Access to Catalogues and On-line Libraries) and Multimatch (Multilingual/Multimedia Access to Cultural

Heritage) projects, both dealing with issues in multilingual access to digital libraries and cultural heritage objects respectively. The technical and conceptual framework employed for the creation of multilingual mappings is work resulting from the TELplus project and the AnnoCultor tool set also utilized for the Europeana Semantic Layer (WP1).

1.2 Aspects of Multilingual Access in Digital Libraries

In order to facilitate the creation of viable solutions for multilingual access, a better understanding of user needs and existing multilingual frameworks was imperative. During the course of the project, a number of questions were analyzed:

- What do we know about multilingual access to digital libraries?
- Which lessons and best practices have been learned from existing information systems dealing with multilingual content and users?
- What do users really want with respect to multilingual access within Europeana?
- Which steps should be taken on the way to a truly multilingual system?
- Which scenarios for multilingual access can we implement for a scalable, operational system?

Both commercial systems and other projects (most EU-funded) were studied. Even in the commercial search engine market, multilingual support is a relatively new phenomenon. Search engines started adding language support in 2004. A study by Zhang and Lin (2007) compared multiple language support features in 21 search engines. The following five aspects were compared: the number of supported languages, visibility of language support, translation ability, result presentation, and interface design.

Supported by evidence from the research literature and other multilingual information systems, we adopted multilingual access strategies more closely aligned with portal capabilities or interaction functionalities. The following five aspects of multilingual access were focused on:

1. Multilingual user interface: This includes the translation of all static content elements on the information system's web sites and a systematic administration of language information for all content elements (called "language-skinning").
2. Multilingual enrichment: This includes both the multilingual enrichment of object metadata (with text in other languages) and the mapping of monolingual knowledge organization systems to a multilingual semantic network, therefore enabling access across different languages.
3. Multilingual search: Multilingual search capabilities can be developed by query translation (the original query is translated into additional languages that the document collection contains), document translation (the documents in the collection are translated into the query language), or an interlingua or pivot language approach (both queries and documents are translated into a single language). Query translation is the most commonly adopted method for multilingual information systems today and was also used in EuropeanaConnect.
4. Multilingual result representation: This includes the representation of results according to language-specific user requirements and means both result filtering features (filter content by language) and result translation features (translate the object itself or at least the object's metadata for user interrogation).
5. Multilingual browsing: Next to search, almost all information systems also offer browsing capabilities to their users. Browsing within an information system is usually provided through a hierarchical classification or subject ontology for content descriptions. Europeana offers browsing capabilities through image suggestions (and their text annotations adapted to the interface language) on the front page.

Other outcomes of multilingual user preferences studies can be found in the report (Agosti et al., 2009).

2 Multilingual Access Strategies in Europeana

Many multilingual access capabilities were already described in the outline functional specification for Europeana, which was one of the outcomes of the predecessor EDLnet project. The specification (Dekkers et al., 2009) outlined potential requirements for a multilingual interface, browsing, multilingual search, and result translation. Almost all requested features are either implemented in the production system already or were demonstrated as a working prototype during the EuropeanaConnect project period. In the following, the individual multilingual features in Europeana will be briefly described.

2.1 Interface Language Change

From the first release of the Europeana portal, a multilingual user interface with language skinning (now up to all EU languages) was provided. Users can choose a language skin (all static content translated) via a pull-down menu from the homepage (see Figure 2). Cookies with this user requirement will be set in order to provide the desired language on a return visit.

Figure 2. Europeana portal interface with Latvian interface language selected

2.2 Multilingual Browsing

Multilingual browsing capabilities are controlled by the selected interface language. The component “People are currently thinking about”, which was placed in the lower part of the homepage (in an earlier version of the interface) offered language-specific query (i.e. browsing) suggestions to users viewing the portal in that particular language (Figure 3). In the new web design, images and their associated texts, which are adapted to the interface language, provide browsing suggestions to the users (see Figure 2).

Figure 3. Query suggestions from Spanish and French interfaces (feature discontinued in new web design).

2.3 Query Result Filtering by Language

When search results are displayed to the user, several filtering ("drill-down") options are presented to the user. Filtering by language allows the user to narrow down search results to a restricted result set that includes objects described in their selected language (Figure 4). The language filter, as it is implemented in the Europeana portal, selects objects according to the language of the content provider the object is provided from. Naturally, a more appropriate filtering implementation would sort according to the language of the object or at least the object's metadata. The language identification module developed in EuropeanaConnect WP2 supports the language identification of metadata fields so that this feature can be provided.

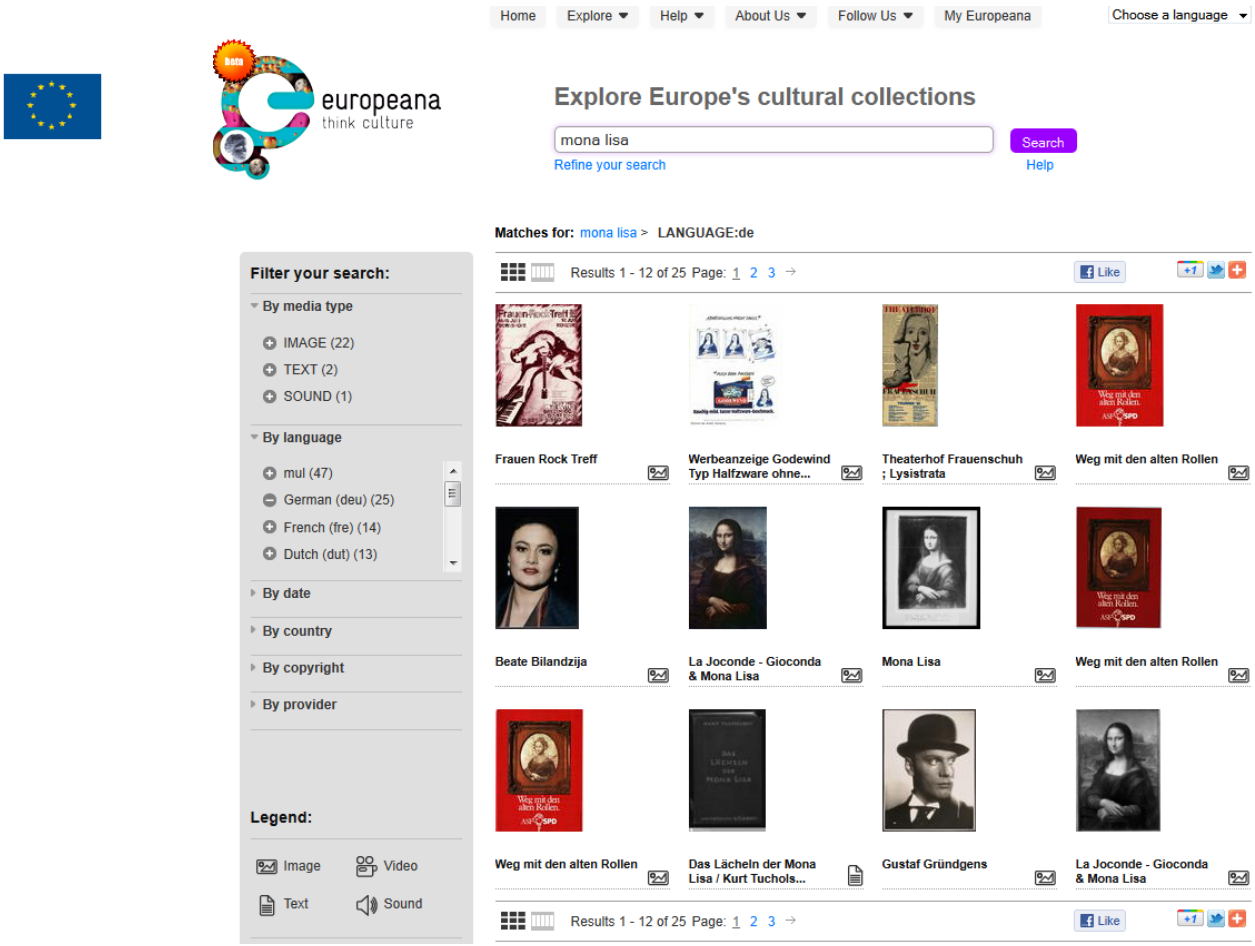

Figure 4. Europeana query results filtered by the German language

2.4 Multilingual Mapping of Vocabularies – Semantic Search & Enrichment

One of the core tasks in EuropeanaConnect work package 2 was the multilingual mapping of knowledge organization systems or other controlled vocabularies like name authority files. This allows multilingual searching or browsing via the linked semantic web of vocabularies that was created during the mapping process. This browsing feature has not been integrated into the Europeana production system yet, but can be tested in the Europeana ThoughtLab and contains data from the Rijksmuseum Amsterdam, the Louvre in Paris and the Netherlands Institute for Art History. The prototype's vocabulary mapping and search is available in English, French and Dutch (Figure 5).

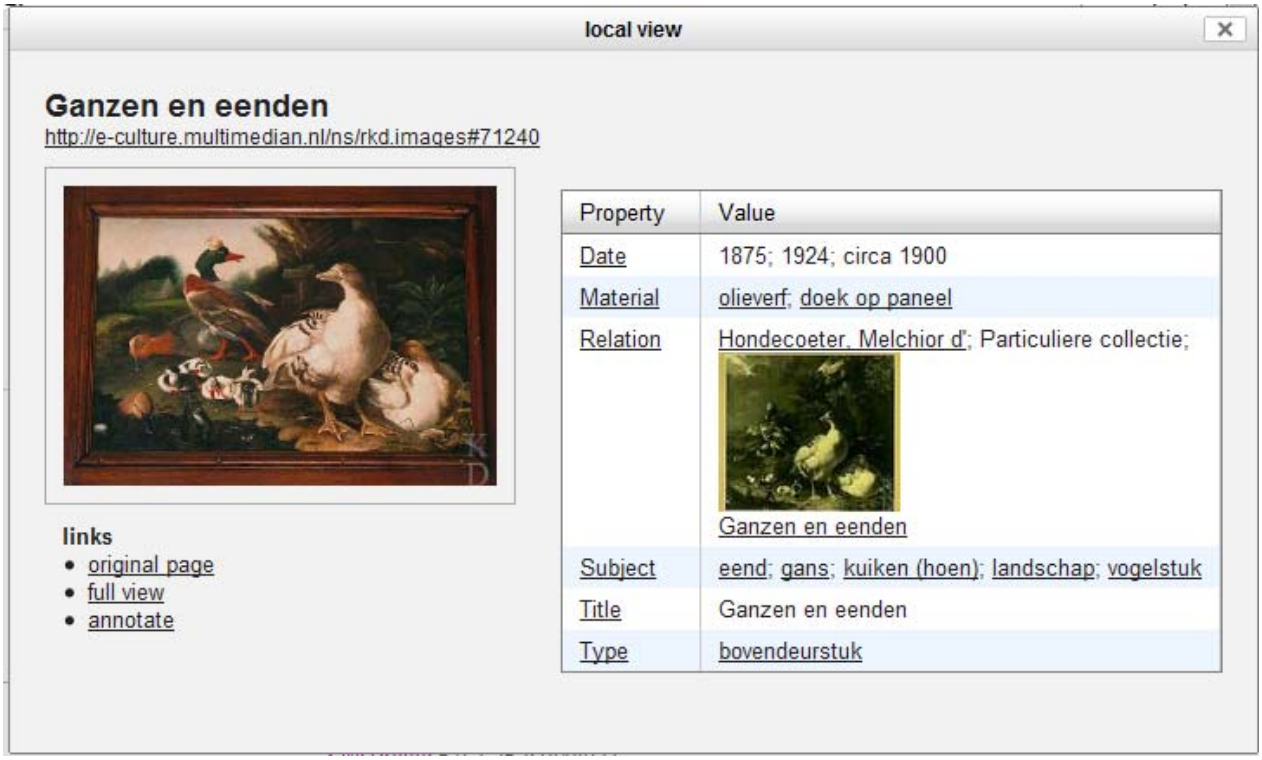

Figure 5. Multilingual query expansion (query = duck) in Europeana Semantic Search Prototype

Metadata object descriptions are now enriched with parallel language versions if the subject keywords, dates, person or place names occur in the linked data web developed. Together with the original metadata, parallel language versions of certain metadata fields can now be searched via the aggregated search fields Who, What, When and Where. When selecting an object, the added multilingual terms can be seen under the auto-tag heading (Figure 6).

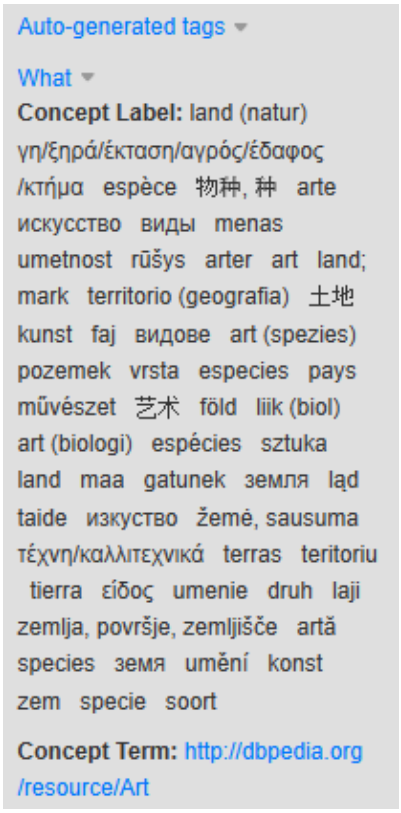

Figure 6. Multilingual tags provided by semantic tagging feature

2.5 Query Translation

The translation of queries for search was another core task in work package 2. A translation prototype incorporating sophisticated natural language processing techniques like named entity recognition (names are not always translated and have to be treated differently) and morphological analyzers (e.g. for part-of-speech tagging) was developed, which is capable of translating queries originating in any of the core or secondary languages to any of the core or secondary languages (a total of 90 translation pairs). The translation module exists as a prototype (Figure 7) and was tested on Europeana data and data from The European Library.

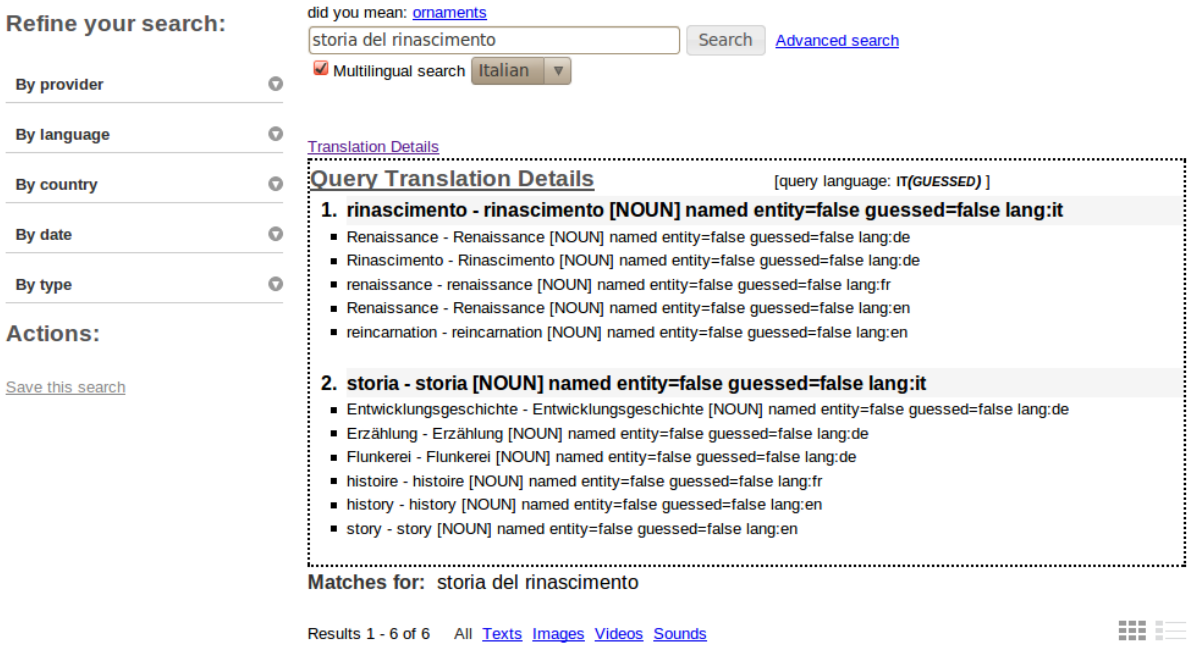

Figure 7. Query translation prototype developed for Europeana

2.6 Document Translation

Document translation is another form of multilingual result presentation. Although this prototype was not developed by partners of work package 2, it will be included here for completeness purposes. When selecting any object from a search result list, users can select to translate the metadata descriptions into their favorite languages (Figure 8). Translation capabilities are offered through the Microsoft translation API and Microsoft-offered languages can be selected for translation by the users.

All multilingual features already offered or developed are important steps towards helping the users navigate a multilingual environment and make informed choices about their available objects. However, some of features or capabilities await integration into the Europeana production system; others require improved interaction patterns for the users. Although different multilingual access options are requested by users (especially translation assistance), their integration into the search process is not always straightforward (see also section 4 on challenges).

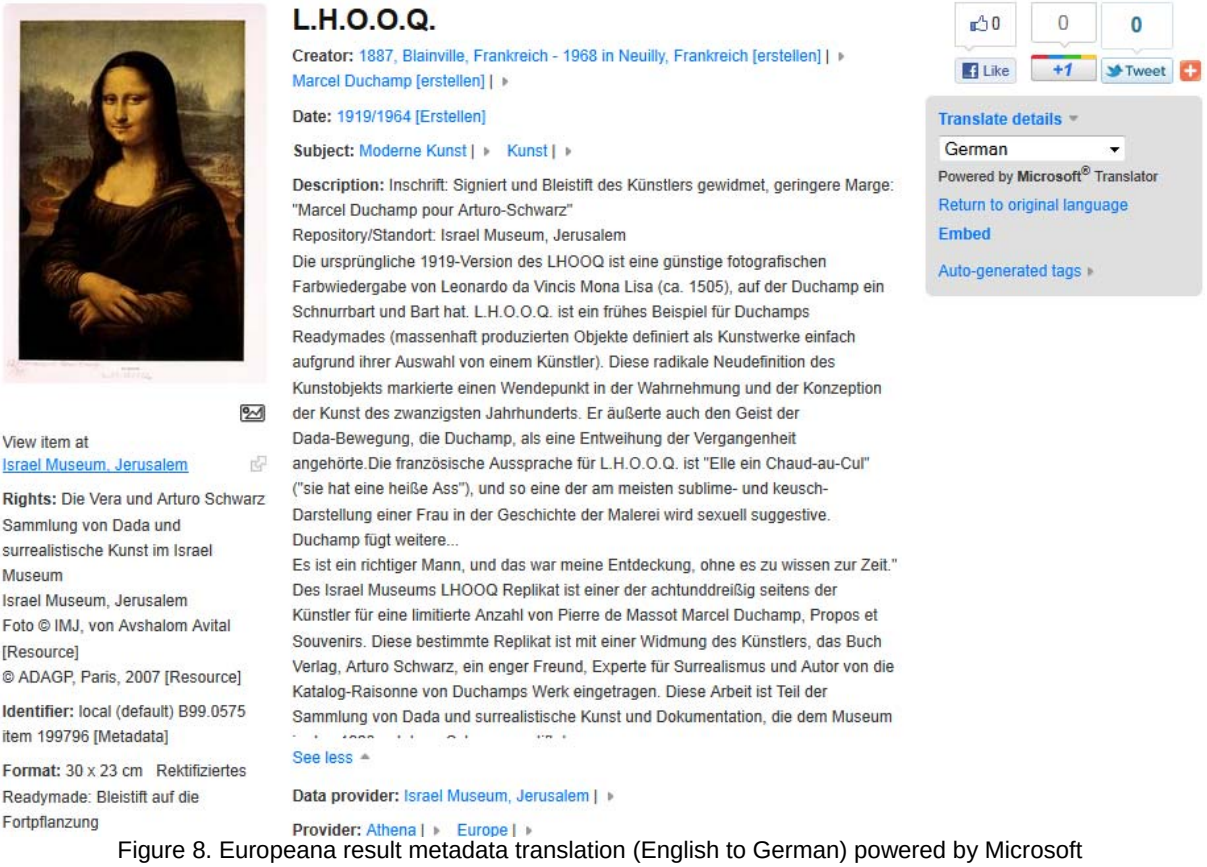

Figure 8. Europeana result metadata translation (English to German) powered by Microsoft

3 Facilitation & Exchange of Multilingual Access Strategies

Throughout the course of the EuropeanaConnect project, work package 2 partners connected with different research and developer communities to exchange ideas and facilitate uptake of applications and frameworks. User needs assessments and the development work were based on predecessor studies and applications and standard quality principles were followed. The work was introduced, presented and vetted in different communities, with a particular focus on the language resources and processing, semantic web and multilingual information retrieval communities. Some examples of these communicative and collaborative endeavours are described here.

3.1 Principles for Multilingual Access in Cultural Heritage

Large-scale EU-funded projects like Minerva and Calimera provided some of the groundwork for this project. Through the Europeana network, partners from these initiatives provided valuable feedback.

Minerva surveyed 657 multilingual websites from across Europe and presented best practices examples for multilingual websites and controlled vocabularies. They found that about three quarters of the analyzed websites and about half of the studied controlled vocabularies were already multilingual and provided useful suggestions for enhancing multilingualism in the cultural heritage sector (Minerva, 2006). Calimera also provided Guidelines for multilingualism in cultural applications (Calimera project, 2005).

Multilingual issues were regularly discussed in the core expert group of the Europeana v1.0 project WP3 (Further Specification of Functionality and Interoperability aspects of Europeana), to which several partners were invited.

3.2 Multilingual Digital Library Projects & Applications

Digital library projects that focus on multilingual information access particularly contributed to the development of the multilingual features in Europeana. Some of the partners overlapped so that a direct information exchange was possible, e.g. with the CACAO project (Levergood et al., 2008) or DISMARC (Koch & Scholz, 2009). Other projects and digital libraries like Multimatch (Amato et al., 2007) and OCLC (Gatenby, 2009) were invited through a workshop organized by work package 2 (MLIA4DL workshop, 2009) to share their experiences.

A summary of evaluation efforts detailing studies within CACAO, Multimatch, The European Library and Europeana was also recently submitted for a handbook on digital library evaluation (Petras et al., 2011).

3.3 Language Resources & Language Processing

In order to develop translation modules, language resources like dictionaries or lemmatizers need to be employed. Early on, contact was made with large European language resources and processing initiatives like CLARIN (Hinrichs, 2009) and Meta-Net (Ananiadou et al., 2011), of which several project partners are associated members. Europeana-specific challenges were also presented at the FLaReNet Forum (Dini & Petras, 2010).

Language processing solutions for translation problems that were integrated into the Europeana translation prototype were also presented to the larger IR community (Bosca & Dini, 2010a & 2010b).

3.4 Multilingual Semantic Web

During the project, several strategies and applications were developed for successful semantic mapping strategies. These were presented at relevant metadata and digital library conferences like Dublin Core (Isaac, 2009), ECDL (Wang et al., 2009) and TPDL (van Ossenbruggen, 2011). The automatic mapping techniques and mapping strategies were also discussed in the traditional library and knowledge organization communities (Petras, 2010). More importantly, they were also discussed and shared in the linked data and semantic web community (e.g. Tordai et al., 2010).

3.5 Multilingual Information Retrieval

Both the project (Crivellari et al., 2011) and specific multilingual requirements (Ferro, 2009; Gäde & Stiller, 2011) were introduced to the general information retrieval community. A particular focus was put on information retrieval evaluation of multilingual information systems as represented by the CLEF (www.clef-campaign.org) and PROMISE (www.promise-noe.eu) initiatives, in which several partners are involved (e.g. Braschler et al., 2010).

In 2008 and 2009, a multilingual information retrieval track with data from The European Library was organized at CLEF (Ferro & Peters, 2010), in which several partners participated in an organizing or experimental participation function (Bosca & Dini, 2010a & 2010b).

A particular focus was also put on logfile analysis of cultural heritage applications. The CLEF track LogCLEF (DiNunzio et al., 2011) is not only organized by EuropeanaConnect partners, but it also uses log data from The European Library (another cultural heritage project) to study multilingual user behaviour). Partners from work package 2 participated in this track to study specific aspects of multilingual digital libraries (Stiller et al., 2010; Gäde et al., 2011a).

A customized click-stream logger was suggested to study multilingual user patterns on Europeana and introduced to the multilingual IR community at the CLEF conference (Gäde et al., 2010).

In 2011, work package 2 partners also organized a CLEF workshop on cultural heritage information system evaluation (Gäde et al., 2011b). The CHiC workshop (CHiC http://www.promise-noe.eu/chic-2011/home) invited experts from the domain to speak about use cases and existing evaluation approaches in the cultural heritage domain. A standardized IR evaluation campaign for cultural heritage information systems was discussed. Europeana and its data were recommended to serve as a case study of a large-scale multilingual cultural heritage information system. The object data just recently (July 2011) released through the Europeana Linked Open Data Pilot (http://version1.europeana.eu/web/lod/) could serve as a data set that could draw attention from the semantic web community (metadata enrichment and semantic search) and the information retrieval community (structured data search). This work will continue after the EuropeanaConnect project has ended.

4 Challenges and Future Work

Even though a lot of progress has been achieved during the course of the EuropeanaConnect project, more work needs to be done to provide a seamless multilingual experience to all Europeana users. Several issues and challenges have to be resolved, specifically in order to improve the user experience of multilingual features.

An important issue for multilingual search, particular translation, is the licensing and availability of language resources for the required languages. With Google's announcement to deprecate the freely available Google Translate API and transfer into a paid service (Google 2011a & 2011b) and other free services possibly following, query and document translation strategies have to be considered and appropriate open source and licensable resources evaluated.

The query translation module has to be integrated into the European system and new languages slowly edited. Query translation quality is still relatively low. Chen & Bao, for example, compared Google Translate and Systran and found that queries were incorrectly translated between 24 and 80% depending on the query type (Chen & Bao, 2009).

Europeana considers multilingual semantic enrichment and indexing of parallel language versions of metadata fields more high priority than query translation and has started to implement those features as described in section 2.4. However, the usability of the auto-generated tag feature and the selection process for multilingual tags to be displayed still has to be improved and considered. If content is added from multilingual external resources, for example Wikipedia, a cultural bias can be transported through the different language versions (e.g. Callahan & Herring, 2011). Careful curation and selection of resources is necessary.

A particular problem is the integration of multilingual search features into the regular search process. Figure 9 shows an example of the Google multilingual search interface, where users manually add languages. An interesting feature is the opportunity for the user to edit suggested translations (e.g. see the wrong French translation of Mona Lisa). Using user input for editing translations is also interesting for Europeana, as a lot of idiosyncratic cultural heritage terminology might not appear in common dictionaries.

A particular user interface challenge is the adaption of (not only) multilingual features to other searching devices, for example, mobile phones with a much smaller screen estate.

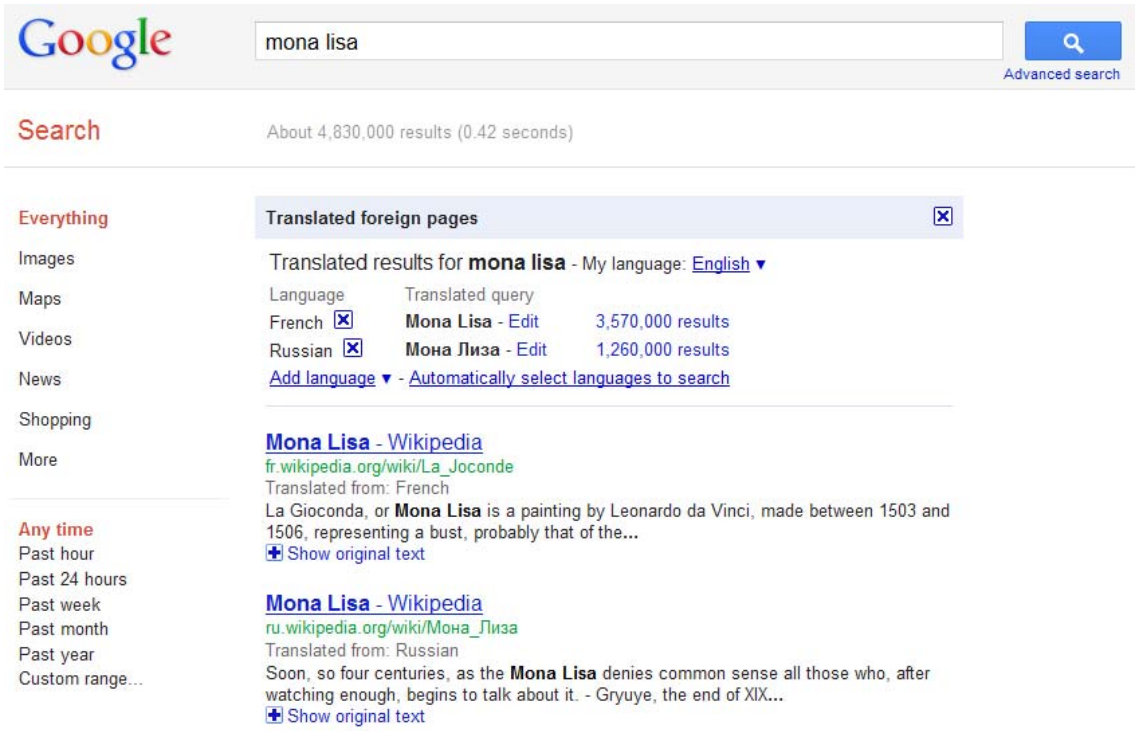

Figure 9. Google query and result translation with possibility to add more than one language and edit incorrect translations (e.g. French translation of Mona Lisa)

References

(All URLs were last tested on September 27, 2011.)

Agosti, M., Crivellari, F., Deambrosis, G., Ferro, N., Gäde, M., Petras, V., et al. (2009). D2.1.1: Report on User Preferences and Information Retrieval Scenarios for Multilingual Access in Europeana - EuropeanaConnect-Project. http://www.europeanaconnect.eu/documents/D2.1.1_eConnect_Report_User_Preferences_MLIA_v1.0_20091222.zip.

Amato, G., Cigarrán, J., Gonzalo, J., Peters, C., & Savino, P. (2007). MultiMatch – Multilingual/Multimedia Access to Cultural Heritage. In L. Kovács, N. Fuhr & C. Meghini (Eds.), Research and Advanced Technology for Digital Libraries. ECDL 2007 (Vol. LNCS 4675, pp. 505-508). Berlin / Heidelberg: Springer http://dx.doi.org/10.1007/978-3-540-74851-9_53.

Ananiadou, S., Thompson, P., Kano, Y., McNaught, J., Attwood, T. K., Day, P. J. R., et al. (2011). Towards Interoperability of European Language Resources. Ariadne: A Web & Print Magazine of Internet Issues for Librarians & Information Specialists, 30(67), 111-121.

Bosca, A., & Dini, L. (2010a). User Logs as a Means to Enrich and Refine Translation Dictionaries. In C. Peters, G. Di Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas & G. Roda (Eds.), Multilingual Information Access Evaluation I. Text Retrieval Experiments. CLEF 2009 (Vol. LNCS 6241, pp. 544-551). Berlin / Heidelberg: Springer http://dx.doi.org/10.1007/978-3-642-15754-7_67.

A. Bosca, L. Dini (2010b). Language Identification Strategies for Cross Language Information Retrieval. In: Martin Braschler; Donna Harman and Emanuele Pianta (Hg.): CLEF 2010 LABs and Workshops, Notebook Papers, 22–23 September 2010, Padua, Italy.

Braschler, M., K. Choukri, N. Ferro, A. Hanbury, J. Karlgren, H. Müller, V. Petras, E. Pianta, M. de Rijke and G. Santucci (2010). A PROMISE for Experimental Evaluation. Multilingual and Multimodal Information Access Evaluation. International Conference of the Cross-Language Evaluation Forum, CLEF 2010. M. Agosti, N. Ferro, C. Peters, M. de Rijke and A. Smeaton Eds.). Padua, Italy, September 2010, Springer. [Lecture Notes for Computer Science; 6360].

Calimera project. (2005). Calimera Guidelines. Cultural Applications: Local Institutions Mediating Electronic Resources. Multilingualism. http://www.calimera.org/Lists/Guidelines/Multilingualism.htm.

Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. Journal of the American Society for Information Science & Technology, 62(10), 1899-1915

Chen, J., & Bao, Y. (2009). Information access across languages on the web: From search engines to digital libraries. Proceedings of the American Society for Information Science and Technology, 46(1), 1-14 http://www.asis.org/Conferences/AM09/contributedpapers/78.pdf.

Chung, W. (2008). Web Searching in a Multilingual World. Communications of the ACM, 51(5), 32-40.

Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions - A New Framework Strategy for

Multilingualism. COM(2005) 596 Brussels, 22.11.2005. (2005). http://ec.europa.eu/education/languages/archive/doc/com596_en.pdf.

Cousins, J. (2010). Europeana Overview. Europeana Open Cultures Conference, 14-15 October Amsterdam.

Crivellari, F. and Deambrosis, G., Di Nunzio, G.M.,Dussin, M., and Ferro, N. (2011). EuropeanaConnect. In Proceedings of the Seventh Italian Research Conference (IRCDL 2011). Springer-Verlag, Heidelberg, Germany.

Dekkers, M., Gradmann, S., & Meghini, C. (2009). Europeana Outline Functional Specification: For Development of an operational European Digital Library. D 2.5 – Europeana Thematic Network Project. http://www.version1.europeana.eu/web/europeana-project/technicaldocuments/

Dini, L. and Petras, V. (2010). The Challenge of Multilinguality in Europeana: Web Services as Language Resources. Position paper for the FLaReNet Forum 2010: Language Resources of the future - the future of Language Resources, Barcelona, Spain, February 11-12, 2010.

DiNunzio, G.M., Leveling, J., Mandl, T. (2011). Multilingual Log Analysis: LogCLEF. In: P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, V. Murdoch (eds.) Proc. 33rd European Conference on IR Research (ECIR 2011), Lecture Notes in Computer Science, Springer.

Dobreva, M., McCulloch, E., Birrell, D., Feliciati, P., Ruthven, I., Sykes, J., et al. (2010). User and Functional Testing. Europeana v1.0 Final report.

European Commission. (2006). Europeans and their Languages. Special Eurobarometer, 243. http://ec.europa.eu/education/languages/pdf/doc631_en.pdf.

European Commission. (2011). User language preferences online. Flash Eurobarometer, 313. http://ec.europa.eu/public_opinion/flash/fl_313_en.pdf.

EuropeanaConnect (2009). Description of Work. ECP 528001.

Ferro, N. (2009). System Architecture and Multilinguality – MLIA Architectures: Standalone Applications and Services to Complex Systems with discussions about the TEL/Europeana relevant examples. Lecture at TrebleCLEF Summer School on Multilingual Information Access. 17 June 2009, Pisa, Italy.

Ferro, N. and Peters, C. (2010). CLEF 2009 Ad Hoc Track Overview: TEL & Persian Tasks. In Peters, C., Di Nunzio, G. M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., and Roda, G., editors. Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments - Tenth Workshop of the Cross-Language Evaluation Forum (CLEF 2009). Revised Selected Papers. Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany.

Gäde, M., Petras, V. and Stiller, J. (2010). Which Log for Which Information? Gathering Multilingual Data from Different Log File Types. Multilingual and Multimodal Information Access Evaluation. International Conference of the Cross-Language Evaluation Forum, CLEF 2010. M. Agosti, N. Ferro, C. Peters, M. de Rijke and A. Smeaton (Eds.). Padua, Italy, September 2010, Springer. [Lecture Notes for Computer Science; 6360].

Gäde, M., Stiller, J., Petras V. & Berendsen, R. (2011a). Interface Language, User Language and Success Rates in the European Library. CLEF (Notebook Papers/Labs/Workshops) 2011.

Gäde, M.,Ferro, N.,Paramita, M.L. (2011). CHiC 2011 – Cultural Heritage in CLEF: From Use Cases to Evaluation in Practice for Multilingual Information Access to Cultural Heritage.CLEF (Notebook Papers/Labs/Workshop) 2011.

Gäde, M., Stiller, J. (2011). Multilingual Interface Usage. Information und Wissen: global, sozial und frei? Proceedings des 12. Internationalen Symposiums für Informationswissenschaft. Boizenburg: Hülsbusch Verlag. 503-504

Gatenby, J. (2009). Metadata and multi-lingualism; OCLC projects and initiatives. Presentation at the MLIA4DL workshop. http://www.europeanaconnect.eu/documents/OCLC_Gatenby.pdf

Google (2011a). Spring cleaning for some of our APIs. The official Google Code blog. June 3, 2011. http://googlecode.blogspot.com/2011/05/spring-cleaning-for-some-of-our-apis.html

Google (2011b). Paid version of Google Translate API now open for business. The official Google Code blog. August 24, 2011. http://googlecode.blogspot.com/2011/08/paid-version-of-google-translate-api.html

Hinrichs, E. (2009). CLARIN - A Common Language Resources and Tools Infrastructure. Presentation at the MLIA4DL workshop. http://www.europeanaconnect.eu/documents/Clarin_Hinrichs.pdf

IRN Research. (2011). EUROPEANA - Online Visitor Survey http://www.version1.europeana.eu/c/document_library/get_file?uuid=334beac7-7fc2-4a4e-ba23-4dcc1450382d&groupId=10602.

Isaac, A. (2009). Linking Data for Europeana. Dublin Core conference (http://www.dc2009.kr/). Seoul, Korea, 13 October 2009.

Koch, W., & Scholz, H. (2009). DISMARC and BHL-Europe: multilingual access to two aggregation platforms for Europeana - CACAO Project. Paper presented at the Workshop on Advanced Technologies for Digital Libraries 2009. Retrieved from http://www.cacaoproject.eu/fileadmin/media/AT4DL/paper-07.pdf

Levergood, B., Farrenkopf, S., & Frasnelli, E. (2008). The Specification of the Language of the Field and Interoperability - Cacao-Project. Paper presented at the International Conference on Dublin Core and Metadata Applications.

Minerva Project. (2006). D6. Final Plan for using and disseminating knowledge and raise public participation and awareness Report on inventories and multilingualism issues: Multilingualism and Thesaurus. http://www.minervaeurope.org/structure/workinggroups/inventor/multilingua/documents/ReportonMultilingualism0512.pdf.

MLIA4DL workshop (2009). Multilinguality in Information Access to Digital Libraries: User Needs and Evaluation of multilingual resources use. MLIA4DL 2009, 9 September 2009, Trento (Italy). Workshop at the International Conference on Digital Libraries and the Semantic Web 2009 (ICSD2009). http://www.europeanaconnect.eu/MLIA4DL09Workshop.php

Petras, V. (2010). Multilingual Access to Content – The Europeana Experience. Invited talk at the EuroVoc conference, Luxembourg, November 19, 2010.

Petras, V., Stiller, J., Gäde, M., (2011). Building for Success (?) – Evaluating Digital Libraries in the Cultural Heritage Domain. In: Recent Developments in the Design, Construction and Evaluation of Digital Libraries. Colleen Cool & Kwong Bor Ng (Eds.) Submitted – publication pending.

Stiller, J., Gäde, M. and Petras, V. (2010). Ambiguity of Queries and the Challenges for Query Language Detection. CLEF 2010 LogCLEF Workshop. In: CLEF 2010 Labs and Workshops Notebook Papers. M. Braschler, D. Harman and E. Pianta (Eds.). Padua, Italy, 22-23 September 2010.

Tordai, A., van Ossenbruggen, J. R., Schreiber, G., and Wielinga, B. (2010). Aligning Large SKOS-Like Vocabularies. In: Proceedings of European Semantic Web Conference 2010 (7) (volume 6088, pages 198 - 212), Lecture Notes in Computer Science, Springer.

van Ossenbruggen, J. R., Hildebrand, M., and de Boer, V. (2011). Interactive Vocabulary Alignment. In: Proceedings of the International Conference on Theory and Practice of Digital Libraries 2011, Springer.

Wang, S., Isaac, A., & Schopman, B. (2009). Matching Multi-lingual Subject Vocabularies. Lecture Notes in Computer Science (Vol. 5714, pp. 125-137). Berlin, Heidelberg http://www.springerlink.com/content/u8u2108543252x41/.

Zhang, J. & Lin, S. (2007). Multiple language supports in search engines. Online Information Review, 31(4), 516-532.