# **MILESTONE**

**Project Acronym:Ev3**
**Grant Agreement number:620484**
**Project Title: Europeana Version 3**

## **MS12:  WHITE PAPER ON BEST PRACTICES FOR MULTILINGUAL ACCESS TO DIGITAL LIBRARIES**

| Revision | Final |
|---|---|
| Date of submission | 30.05.2015 |
| Author(s) | Juliane Stiller (Humboldt-Universität zu Berlin)<br>Vivien Petras (Humboldt-Universität zu Berlin)<br>Maria Gäde (Humboldt-Universität zu Berlin) |
| Dissemination Level | [Public] |

**Project co-funded by the European Commission within the ICT Policy Support Programme**

# REVISION HISTORY AND STATEMENT OF ORIGINALITY

## Revision History

| Revision No. | Date | Author | Organisation | Description |
|---|---|---|---|---|
| 1 | February 5, 2015 | Juliane Stiller | Humboldt-Universität zu Berlin | 1st Draft |
| 2 | March 15, 2015 | Vivien Petras | Humboldt-Universität zu Berlin | Strcuture and Table of Content |
| 3 | May 6, 2015 | Juliane Stiller | Humboldt-Universität zu Berlin | 2nd Draft |
| 4 | May 15, 2015 | Maria Gäde | Humboldt-Universität zu Berlin | 3rd Draft |
| 5 | May 16, 2015 | Vivien Petras | Humboldt-Universität zu Berlin | 4th Draft |
| 6 | May 20, 2015 | Valentine Charles | Europeana Foundation | Comments and Edits |
| 7 | May 26, 2015 | Juliane Stiller, Vivien Petras | Humboldt-Universität zu Berlin | Final Draft |
| 8 | May 27, 2015 | Antoine Isaac | Europeana Foundation | Comments and edits |
| 9 | May 29, 2015 | Juliane Stiller | Humboldt-Universität zu Berlin | Final version |

## Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# Table of Contents

# 1. **Introduction**

Cultural heritage information systems are digital libraries that aggregate digitized or born-digital cultural heritage objects and present these and/or their representations to users through various access channels (Petras et al, 2013). The main goal of a cultural heritage information system is to let users find, explore and engage with cultural heritage objects. Often, as it is the case for Europeana, these objects are in different languages, most of which the user cannot understand. The information system should ideally bridge this gap letting users find objects in languages different from their native one. This White Paper explores the different dimensions of multilinguality in cultural heritage digital libraries and provides recommendations and best practices for implementing multilingual access to digital cultural heritage content.

## 1.1 Levels of Multilinguality in a Cultural Heritage Information System

Most digital cultural heritage objects are not text-, but image-based and depict a painting, a statue or any other item of cultural value. These objects do not have a language and could be enjoyed regardless of the user's language skills. Still, retrieving them can be tricky as they are searched through their metadata, which is text in a certain language. The metadata language can correspond with the language of the objects (if they have a language like full texts of monographs, for example) but does not have to. It therefore constitutes another level of multilinguality in a cultural heritage information system. The access system, which is often a search engine, determines how the metadata and therefore the object is retrieved. This is usually initiated by the query that users formulate to articulate their information need. This query - search - result list interaction is one of the interactions users will perform in the system but many more actually occur. Another major interaction is the actual navigation of a user through the information system, which is a language-dependent activity. The user interface needs to be localized by language to be understandable and therefore usable for users. As this is the first encounter of users with the system and the content, special care needs to be put into its multilingual interaction design. If it fails at this point, the other levels cannot be explored (Bates, 2002). Figure 1 shows the four different layers of a typical information system with all of these levels having a multilingual perspective, namely the user interface, the user interactions, the access system and the underlying data (metadata and / or objects).



*Figure 1: The different layers of accessing information in a cultural heritage information system.*

## 1.2 Structure and Objective of Paper

The White Paper is structured in four big parts which group issues, best practices, solutions and findings to one of the main components for successful cultural heritage systems in a multilingual environment: the underlying data (chapter 2), the user interface (chapter 3), the user interactions (chapter 4). Chapter 5 highlights specific problems and challenges in multilingual systems and chapter 6 deals with the evaluation component.

# 2. Making your Data Multilingual

Providing multilingual access to content does not only mean offering objects in several languages but also their describing metadata. Multilingual metadata descriptions help in crossing the language barrier between the object's and the user's language. Translating metadata is cost- and labour-intensive but it is not uncommon for cultural heritage institutions serving communities which are bi- or multilingual (see also 2.4). This chapter describes multilingual options to make the content - metadata and objects alike - more multilingual to increase the options for user access.

## 2.1 Language Attributes

A first step for increasing multilingual access to cultural heritage content is to indicate the language of your metadata.

Motivation:
Adding language properties to metadata identifies the language of the text and supports re-use and processing of the metadata in a multilingual environment. If the metadata is available in several languages, a language attribute helps to display the right language to the user. Language attributes also support the translation of the text in other languages for searching or browsing. Language identification of metadata is also crucial for mapping fields to multilingual vocabularies, which eventually increases the searchable data for an object.

Study Findings & Further Reading:

| Findings and Summaries | Source |
| --- | --- |
| Europeana suggests to add language tags to identify multiple records in different languages for the same object. | http://pro.europeana.eu/share-your-data/data-guidelines/edm-case-studies/data-multilinguality |

Best Practices:
- Language tags can be added to the metadata or to individual text string values within the metadata. This makes sense especially if controlled vocabularies describe the objects and have different language variants. In this case, each keyword from the vocabulary would have the language tag attached. Language agnostic information systems could then display the appropriate language according to the user's preferences.
- To identify the language of metadata, automatic language detection could be considered, but only if the text is sufficient for accurate detection.

Example:

```xml
<?xml version="1.0"?>
- <rdf:RDF xmlns:local="#local-functions" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
   - <skos:Concept rdf:about="http://www.mimo-db.eu/InstrumentsKeywords/2235">
       <skos:prefLabel xml:lang="">Harmonium</skos:prefLabel>
       <skos:prefLabel xml:lang="en">Harmonium</skos:prefLabel>
       <skos:prefLabel xml:lang="fr">Harmonium</skos:prefLabel>
       <skos:prefLabel xml:lang="it">Harmonium</skos:prefLabel>
       <skos:prefLabel xml:lang="de">Harmonium </skos:prefLabel>
       <skos:prefLabel xml:lang="nl">Harmonium </skos:prefLabel>
       <skos:prefLabel xml:lang="sv">Harmonium</skos:prefLabel>
       <skos:prefLabel xml:lang="">Harmònium</skos:prefLabel>
    - <skos:broader>
       - <skos:Concept rdf:about="http://www.mimo-
         db.eu/InstrumentsKeywords/2233">
           <skos:prefLabel>Reed organs</skos:prefLabel>
           <skos:inScheme rdf:resource="http://www.mimo-
             db.eu/InstrumentsKeywords/"/>
       </skos:Concept>
    </skos:broader>
   </skos:Concept>
</rdf:RDF>
```

*Figure 2: Language tag of a concept in the MIMO thesaurus[1].*


## 2.2 Creating Multilingual Vocabularies through Mapping and Translation

Creating multilingual vocabularies can be necessary to provide multilingual access to specialized collections. This can happen through the mapping of monolingual vocabularies in different languages or through the translation of monolingual vocabularies in other languages.

Motivation:
Mapping your controlled monolingual vocabulary to multilingual vocabularies will enable cross-lingual search for users, which do not speak the language(s) your content is offered in. Adding controlled multilingual vocabularies will help users retrieve objects and determine their relevance. Furthermore, manual translation of existing vocabularies and manual term translation is beneficial for very specialized domains where no other language resources exist.

Study Findings & Further Reading:

| Findings and Summaries | Source |
| --- | --- |
| Different workflows for multilingual mapping of vocabularies for building new multilingual vocabularies can be found in various reports and deliverables of cultural heritage projects. | EuropeanaConnect D2.3.1, 2011; PartagePlus D3.1, 2012; EuropeanaPhotography D4.1, 2013; Europeana Fashion 3.3 (a), 2013, |
| List of Vocabularies in the cultural domain are available, which can be used for data enrichment and mapping resulting in new multilingual vocabularies. | EuropeanaConnect WP2.3, 2011 |

---

[1] http://www.mimo-international.com/vocabulary.html

Best Practices:
- Use controlled vocabularies or authority lists that have multilingual elements.
- Use open vocabularies such as VIAF[2] for names, EuroVoc[3], AAT[4], or MACS (mapped LCSH/Rameau/SWD) (Landry, 2009) for subject headings.
- Browse the FLOSS[5] inventory for finding suitable mapping tools and software.
- There is no strict mapping strategy but it seems in general easier/more productive to map from vocabularies in few languages with specialized coverage to more general vocabularies that cover more languages.

Example:

**Deportation**

**Institution:**
Europeana Foundation

**Collection:**
Europeana

**URI:**
http://data.europeana.eu/concept/loc/sh85037040

**SKOS Class:**
http://www.w3.org/2004/02/skos/core#Concept

**inScheme**

- http://data.europeana.eu/concept/loc

**LexicalLabels**

skos:prefLabel

- Deportation (en)
- Déportation (fr)
- Deportazione (it)
- Deportatie (nl)
- Deportation (da)
- Deportation (de)
- Deportacija (sr)

**SemanticRelations**

*Figure 3: For Europeana 1914-1918, translations of concepts were added to the original LCSH[6].*

## 2.3 Multilingual Semantic Enrichment

If your own metadata is monolingual, multilinguality can be added by linking and mapping your metadata elements to multilingual vocabularies and authority files. Multilingual semantic enrichments add equivalent or semantically related (e.g. broader or narrower concepts) concepts (in several languages) to the metadata. The links created can be further exploited and semantically related keywords and/or translations added to improve retrieval.

Motivation:
Next to the multilingual dimension, adding additional terms to the metadata helps to contextualise the cultural heritage objects and makes them easier to retrieve.

---

[2] https://viaf.org/

[3] http://eurovoc.europa.eu/drupal/

[4] http://www.getty.edu/research/tools/vocabularies/aat/

[5] 300 Free, Libre Open Source Software relevant for the cultural heritage domain:
https://docs.google.com/spreadsheet/ccc?key=0Ag_7rVJwt0CpdFRJOEJxdEk4ZEMxQ01jaDgxQXFSTkE#gid=0

[6] http://id.loc.gov/authorities/subjects.html

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| A semantic extraction market study lists the technical options for semantic feature extraction. | https://www.assembla.com/spaces/europeana-r-d/wiki/Semantic_feature_extraction_-Market_Study |
| Enrichment workflows in cultural heritage digital libraries are described in several case studies. | Manguinhas, 2014; Freire, 2013 |
| The quality of enrichments in the cultural heritage domain has been described and evaluated in several studies. | Stiller, Isaac & Petras (eds.) 2014; Olensky et al. 2012; Stiller et al, 2014; Stiller et al, 2014 |

Best Practices:
- Enrich monolingual metadata with multilingual vocabularies through mapping and linking terms.
- Establish an enrichment process and a sustainable enrichment strategy for continuous updates.
- Establish criteria for selecting suitable vocabularies for your particular content.
- Establish enrichment rules for your particular content.
- Match the language of the metadata with the language of the vocabulary.

Example:



*Figure 4: Enrichment of the dc:creator field with language variants for "Johannes Vermeer" from DBpedia[7].*

## 2.4 Translating Multilingual Metadata and Multilingual Objects

Creating multilingual metadata is not uncommon. Titles, descriptions or abstracts and sometimes keywords are stored in the native institutional language and another language - most often English - in order to comply with institutional objectives and requirements. It is rare that textual content in cultural heritage information system is multilingual. Sometimes, texts are provided together with translated versions. Images and nonverbal videos or audio recordings (e.g. music)

---

[7] http://wiki.dbpedia.org/

are inherently multilingual. Object translation enables a cultural heritage information system user to access content in a non-native language after it has been found or selected. Multilingual metadata and multilingual objects have also implications for search result representation.

Motivation:

Often the motivation for multilingual metadata is rooted in the number of official languages spoken in the country the cultural heritage institution is residing. Countries like Belgium or Switzerland often need to provide their metadata in all the officially spoken languages.
After users identified a relevant object to satisfy their information need, they also should be able to understand this object even if it is not in their preferred or native language.

Study Findings & Further Reading:

| Findings and Summaries | Source |
| --- | --- |
| Metadata translation is often sufficient in order to support a user's decision if an object is relevant or not. | Oard et al., 2004; Gonzalo et al., 2008; Minelli et al., 2006; Clough and Sanderson, 2006 |
| Result representations depend on the user's language skills and information need. Merged or language separated result lists should be available. | Gonzalo et al, 2008 |
| Clear separation of languages in result lists is preferred. | Steichen and Freund, 2015 |

Best Practices:
- At least the metadata should be displayed in a language the user understands.
- Automatic translations on object level can be offered using external translation services.

Examples:



*Figure 5: Example of metadata translation on the object label. Here from Dutch to French using the external Microsoft translator.*



*Figure 6: Example of a Spanish book in the International Children´s Digital Library with volunteer translations in 5 additional languages (see dropdown menu).*

# 3. Making your User Interface Multilingual

The user interface is the first encounter the users have with the cultural heritage information system. It is what greets them and invites to explore and engage with the cultural collection - ideally in a language the users understand. This chapter describes functionalities to make the interface more multilingual. The language options of the user interface include the display language of all menu items and the static content.

## 3.1 Multilingual Static Pages

A first and simple step in achieving multilinguality is the translation of all static pages in your cultural heritage information system. Because these do not change often, the effort required is relatively low and non-native speakers can determine the relevance of the site for their uses when switching to their own preferred language.

Motivation:
Through the translation of the static pages and menu items of the system, the user can enter and navigate it in their preferred language.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| The most frequently used interface language is English. | Gäde, 2014;  Angelaki, 2007; Agosti et al., 2007; Clark et al., 2011; Oakes et al., 2009; Keegan and Cunnigham, 2005 |

Best Practices:
- Make sure the different language versions are accessible through search engines.
- Follow design convention for menu items and navigation to make access to your system easier, e.g consistent labelling.
- Make sure the user can switch between the different language versions you are offering at any point.

Example:



*Figure 7: Homepage of Europeana1914-1918 in German.*

## 3.2 User Language Detection

Detecting the user's native or preferred language is a first step in providing customized multilingual services to users.

Motivation:
By identifying the user's preferred language the appropriate interface language version can be served and customized language and / or location content could be provided.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| Users prefer their site in their native language and they are also more likely to visit a site in their preferred language. | Agosti et al., 2007; Agosti et al., 2009; Dobreva et al., 2010; Gäde & Petras, 2014 |

Best Practices:
- Use automatic detection of the user language if at all possible.
- If users indicate a language preference, a cookie should be set and the preference should be stored for future interactions.
- Language preferences should be clearly demarcated in the user profile.
- Users should always be able to easily switch their language preferences even when it is automatically detected.

Example:



*Figure 8: Automatic detection of language preferences through the browser locale in Europeana.*

## 3.3 Interface Language Change

Changing the interface language - and with it all static content and interaction functionalities (e.g. search buttons) - when a user accesses the site provides a multilingual starting point.

Motivation:
Letting users adapt their interface language improves the overall user experience by making the portal more familiar and usable.

Study Findings & Further Reading:

| Findings and Summaries | Source |
| --- | --- |
| Flags instead of language names can lead to confusion, although they might give an indication where on the page the language change button is located. | http://flagsarenotlanguages.com/blog/best-practice-for-presenting-languages/ |
| Users prefer automatic solutions where their native language is detected, they hardly trigger the language change themselves. | Agosti et al., 2007; Agosti et al., 2009; Dobreva et al., 2010; Clark et al., 2011; Oakes et al., 2009; Keegan and Cunnigham, 2005 |

Best Practices:
- It should be very clear what a language change is impacting - the interface language, the language of the search or the language of the collection searched in.
- Languages should appear in the local name or be displayed according to the conventions for language codes (e.g. ISO 639-2).
- Flags as representations of a language should not be used although they could hint at multilingual content and culturally responsive interactions.
- Avoid language mixes which might occur when static content is translated but dynamic content stays in the original language.
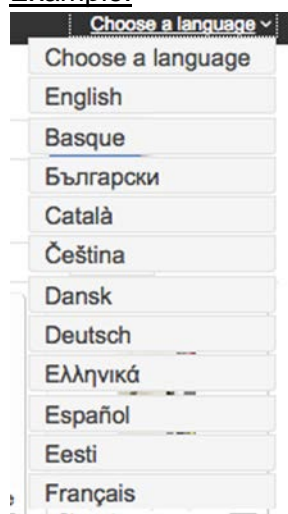
Example:



*Figure 9: Language drop-down menu of Europeana showing languages in their respective language.*

# 4. Making your User Interactions Multilingual

This chapter focuses on the user interactions in the information system. If the cultural heritage information system has multilingual users and multilingual content, the system providers need to put extra care in crafting their user interactions. Making user interactions multilingual is supported by creating access points for content in languages different than the user's preferred language. This does not only cover search across different languages but also features for browsing and engaging the user. On the other hand, there are also several features for content discovery which can be considered language independent and are therefore recommended for use when multilinguality through translation or other options is not guaranteed.

## 4.1 Query Auto-Completion & Query Suggestions

Query completion and query suggestions can show the searcher what queries will be successful and what content can be expected when accessing the cultural heritage information system.

Motivation:
Query completion or query suggestions are more targeted and helpful when provided in a language the user understands. Query suggestions can support users in formulating queries, recommend search terms and avoid spelling mistakes.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| Dynamic query suggestions and auto-completions are becoming standard in search engine interfaces. | Hearst, 2014, chapter 4 |
| Query suggestion and recommendation services help users in finding what they are looking for. | Assets D2.2.1, 2012 |

Best Practices:
- Dynamic query suggestions should be timely.
- The query suggestion should only suggest queries which retrieve objects.
- Auto-suggestions should be in the user's preferred language.

Example:



*Figure 10: Query suggestion in Europeana with number of results the query would retrieve.*

## 4.2 Automatic Query Translation

Query translation is a major step for digital libraries to expose users to content they otherwise would never find. Correctly translating the query often requires to identify the query language beforehand. It is also very helpful in constructing queries with language variants. A query expanded by the translations of the query can be generalized by adding language variants to a query component using the Boolean OR operator.

Motivation:
Automatic query translations help to cross the language barrier and retrieve objects in the languages they are described in.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| Users often struggle with the selection of appropriate translation candidates. | Petrelli et al., 2002; Gonzalo et al., 2008 |
| Automatic language detection for query translation can be harmful as user queries are often very short and especially in the cultural heritage domain dominated by named entities. | Stiller et al, 2013 |
| Important aspects regarding the implementation of query formulation and translation to multilingual information systems can be enumerated. | Peters et al, 2012 |
| Europeana implemented a query translation process using parallel language Wikipedia versions. | Kiraly, 2015 |

Best Practices:
- Offer automatic query translation with limited suggestions for when the translation process fails.
- For query expansion, compound and phrase queries (e.g. "apple tree") need to be identified correctly.
- Controlled vocabularies and named entity recognition tools should be incorporated in the query translation process.
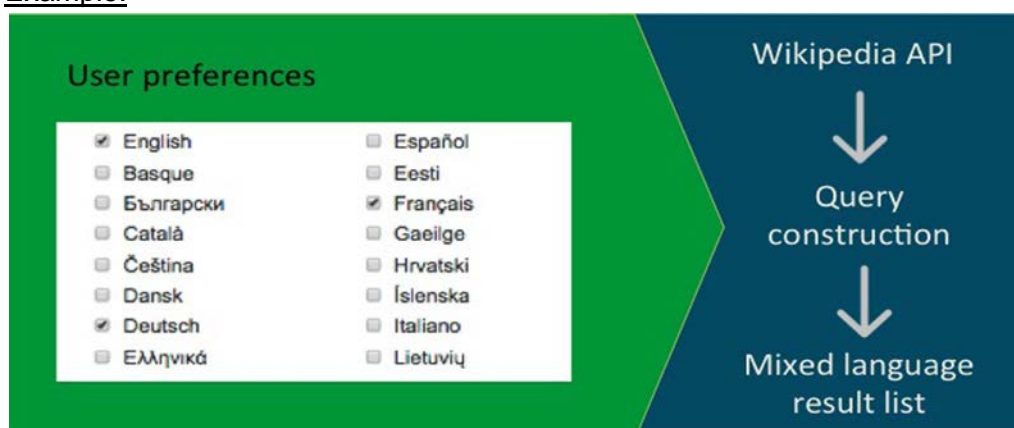
Example:



*Figure 11: Visualization of the query translation workflow in Europeana.*

## 4.3 User-Assisted Query Translation

User-assisted translation either makes use of indirect user input such as query logs or directly involves the user into the translation process. It is still an open issue how the quality of user-generated input should be controlled and measured. Interactive systems need to support and encourage the user to participate in the search process. User-assisted translation is a multi-level process that includes several steps where user input can be leveraged, such as determining the source query language, determining the target language(s), select translation offered by the system. Therefore, it is essential to find the balance between transparency of the system and overloading the interface or the user with too complex interaction steps.

Motivation:
User-assisted translations do not only support the system in adding domain-specific translations to their dictionary, but also provides the user with more control over the system functionalities.

Study Findings & Further Reading:

| Findings and Summaries | Source |
| --- | --- |
| Users want to control the query translation process, advanced search functionalities should support the user-assisted query translation. | Agosti et al., 2009; Gonzalo et al., 2008 |
| Users tend to search in their native language and only repeat queries in foreign languages if the result set is not satisfying. | Srinivasarao et al., 2008; Aula and Kellar, 2009; Trojahn and Siciliano, 2009; Ghorab et al., 2010; Leveling et al., 2010; Marlow et al., 2008 |

Best Practices:
- The translation workflow should not require too much effort from the user's side and required clicks need to be minimized.
- Options for users to edit the translation should be easily visible in the system.

Example:



*Figure 11: Mock-up of different possibilities for user-assisted query translation interactions. Taken from Europeana v2.0 D7.7*

## 4.4 Browse

Searching might not be the optimal access interaction for cultural heritage information systems. Browsing allows users to receive an overview of what the cultural heritage information system contains, it helps to provide more guided access to the content and supports serendipity.

Motivation:
Browsing allows users to get an overview of the offered collections and objects within an information system. It also helps to access the collection without a clear information need. Browsing functionalities also help to present data from different viewpoints letting users explore relations between different items.

Study Findings & References:

| Findings and Summaries | Source |
|---|---|
| Information needs and strategies are influenced by the user´s language background, the system should provide different access and assistance points | Lamm et al., 2010; Keegan and Cunnigham, 2005; Wu et al., 2012 |

Best Practices:

- When providing browsing access, make sure to provide multilingual options, such as multilingual user interfaces and vocabularies for browsing.
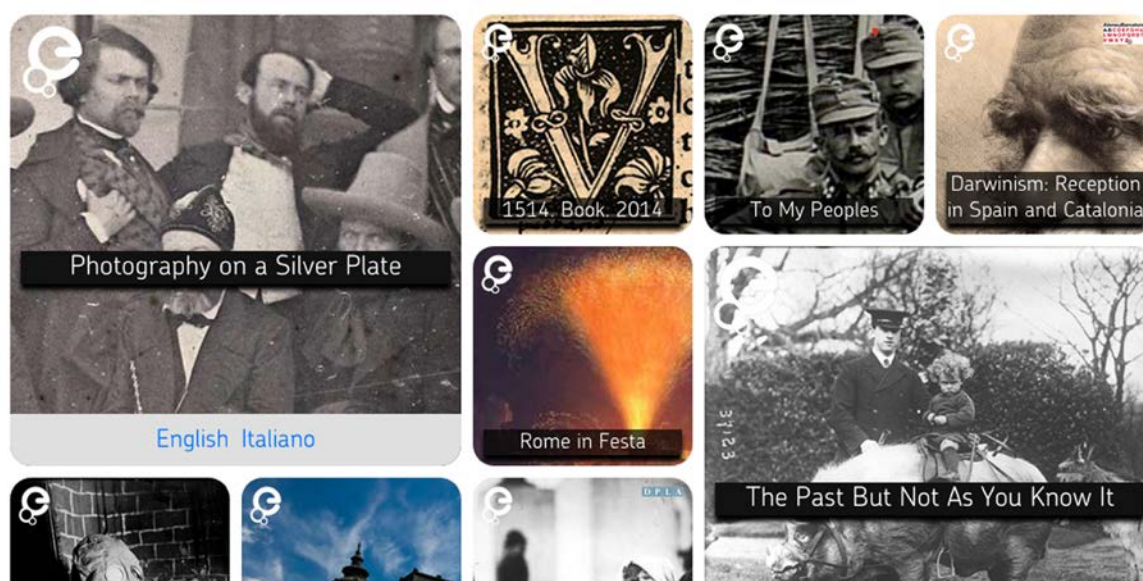
Example:



*Figure 12: Browsing interface for curated exhibitions in Europeana - the first one offered in English and Italian.*

## 4.5 Search and Browse Result Filtering

When reaching a certain scale, all digital libraries need to provide filter or drill-down options for users, so that the number of displayed objects can be reduced to a manageable amount.

Motivation:
Options for filtering by language are a natural way for users to reduce their result set in a sensible way.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| Users tend to refine results by language. | IRN Research, 2009; Bilal and Bachir, 2007; Gäde, 2014 |
| Users rarely refine results by country information. | Gäde, 2014 |

Best Practices:
- Provide facets, which let users refine results by language.
- Make clear, what is meant by the language filter - objects or the metadata.
- Provide language refinement options in advanced search interface as well as a facet.
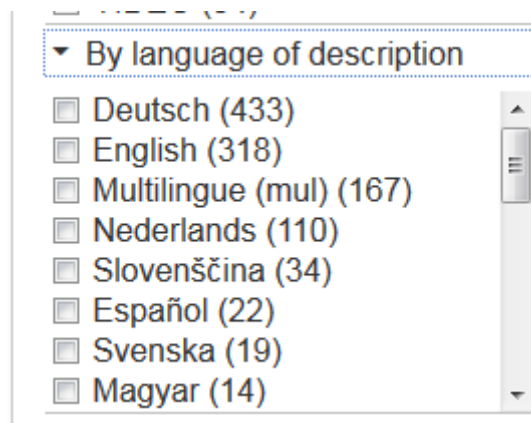
Example:



*Figure 13: Europeana's language filter describes what is filtered: the metadata language.*

## 4.6 Language-Independent Access Options

Besides searching or browsing a text-based categorization system, other access options provide alternative entry points into digital libraries. Features like timelines or map displays, which are not text-based, are language-independent and can be provided in any multilingual cultural heritage information system.

Motivation:
Language-independent access option are an optimal way for access and presenting information in a different way than via a search box or through browsing facilities. It allows the user to discover new aspects and present the data from a new perspective.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| The majority of users do not understand the relation between query and object language, language independent access points can overcome language barriers. | Peinado et al., 2008 |

Best Practices:
- Probably most common way to present information is through its spatial or temporal aspects. One thing to remember is that spatial information can refer to several aspects in the life of a digital object, e.g. to its place of creation or the place it is representing.
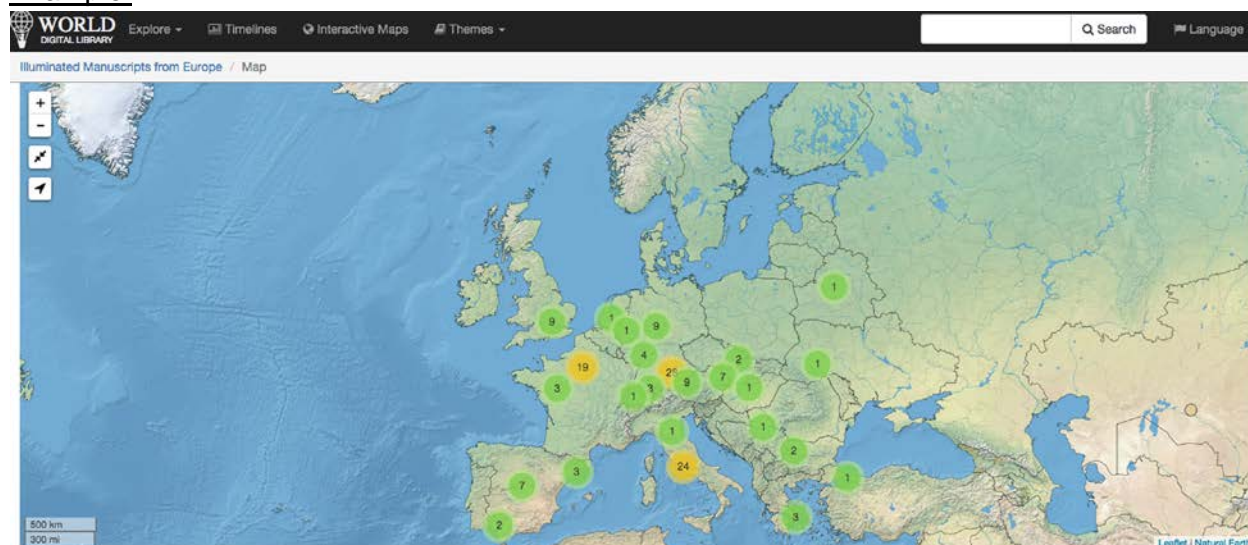
Example:



*Figure 14: Map Browsing of the World Digital Library[8].*


## 4.7 Site Structure and Search Engine Landing Pages

You do not only want users to find their preferred language version of your site but also design your site search engine friendly.Making sure web search engines know which language version you are offering helps in serving users the right one in the search results.

Motivation:
For users coming from search engines to a landing page they do not understand, the language switch should be easy to find and identifiable.

Study Findings & Further Reading:

| Findings and Summaries | Source |
|---|---|
| The choices made with regard to different language versions of the site also impact search engine findability. | Google:<br>https://support.google.com/webmasters/answer/182192?hl=en#1<br>Bing:<br>http://www.bing.com/webmaster/help/geo-targeting-your-website-b7629197 |

Best Practices:
- Offer pages and object landing pages in different language versions.

---

[8] http://www.wdl.org/

Example:



*Figure 15: Different language version of Europeana 1914-1918[9] in Google search results.*

## 4.8 Multilingual User-Generated Content

In interactive cultural heritage information systems, users contribute metadata or content themselves to the system via social media or tagging and other forms of annotations, for example.

Motivation:
Users might help in translating objects or adding tags in multiple languages to existing metadata. Both can help to improve multilingual access in an information system.

Study Findings & Further Readings:

| Findings and Summaries | Source |
|---|---|
| User-generated multilingual data can be used to improve the user search experience. | Stiller et al., 2011 |
| For multilingual user tagging, cultural perspectives are expressed in less frequent tags. | Eleta and Golbeck, 2012 |

---

[9] www.europeana1914-1918.eu/

Best Practices:
- Design interfaces, which let users determine the language of their contribution.
- Be aware of cultural diversity of user contributions.

Example:



*Figure 16: Tagging feature of Steve.Museum where users determine the language of their added tag.*

# 5. Overcoming Challenges in Achieving Multilinguality

When building and implementing a multilingual system, system providers can follow the suggestions outlined above and try to follow and implement best practices. Nevertheless, in multilingual environments, one still has to look out for some common issues and impracticalities which might have a negative influence on the user experience. Some of these overarching concerns and issues are addressed in this chapter.

## 5.1 Avoiding the Language Mix

When thinking about language displays, several things need to be considered to avoid a language mix on the site which will reduce the user's satisfaction and interfere with the usability of the whole site. When thinking about a translation of the content and structure of the site, providers should consider all the different elements on the pages which need to be translated. This could be applicable to:
- the menu items,
- automatically pulled content coming from blogs, news feeds, etc. also dynamic content,
- metadata of objects,
- query suggestions and auto-completion features.



*Figure 17: Europeana search result page with language mix in the option and filtering panel.*

## 5.2 Distinguishing between the Object and the Metadata Language

Users often do not understand institutional conventions regarding the language of objects or metadata. In the case of Europeana, the language of a digital object - if unknown - is determined by the language of the providing institution which can be very different from the language of the metadata and the language of the object itself.
Especially if this information is used in filtering results, it should be made very clear what it refers to.

## 5.3 Managing Expectations for Automatic Processing

As shown in the previous chapters, many of the solutions for offering multilingual access to cultural heritage content include automatic processing of large amounts of data. Often these automatic processes will work with well-curated data applying techniques that make this data more accessible across languages. Automatic processes work best on a large scale but might flatten information or even introduce errors where data is ambiguous, not explicit or just too complex. Keeping these errors rates low and monitoring the effects of technical solutions is crucial (chapter 6). Overall, the benefits of bridging the language gap might outweigh the negative impact automatic solutions might have on parts of the data – providers need to check regularly on what goals and objectives can be achieved with it.

## 5.4 Providing Sustainable Multilinguality

Offering multilingual solutions, which bridge the language gap in an information system, is not only a decision made at the beginning of an information system development project but an ongoing endeavor adjusting to the needs of users and the content offered over the course of time. If language resources are used at any point, they need to be updated and adapted to the evolving needs of the collections and / or the users. As this can be resource-intensive, the objectives of the platform should guide and steer the decisions made in this regard. These objectives will decide if the commitment to multilinguality is a one-timer or a permanent effort. For example, the using an external translation service for metadata translation on object level might come with fees. On the one hand, an external translation services can lower costs through less maintenance, on the other hand, it might be expensive if fees apply which are based on the amount data processed. One should also keep in mind that language technology is constantly evolving; technical solutions which were sufficient a couple of years ago might be outdated and insufficient for today's requirements.

# 6. Evaluating your Multilingual Components

The approaches and best practices recommended in this report need to be evaluated and tested in any particular cultural heritage information system implementation. This chapter introduces and references methodologies and measures that can be used in evaluating components of your multilingual cultural heritage information system.

## 6.1 Evaluating your Data

Evaluating data - the basis of each information system - is of utmost importance. If the quality of the underlying data is insufficient, the system cannot be used as intended and it will lead to bad user experiences. Europeana has launched a series of initiatives to tackle the issue of metadata quality in its portal. A task force was launched which recently published its results and gave recommendations on how the quality of the metadata can be improved (Dangerfield and Kalshoven, 2015).

To raise awareness of metadata quality, a special issue of the EuropeanaTech Insight journal focused solely on metadata quality (link) and the EuropeanaTech Meeting 2015[10] dedicated a whole session with round table discussion to this issue[11]. One of the main findings was that the quality of the metadata is defined by its purpose. So far, several frameworks tried to define the quality of metadata, but a consensus what constitutes high-quality data has not yet been reached (Calhoun et al., 2011; Baierer et al., 2014).

Initiatives for mapping vocabularies have been evaluated several times, both for manual (e.g. Mayr & Petras, 2008) and automatic mappings (e.g. Isaac et al., 2009). The OAEI library track (e.g. Dragisic et al., 2014, http://oaei.ontologymatching.org/2014/library/results.html) regularly evaluates linked data-based systems in their capabilities to map RDF-based vocabularies. The evaluation of automatic enrichment, which often helps in bridging the language gap through the use of language variants, is also not yet standardized but has helped realizing that both the quality of the enrichments themselves as well as their impact on the information retrieval output should be of interest (Stiller et al., 2014a, 2014b; Olensky et al., 2012).

## 6.2 Evaluating your User Interface

Evaluating the user interface and the offered information system with regard to usability has become one of the core areas of information system evaluation also in the multilingual cultural heritage domain. It can hardly be separated from evaluating the user interactions but it often focuses on the design, layout and informational structure of the website.

## 6.3 Evaluating your User Interactions

Many examples of studies evaluating user interactions exist but they often focus on a single isolated multilingual feature, for example image search (Vassilakaki et al., 2012).
The Conference and Labs of the Evaluation Forum (http://www.clef-initiative.eu/) focuses on multilingual and multimedia systems providing a framework for evaluation efforts particularly in the domain of information retrieval. The LogCLEF track was launched in 2009 with the aim to study user behavior in multilingual search systems through the analysis of activities and search queries. In 2009 and 2010, log files from different providers were evaluated intending to analyze

---

[10] http://www.europeanatech2015.eu/
[11] http://pro.europeana.eu/blogpost/we-want-good-quality-data-and-we-want-it-now

and classify user queries in order to understand search behavior in multilingual contexts and to improve search systems (Mandl et al., 2010,a 2010b).

From 2011 - 2013 the Cultural Heritage in CLEF (CHiC) lab Europeana data was used to identify and establish standardized evaluation procedures for multilingual cultural heritage information systems (Petras et al., 2013). Within CHiC, the iCLEF interactive task focused on user interactions and experience using Europeana data (Toms and Hall, 2013).

In 2014, the Interactive Social Book Search Tasks was introduced as part of the INEX lab at CLEF with the aim to investigate book search behavior with regard to metadata usage (Hall et al., 2014). Two different interfaces were provided, one basic interface and one multistage interface that focuses on browsing features with user-generated metadata such as ratings or reviews. Especially in open tasks such as casual leisure situations, browsing accesses are an important user experience aspect.

# 7. References

Agosti, M., Angelaki, G., Coppotelli, T., Di Nunzio, G. M. (2007). Analysing HTTP Logs of a European DL Initiative to Maximize Usage and Usability. In Goh, D. H.-L., Cao, T.H., Solvberg, I.T., Rasmussen, E. (Eds.), Proceedings of the 10th International Conference on Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers (Lecture Notes in Computer Science, Vol. 4822, pp. 35-44). Berlin, Heidelberg: Springer.

Agosti, M., Crivellari, F., Di Nunzio, G.M., Ioannidis, Y., Stamatogiannakis, E., Triantafyllidi, M.L., Vayanou, M. (2009). Report on Search Engines and HTTP Log Analysis - D5.2 (TELplus project). Retrieved from http://www.theeuropeanlibrary.org/portal/organisation/cooperation/telplus/documents/TELplus_ D5.2_15102009.pdf

Angelaki, G. (2007). Interim Report on Usability Developments in The European Library - M1.4 (EDLproject). Retrieved from http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/downloads /M1.4_Interim%20Report%20on%20Usage%20and%20Usability.pdf

Assets (2012).D2.2.1  Specifications of post querying processing functionalities. Available here: http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ASSETS/Deliverables/D2.2.1%20Specification%20of%20post%20querying%20processing%20functionalities.pdf

Aula, A., Kellar, M. (2009). Multilingual Search Strategies. In Olsen Jr., D. R., Arthur, R.B., Hinckley, K., Ringel Morris, M. Hudson, S.E., Greenberg, S. (Eds.), CHI '09 Extended Abstracts on Human Factors in Computing Systems (pp. 3865-3870). Boston, USA: ACM

Baierer, K., Dröge, E., Petras, V., & Trkulja, V. (2014, October). Linked data mapping cultures: an evaluation of metadata usage and distribution in a linked data environment. In *Proceedings of the 2014 International Conference on Dublin Core and Metadata Applications* (pp. 1-11). Dublin Core Metadata Initiative.

Marcia J. Bates. 2002. The cascade of interactions in the digital library interface. *Inf. Process. Manage.* 38, 3 (May 2002), 381-400. DOI=10.1016/S0306-4573(01)00041-3 http://dx.doi.org/10.1016/S0306-4573(01)00041-3

Bilal, D., Bachir I. (2007). Children's Interaction with Cross-Cultural and Multilingual Digital Libraries I: Understanding Interface Design Representations. Information Processing & Management, 43(1), 47-64.

Calhoun, K. S., & Patton, G. (2011). *WorldCat quality: an OCLC report.* OCLC.

Clark, D. J., Nicholas, D., Rowlands, I. (2011). Publishable Report on Best Practice and How Users are 151 Using the Europeana Service - D3.1.3 (EuropeanaConnect project). Retrieved from http://www.europeanaconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport_v1.0.pdf

Clough, P., Sanderson, M. (2006). User Experiments with the Eurovision Cross-Language Image Retrieval System.Journal of the American Society for Information Science and Technology, 57(5), 697-708.

Dangerfield, Kalshoven (eds.) (2015): Report and Recommendations from the Task Force on Metadata Quality. Available at: http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf

Dobreva, M., Chowdhury, S. (2010).A User-Centric Evaluation of The Europeana Digital Library. In Chowdhury, G., Khoo, C., Hunter, J. (Eds.), The Role of Digital Libraries in a Time of Global Change, 12th International Conference on Asia-Pacific Digital Libraries (Lecture Notes in Computer Science, Vol. 6102, pp. 148-157). Berlin, Heidelberg: Springer

Dragisic,Z., Eckert, K., Euzenat, J., Faria, D., Ferrara, A., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Kempf, A. O., Lambrix, P., Montanelli, S., Paulheim, H., Ritze, D., Shvaiko, P., Solimando, A., Trojahn, C., Zamazal, O. and Cuenca Grau, B. (2014) Results of the Ontology Alignment Evaluation Initiative 2014. In: *CEUR Workshop Proceedings OM 2014 : Proceedings of the 9th International Workshop on Ontology Matching co-located with the 13th International Semantic Web Conference (ISWC 2014) ; Riva del Garda, Trentino, Italy, October 20, 2014*; 61-104. RWTH, Aachen, 2014.

Eleta, I.; Golbeck, J.: A study of multilingual social tagging of art images: cultural bridges and diversity. CSCW '12 Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 2012

EuropeanaConnect (2011). D2.3.1 Multilingual mapping of schemes and vocabularies 2011. Final Report. Avalaible at: http://www.europeanaconnect.eu/documents/D2.3.1%20Multilingual%20mapping%20of%20schemes%20and%20vocabularies-Draft.pdf

EuropeanaConnect (2011).WP2.3 Multilingual Mapping of Controlled Vocabularies – Languages and Vocabularies for Selection. Available at: http://www.europeanaconnect.eu/documents/09_Multilingual_Mapping_of_Controlled_Vocabularies.pdf

Europeana Fashion (2013). D3.3(a) Enrichment Tool v1.0 Report. Available at: http://cordis.europa.eu/docs/projects/cnect/7/297167/080/deliverables/002-EuropeanaFashionDeliverable33aEnrichmentToolReportv1.pdf

EuropeanaPhotography (2013).D4.1 EuropeanaPhotography Vocabulary Definition. Available at: http://www.europeana-photography.eu/getFile.php?id=219

Europeana v2.0 (2013). D7.7 Midterm-report on innovative multilingual information access. Available at: http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/Readings/D7.7%20Midterm%20Report%20on%20Innovative%20Multilingual%20Information%20Access.pdf

Freire, N, 2013, 'Visualization and navigation of knowledge in pan-European resources: the case of The European Library' in proceedings of International UDC Seminar on Classification & Visualization: interfaces to knowledge.

Gäde, M. (2014). Country and language level differences in multilingual digital libraries; Dissertation, Humboldt-Universität zu Berlin, Philosophische Fakultät I. Thesis.

Gäde, M., Petras, V. (2014).Multilingual interface preferences.In Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14). ACM, New York, NY, USA, 231-234.

Ghorab, M. R., Leveling, J., Zhou, D., Jones, G.J.F., Wade, V. (2010). Identifying Common User Behaviour in Multilingual Search Logs. In Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G (Eds.), Multilingual Information Access Evaluation I. Text Retrieval Experiments. Proceedings of the 10th Cross-Language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments (Lecture Notes in Computer Science, Vol. 6241, pp. 518-525). Berlin, Heidelberg: Springer

Gonzalo, J., Penas, A., Verdejo, F., Peters, C. (2008). Workshop on Best Practices for the Development of Multilingual Information Access Systems: The User Perspective – D3.2 (TrebleCLEF project). Retrieved from http://www.trebleclef.eu/getfile.php?id=249

Hall, M. M., Huurdemann, H., Koolen, M., Skov, M., & Walsh, D. (2014).Overview of the INEX 2014 interactive social book search track.

Manguinhas, H (editor, 2014): Semantic Enrichment Framework. Available at: https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y

Isaac, A., Wang, S., Van der Meij, L., Schlobach, S., Zinn, C., & Matthezing, H. (2009). Evaluating thesaurus alignments for semantic interoperability in the library domain.*IEEE Intelligent Systems,24*(2), 76-86.

ISO 639-2:1998, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

IRN Research. (2011). EUROPEANA - Online Visitor Survey Research Report Version 3. Retrieved from http://pro.europeana.eu/c/document_library/get_file?uuid=334beac7-7fc2-4a4e-ba23- 4dcc1450382d&groupId=10602

Keegan, T. T., Cunningham, S.J. (2005). Language Preference in a Bi-Language Digital Library.In Marlino, M., Sumner, T., Shipman, F. (Eds.), Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (pp. 174-175). New York, NY: ACM

Király, Péter: Query Translation in Europeana. Code4Lib Journal. Issue 27. AVailable here: http://journal.code4lib.org/articles/10285

Lamm, K., Mandl, T., Koelle, R. (2010). Search Path Visualization and Session Performance Evaluation with Log Files. In Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G. (Eds.), Multilingual Information Access Evaluation I. Text Retrieval Experiments. Proceedings of the 10th Cross-Language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments (Lecture Notes in Computer Science, Vol. 6241, pp. 538-543). Berlin, Heidelberg: Springer

Landry, Patrice (2009). Multilingualism and subject heading languages: how the MACS project will be providing multilingual subject access in Europe. *Catalogue & Index* 157 p. 9-11.

Leveling, J., Ghorab, R., Magdy, W., Jones, G. J. F., Wade, V. (2010). DCU-TCD@LogCLEF 2010: ReRanking Document Collections and Query Performance Estimation. In Braschler, M., Harman, D., Pianta, E. (Eds.), CLEF 2010 LABs and Workshops, Notebook Papers. Padua, Italy. Retrieved from http://ims-sites.dei.unipd.it/documents/71612/86374/CLEF2010wn-LogCLEFLevelingEt2010.pdf.

Mandl, T., Agosti, M., Di Nunzio, G.M., Yeh, A., Mani, I., Doran, C., Schulz, J.M. (2010). LogCLEF 2009: The CLEF 2009 Multilingual Logfile Analysis Track Overview. In Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda G. (Eds.), Multilingual Information Access Evaluation I. Text Retrieval Experiments. Proceedings of the 10th Cross-Language Evaluation Forum Conference on Multilingual Information Access Evaluation: Text Retrieval Experiments (Lecture Notes in Computer Science, Vol. 6241, pp. 508-517). Berlin, Heidelberg: Springer.

Mandl, T., Di Nunzio, G.M., Schulz, J.M. (2010a). LogCLEF 2010: The CLEF 2010 Multilingual Logfile Analysis Track Overview. In Braschler, M., Harman, D., Pianta, E. (Eds.), CLEF 2010 LABs and Workshops, Notebook Papers (pp. 22-23). Padua, Italy. Retrieved from http://imssites.dei.unipd.it/documents/71612/86374/CLEF2010wn-LogCLEF-MandlEt2010.pdf.

Marlow, J., Clough, P.D., Recuero, J.C., Artiles, J. (2008). Exploring the Effects of Language Skills on Multilingual Web Search. In Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (Eds.), Advances in Information Retrieval. Proceedings of 30th European Conference on IR Research (ECIR 2008) (Lecture Notes in Computer Science, pp. 126-137). Berlin, Heidelberg: Springer

Mayr, Philipp, and Vivien Petras. 2009. "Cross-concordances: terminology mapping and its effectiveness for information retrieval." *International Cataloguing and Bibliographic Control* vol. 38 , no. 3 43-52.

Minelli, S., Del Secco, I., Naldi, G. . (2006). User Requirements Analysis - D 1.2 (Multimatch project). Retrieved from http://www.multimatch.org/docs/publicdels/D1.2Final.pdf

Oakes, M., Xu, Y. (2009). A Search Engine based on Query Logs and Search Log Analysis at the University of Sunderland. In Peters, C. (Ed.), Results of the CLEF 2009 Cross-Language System Evaluation Campaign. Corfu, Greece. Retrieved from http://imssites.dei.unipd.it/documents/71612/85150/CLEF2009wn-LogCLEF-OakesEt2009.pdf.

Toms, E. G., Hall, M.M. (2013).The CHiC Interactive Task (CHiCi) at CLEF2013.In Forner, P., Navigli, R., Tufis, D. (Eds.), CLEF 2013 Evaluation Labs and Workshop, Online Working Notes.Valencia, Spain. Retrieved from http://imssites.dei.unipd.it/documents/71612/430938/CLEF2013wn-CHiC-TomsEt2013.pdf.

Oard, D. W., Gonzalo, J., Sanderson, M., López-Ostenero, F., Wang, J. (2004).Interactive Cross-Language Document Selection.Information Retrieval, 7(1-2), 205-228.

Olensky, M., Stiller, J., & Dröge, E. (2012). Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy. In *Metadata and Semantics Research* (pp. 252-263). Springer Berlin Heidelberg.

PartagePlus (2012) D3.1  - Multilingual terminology Of Art Nouveau 'Micro' object types. Available at: http://www.partage-plus.eu/en/servefile?id=18

Peinado, V., Artiles, J., Gonzalo, J., Barker, E., López-Ostenero, F. (2008). FlickLing: A Multilingual Search Interface for Flickr. In Peters, C. (Eds.), Results of the CLEF 2008 Cross-Language System Evaluation Campaign. Aarhus, Denmark. Retrieved from http://imssites.dei.unipd.it/documents/71612/86371/CLEF2008wn-iCLEF-PeinadoEt2008a.pd.

Petras, V., Bogers, T., Toms, E., Hall, M., Savoy, J., Malak, P., ... & Masiero, I. (2013).Cultural heritage in clef (chic) 2013.In *Information Access Evaluation.Multilinguality, Multimodality, and Visualization* (pp. 192-211).Springer Berlin Heidelberg.

V. Petras, J. Stiller, and M. Gäde (2013). Building for Success (?) - Evaluating Digital Libraries in the Cultural Heritage Domain. In: Recent Developments in the Design, Construction and Evaluation of Digital Libraries: Case Studies, ed. by Cool, C., NG, K. B., ICI Global.

Petrelli, D., Beaulieu, M., Sanderson, M. (2002).User Requirement Elicitation for Cross-Language Information Retrieval.The New Review of Information Behaviour Research, 2, 17-35.

Srinivasarao, V. (2008). Mining the Behavior of Users in a Multilingual Information Access Task. In Peters, C. (Eds.), Results of the CLEF 2008 Cross-Language System Evaluation Campaign. Aarhus, Denmark. Retrieved from http://imssites.dei.unipd.it/documents/71612/86371/CLEF2008wn-iCLEF-Vundavalli2008.pdf.

Stiller,J., Gäde, M., Petras, V. (2011). Is tagging multilingual?: a case study with BibSonomy. In Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries (JCDL '11). ACM, New York, NY, USA, 421-422.

Stiller, J., Olensky, M., & Petras, V. (2014a).A Framework for the Evaluation of Automatic Metadata Enrichments. In *Metadata and Semantics Research* (pp. 238-249). Springer International Publishing.

Stiller, J., Petras, V., Gäde, M., & Isaac, A. (2014b). Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection* (pp. 238-247). Springer International Publishing.

Stiller, Isaac & Petras (eds.) (2014): EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report. Available: http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/MultilingualSemanticEnrichment//Multilingual%20Semantic%20Enrichment%20report.pdf

Steichen, B., and Freund, L.(2015). Supporting the Modern Polyglot: A Comparison of Multilingual Search Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 3483-3492. DOI=10.1145/2702123.2702541 http://doi.acm.org/10.1145/2702123.2702541

Vassilakaki, E., Hartley, R. J., & Johnson, F. (2012). Image seeking in multilingual environments: a study of the user experience. *Information research*, *17*(4).

Wu, D., He, D., Lou, B. (2012). Multilingual Needs and Expectations in Digital Libraries: A Survey of Academic Users with Different Languages. The Electronic Library, 30(2), 182-197.