White Paper on Best Practices for Multilingual Access to Digital Libraries

23/06/2016

Editor
- Juliane Stiller

Contributors (alphabetical order)
- Anila Angjeli
- Emmanuelle Bermès
- Dean Birkett
- Maarten Brinkerink
- Valentine Charles
- Mariana Damova
- Cécile Devarenne
- Maria Gäde
- Sergiu Gordea
- David Haskiya
- Timothy Hill
- Antoine Isaac
- Lukas Koster
- Hugo Manguinhas
- Gregory Markus
- Mark A. Matienzo
- Vivien Petras

Executive Summary

This White Paper aggregates resources and best practices for realizing multilingual access to cultural heritage content in digital libraries. It offers recommendations and resources for overcoming challenges in letting users access content they might not understand. It addresses cultural heritage professionals and wants to give practical advice for common problems. It shows examples from the cultural heritage domain, lists research that addresses the implementation of multilingual access and gathers best practices from various projects within the domain.

The paper focuses on three different components of multilingual cultural heritage information systems – the data, the interactions and the interface. Additionally, common pitfalls are addressed and suggestions for evaluating multilingual components are gathered.

It is an extended version of a deliverable published within the EU-funded project Europeana v3.0. Many cultural heritage professionals from the Europeana Network followed the open call to contribute their expertise to this work, resulting in a rich resource reflecting multiple perspectives.

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 coming from museums, libraries and archives. Often, as is the case for Europeana, these objects are or are represented in different languages, most of which a 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 specific 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 particular language.

The metadata language can correspond with the language of the objects (if they have a language like full texts of monographs, for example, or audio-visual material like films) but it does not have to. The object language corresponds to the language of the text for textual-based objects, but other content types have either no language - such as paintings or instrumental music - or might have several languages. Catering for these differences is a challenge. 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 process is one important interaction users will perform in the system but many more actually occur.

Another major interaction is the navigation through the information system, which is also a language-dependent activity. The graphical user interface is the tool users interact with. 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 and its potential localisation. If it fails at this point, the other levels of the information system cannot be explored (Bates, 2012). Figure 1 shows the four different layers of a typical information system, namely the user interface, the user interactions, the access system and the underlying data (metadata and / or objects). All these levels have a multilingual perspective. Diekema (2012) lists the following areas in which challenges for the development of multilingual digital libraries arise: translation, language processing, user interface, system architecture and user

project management. Although this White Paper presents a different structure, all these areas will also be covered.
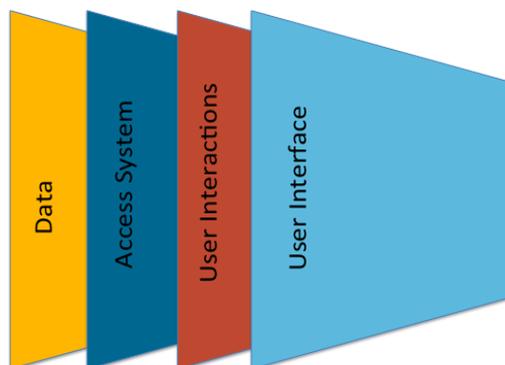
Figure 1: The different layers of accessing information in a cultural heritage information system.

1.2.   Structure and Objective of the Paper

This White Paper aggregates best practices and common procedures on how digital libraries present multilingual cultural heritage content and how it can be effectively offered to users. Many examples are taken from Europeana but the recommendations target digital libraries, archives and museums in general.

The White Paper is structured in four major parts which relate 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), and the functionalities user interact with (chapter 4). Chapter 5 highlights specific problems and challenges in multilingual systems and chapter 6 discusses the evaluation of these components.

2. Making your Data Multilingual

Multilingual access to content does not only mean offering objects with textual content in several languages but also providing their describing metadata in ways that support multilingual access. 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. 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. 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 Tags

A first step towards 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 metadata text and supports re-use and processing of the metadata in a multilingual environment. If the metadata is available in several languages, a language tag helps to display the most suitable language to the user. Language attributes also support the translation of the text to other languages for searching or browsing. Language identification of metadata is crucial for mapping metadata elements to multilingual vocabularies, named entity recognition and data mining with natural language processing (NLP), which eventually increases the searchable data for an object.

Study Findings & Further Reading:

Findings and Summaries	Source
Europeana suggests adding 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
Language-related elements of the HOPE project[1] data model.	http://hopewiki.socialhistoryportal.org/index.php/Semantic_Model:_Supporting_Multilingual_Description

Best Practices:
- Language tags can be added to the metadata or to individual text string values within the metadata. In XML and RDF metadata, the use of the xml:lang attribute to

---

[1] http://www.peoplesheritage.eu/

specify the language of the specific metadata value is suggested. 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 corresponding language tag attached. Language-aware 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 sufficient textual, language-dependent description is available for accurate language 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[2].

2.2.   Using Multilingual Vocabularies

The metadata for objects can be made multilingual by using multilingual vocabularies for description. There are multilingual vocabularies (i.e. controlled vocabularies for subject description or name authorities for entities) which use identifiers for concepts with different labels for multiple languages, which can be used to cross the language barrier. One might also want to create multilingual vocabularies to provide multilingual access to specialized collections. This can be achieved by mapping (monolingual) vocabularies available in different languages or by translation of (monolingual) vocabularies to 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 multilingual controlled 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 and technical approaches 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,
Lists 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 Europeana Sound D1.3
Models for mapping and guidelines for mapping can be found in the ISO 25964 - the international standard for thesauri and interoperability with other vocabularies.	http://www.niso.org/schemas/iso25964/

Best Practices:
- Use controlled vocabularies or authority lists that have multilingual elements and ideally use identifiers for concepts instead of textual strings.
- Use open vocabularies such as VIAF[3] for names, EuroVoc[4], AAT[5], or MACS (mapped LCSH/Rameau/SWD) (Landry, 2009) for subject headings.
- Browse the FLOSS[6] inventory for finding suitable mapping and matching tools and software.
- In general it seems easier/more productive to map from vocabularies in a few languages with specialized coverage to more general vocabularies that cover more languages.
- ISO 25964[7] provides models for mapping.

---

[3] https://viaf.org/

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

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

[6] 300 Free, Libre Open Source Software relevant for the cultural heritage domain:
http://pro.europeana.eu/page/floss-inventory

[7] ISO (2013). ISO 25964-2 – the international standard for thesauri and interoperability with other vocabularies. Interoperability with other vocabularies

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.[8]

2.3.   Multilingual Semantic Metadata Enrichment

If the metadata associated with cultural heritage objects is monolingual, multilinguality can be added by linking your metadata elements to (multilingual) vocabularies and authority files. Multilingual semantic enrichments add links to equivalent or semantically related (e.g. broader or narrower) resources to the metadata, their descriptions being available in different languages. The newly created links can be further exploited and semantically related keywords and/or translations will be used to improve retrieval.

Motivation:
Besides supporting the multilingual dimension, linking or adding additional terms to the metadata also helps to contextualise the cultural heritage objects and makes them easier to retrieve.

---

[8] 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.	Isaac et al. (eds), 2015; Stiller, Isaac & Petras (eds.) 2014; Olensky et al., 2012; Stiller, Olensky & Petras, 2014; Stiller et al, 2014)
A list of vocabularies, which Europeana can dereference, is provided.	https://docs.google.com/spreadsheets/d/1BoDNolkcp_qfvVShdOZyGcf61XslcwKF2MdGcjgYs20
A detailed description of the Europeana enrichment process has been provided.	Stiller, Isaac & Petras (eds.) 2014, App.: 1 http://pro.europeana.eu/page/europeana-semantic-enrichment

Best Practices:
- Rather than adding strings to the metadata, the enrichment process should be designed in a way that URIs or identifiers are used to link to the multilingual vocabulary.
- If providing URIs or other identifiers is not possible, then enrich metadata with multilingual vocabularies by linking to vocabulary terms. Use mapping techniques to create multilingual vocabularies.
- Establish an enrichment process and a sustainable enrichment strategy for continuous updates (from vocabularies and metadata alike).
- 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[9] (accessed September 2015).

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 the 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) are inherently multilingual (or non-lingual). 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 also have implications for a localised and language-aware search result representation.

Motivation:

When users identify a relevant object satisfying their information need, they should be able to understand the description of the given object even if it is not in their preferred or native language.

---

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

Study Findings & Further Reading:

Findings and Summaries	Source
Metadata translation is often sufficient to support a user's assessment whether an object is relevant or not.	Oard et al., 2004; Gonzalo et al., 2008; D1.2 MultiMatch, 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
Multilingual subtitles in audio or video files added using editorial software such as Amara.	http://www.amara.org/
For indexing metadata, transliteration needs also to be taken into account.	D2.2 HOPE Project., 2011
How to build your own machine translation system for digital collections	Chen, 2016

Best Practices:
- The metadata should be displayed in a language the user understands.
- Automatic translations at the object level can be offered using external translation services.
- Machine translations should be used with care especially for highly specialized and curated content.
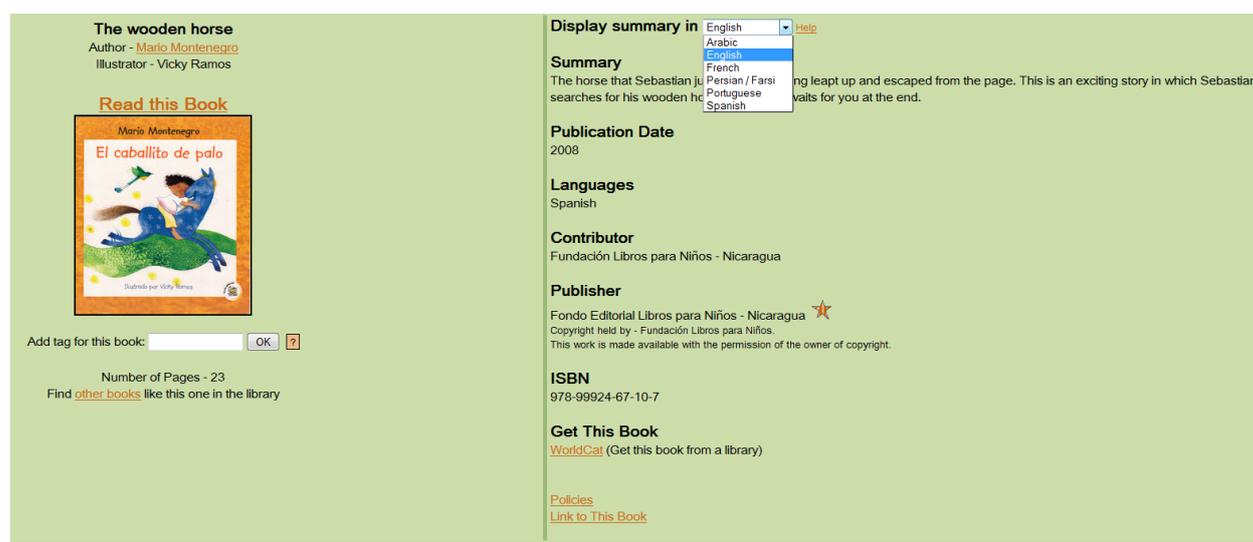

Figure 5: Example of a Spanish book in the International Children´s Digital Library[10] with volunteer translations in 5 additional languages (see dropdown menu).

---

[10] http://en.childrenslibrary.org/

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. Going beyond a mere translation of words[11] involves an adaptation to culture-specific requirements and can go as far as displaying different items to different language groups - a practice often referred to as 'localisation'.

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. Static pages and content are help sections, contact site, navigational items and 'terms of use' and 'about us'-pages, for example. Because these do not change often, the translation 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. If resources are limited, an indication of a minimum of desired translations could be achieved through usage statistics. One should also keep in mind that users of cultural heritage information systems are scattered around the world and one might also want to cater for right-to-left languages such as Arabic. The most common case for translating websites and static content is to hire a professional translator. To find a suitable candidate, one can look in the directories of professional associations for translators - most of the European countries have one. When it comes to selecting an appropriate translator, it is important to ensure to find translators who translate into their mother tongue, i.e. the target language for the translation and the translator's native language should match. That being said, it is also important to have translations of interfaces proofread by a native speaker if a professional translator is not an option. There is also an International Standard, which provides requirements for translation services. The ISO 17100:2015 standard[12] covers all aspects of the translation process.

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

---

[11] https://en.wikipedia.org/wiki/Language_localisation
[12] http://www.iso.org/iso/catalogue_detail.htm?csnumber=59149

Study Findings & Further Reading:

Findings and Summaries	Source
The most frequently used interface language is English.	Gäde, 2014; M1.4 EDL-project, 2007; Agosti et al., 2007; D3.1.3 EuropeanaConnect; Oakes and Yu., 2009; Keegan and Cunningham, 2005
The default interface language influences user behavior and affects the number of users.	Keegan and Cunningham, 2008
Tips for responsive web design for multilingual websites can be listed.	http://responsivenews.co.uk/post/1231045 12468/13-tips-for-making-responsive-web-design
Google provides advice on bidirectionality.	https://www.google.com/design/spec/usabi lity/bidirectionality.html
The W3C provides tips on Internationalization.	http://www.w3.org/International/quicktips/
Google offers several recommendations on languages and language-related SEO.	https://support.google.com/webmasters/to pic/2370587?hl=en&ref_topic=4598733

Best Practices:
- Make sure that the different language versions are accessible through search engines, e.g. sending English users to the English version of your site.
- Follow design convention for menu items and navigation to make access to your system easier, e.g., consistent labeling.
- Make sure that the user can switch between the different language versions you are offering at any point of an interaction.

Example:
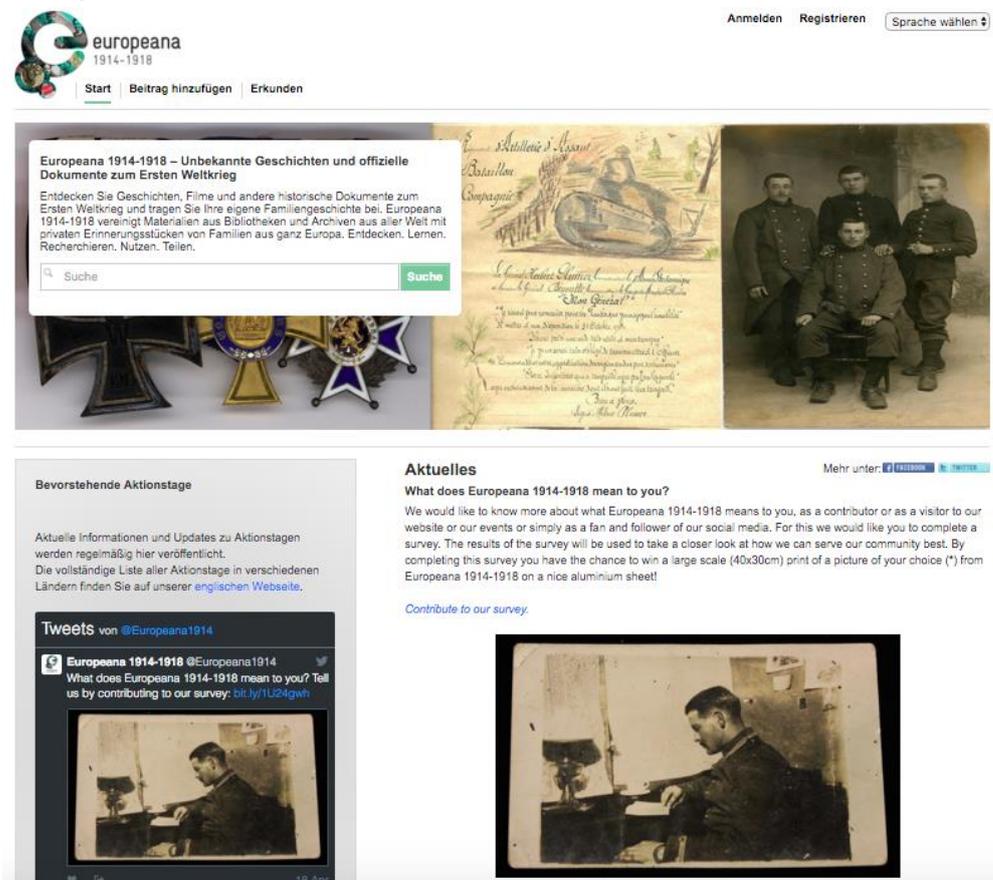
Figure 6: Homepage of Europeana1914-1918 in German.

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 (assuming that the quality of the translation content is good).	Agosti et al., 2007; D5.2 TELplus, 2009; Dobreva & Chowdhury, 2010; Gäde & Petras, 2014

Best Practices:
- Use automatic detection of the user language if at all possible: language of the browser or operating system should be preferred over automatic detection via IP-address.
- If users indicate a language preference, a cookie should be set and the preference should be stored for future interactions as a user profile.
- 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 or predefined in the user settings.

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
Using flags instead of language names is a common practice although flags do not represent languages and therefore lead to confusion.	http://ux.stackexchange.com/questions/2472/how-to-graphically-represent-a-language
Flags instead of language names 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; D5.2 TELplus, 2009; Dobreva & Chowdhury, 2010; D3.1.3 EuropeananConnect, 2011; Oakes et al., 2009; Keegan & 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.

- Avoid language mixes, which might occur when static content is translated but dynamic content stays in the original language (see section 5.1).

Example:

Figure 7: Language drop-down menu of Europeana showing languages in the default English language, which is not recommended practice (this drop-down menu is accessible on a settings page one click away from the homepage).

3.4.   Multilingual Result Display

If search across languages is enabled, the question how to present results in different languages to the user needs to be answered. The options range from a single merged search result page, which integrates all different language results, to lists of results separated by language.

Motivation:

If users are presented with different languages in their search result list, purposeful interface design might help to identify relevant results.

Study Findings & Further Reading:

Findings and Summaries	Source
Users prefer the panel interface where results separated by language are displayed in equal sized panels while the merged result list was least preferred.	Steichen & Freund, 2015
Clear separation of languages in result lists is preferred.	Steichen & Freund, 2015

Best Practices:
- Avoid a merged list of results in different languages.
- Offer the user visual clues, separation by language or highlighting of languages, to enable them to quickly identify results they can understand.

4. Making your User Interactions support Multilinguality

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. If query completion or query suggestion features are language-aware, users can be served more appropriately, but it can also be offered independent of language.

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, avoid spelling mistakes and help disambiguate terms.

Study Findings & Further Reading:

Findings and Summaries	Source
Dynamic query suggestions and auto-completions are becoming standard in search engine interfaces.	Hearst, 2009, chapter 4
Query suggestion and recommendation services help users in finding what they are looking for.	D2.2.1 Assets, 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 8: Query suggestion and auto-complete functionality in Smithsonian Collections.[13]

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 identification of 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, Beaulieu & Sanderson, 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, Gäde & Petras, 2013
Important aspects regarding the	Peters, Braschler & Clough, 2012

---

[13] http://collections.si.edu/search/

implementation of query formulation and translation to multilingual information systems can be enumerated.	
Europeana API supports a query translation process using parallel language Wikipedia versions.	Király, 2015; http://labs.europeana.eu/api/query-translation

Best Practices:
- Offer automatic query translation with limited suggestions or variants for when the translation process fails.
- Offer users the ability to turn on/off the query translation feature according to their needs.
- 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.

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 in 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 overly complex interaction steps.

Motivation:
User-assisted translations do not only support the system in adding domain-specific translations to their dictionary, but also provide 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.	D5.2 TELplus, 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, 2008; Aula and Kellar, 2009; D7.4 Cacao Project, 2009; Ghorab et al., 2009; Leveling et al., 2010; Marlow et al., 2008

Best Practices:
- The translation workflow should not require too much effort from the user's side. Required clicks need to be minimized.
- Options for users to edit the translated query should be easily visible in the system (more on user-generated content which can help to improve multilingual access can be found in section 4.8).
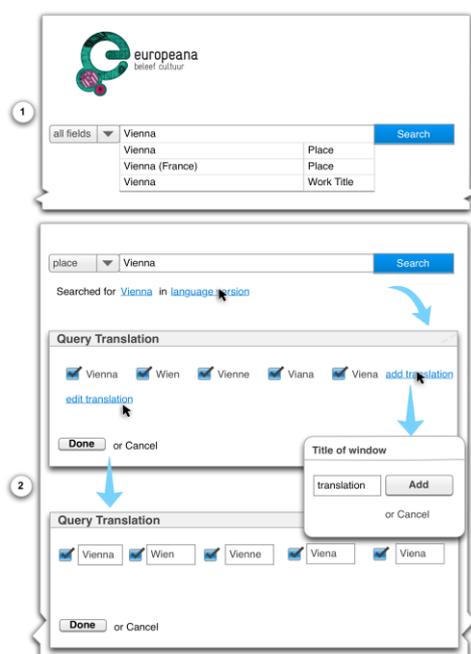
Example:

Figure 9: 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. Browsing provides a possibly language-independent interaction, when visual representations of objects or metadata options are used - another way to cross the language barrier.

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, Mandl & Koelle, 2009; Keegan and Cunningham, 2005; Wu, He, Luo, 2012
Assessing cultural collections can benefit from the principles of rich-prospect browsing, which displays a visual representation of every object in the collection.	Ruecker, Radzikowska &Sinclair, 2011

Best Practices:
- When providing textual browsing access, make sure to provide multilingual options, such as multilingual user interfaces and vocabularies for browsing.
- If possible, also provide language-independent features (e.g. visual cues) for browsing.

Example:

Colours - Europeana Collections

Search for images based on their colour. Simply click a coloured box below to discover images containing that colour!

Dark Slate Gray #2F4F4F	Dim Gray #696969	Dark Gray #A9A9A9	Black #000000	Web Gray #808080
Silver #C0C0C0	Gainsboro #DCDCDC	Light Gray #D3D3D3	White Smoke #F5F5F5	Tan #D2B48C
Maroon #800000	White #FFFFFF	Dark Olive Green #556B2F	Lavender #E6E6FA	Dark Khaki #BDB76B
Rosy Brown #BC8F8F	Beige #F5F5DC	Wheat #F5DEB3	Pale Goldenrod #EEE8AA	Snow #FFFAFA

Figure 10: Browsing by colour functionality offered by Europeana.

4.5.   Search and Browse Result Filtering

When reaching a certain scale, all digital libraries need to provide filter or drill-down options for search results, so that the number of displayed objects can be reduced to a manageable amount. These filter and browsing options should be presented in desired languages enabling the users to filter for subjects or material type, for example. A specific case of filtering would be the option to filter result sets by language.

Motivation:
Options for filtering are a natural way for users to reduce their result set in a sensible way. A language filter allows users to filter out results in languages they do not understand.

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 options for users to display browsing and filter mechanisms in their desired languages.
- Provide facets, which let users refine results by original language of the object in case of textual or audiovisual objects.
- Make clear, what is meant by the language filter - the language of the objects or the language of the metadata.
- Provide language refinement options in advanced search interface as well as a facet.

Example:

LANGUAGE

- Deutsch (1,061,374)
- Multilingual (31,688)
- Polski (18,128)
- Magyar (6,706)
- Nederlands (6,260)

Figure 11: Europeana's language filter. Note that it is not clear if the filter targets the object or 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 options 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
Overcoming the search paradigm and offering visual access to cultural content can provide better access control to users.	Glinka, Meier & Dörk, 2015

Best Practices:
- Probably the most common non-text-based way to present information is through its spatial or temporal aspects. One thing to remember is that spatial as well as temporal 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.
- Spatial and temporal browsing offers new access points for cultural collections but requires targeted data curation and design.
- Browsing by visual aspects such as the colour of objects is also an option for language independent access, see section 4.4.
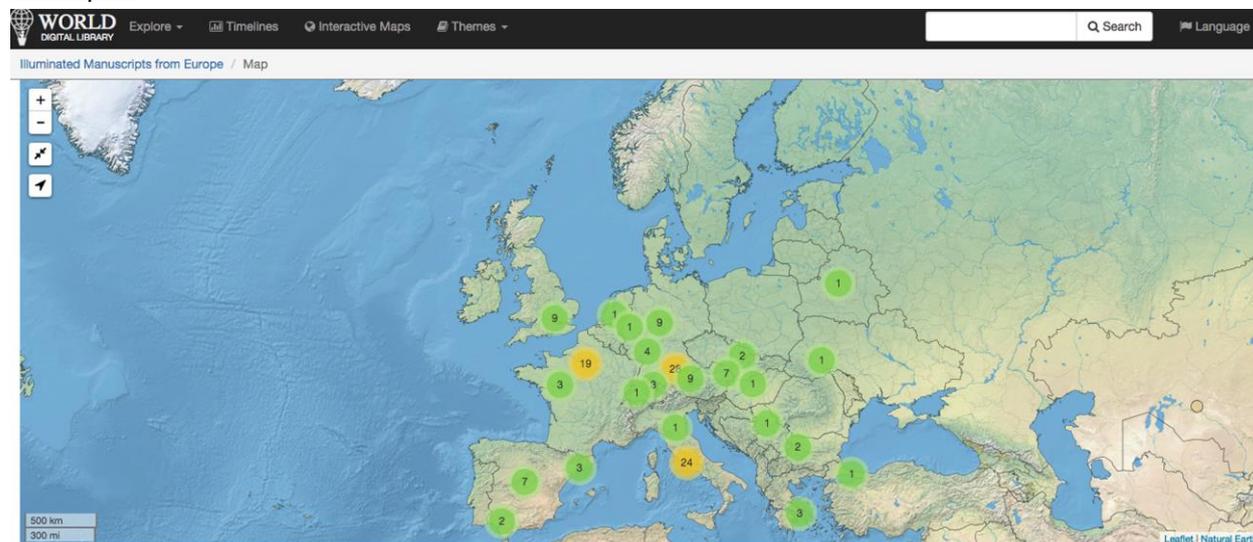
Example:

Figure 12: Map Browsing of the World Digital Library[14].

---

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

4.7.   Site Structure and Search Engine Landing Pages

You do not only want users to find the version of your site in their preferred language but also design your site to be search engine friendly.

Motivation: Making sure web search engines know which languages you are supporting helps in serving users by returning the page version with appropriate language in the search results.

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: https://support.google.com/webmasters/answer/182192?hl=en#1 Bing: http://www.bing.com/webmaster/help/geo-targeting-your-website-b7629197

Best Practices:

- Offer pages and object landing pages in different language versions.
- For users coming from search engines to a landing page they do not understand the language switch should be easy to find and identifiable.

Example:

Europeana 1914-1918 - untold stories & official histories of ...
www.europeana1914-1918.eu/ ▾ Diese Seite übersetzen
Historic lost footage of Chicago Disaster found by chance on Europeana. A phD candidate from the University of Illinois in Chicago has uncovered previously ...

Europeana 1914-1918 - Ihre Familiengeschichte zum ...
www.europeana1914-1918.eu/de ▾
Europeana 1914-1918 – Unbekannte Geschichten und offizielle Dokumente zum Ersten Weltkrieg. Entdecken Sie Geschichten, Filme und andere historische ...

Europeana 1914-1918 - skrite zgodbe & uradna zgodovina ...
www.europeana1914-1918.eu/sl ▾ Diese Seite übersetzen
Europeana 1914-1918 – skrite zgodbe & uradna zgodovina prve svetovne vojne. Raziskujte zgodbe, filme in zgodovinsko gradivo o prvi svetovni vojni ...

Europeana 1914-1918 - untold stories & official histories of ...
www.europeana1914-1918.eu/hu ▾ Diese Seite übersetzen
Europeana 1914-1918 - Az első világháború elmondatlan és hivatalos történetei. Fedezd föl az első világháború történeteit, filmjeit és más történeti anyagait, ...

Europeana 1914-1918 - histoires inédites et histoires ...
www.europeana1914-1918.eu/fr ▾ Diese Seite übersetzen
Europeana 1914-1918 – histoires inédites et histoires officielles de la Première Guerre mondiale. Découvrez des histoires, des films et des documents ...

Figure 13: Different language version of Europeana 1914-1918[15] in Google search results.

---

[15] www.europeana1914-1918.eu/

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. This added content can be very helpful supporting access to the objects, but multilinguality needs to be considered when adding content, which is potentially multilingual.

Motivation:

Besides adding descriptions to content, users might help in translating objects or adding tags in multiple languages to existing metadata, thereby alleviating multilingual challenges. 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, Gäde & Petras, 2011
For multilingual user tagging, cultural perspectives are expressed in less frequent tags.	Eleta & Golbeck, 2012
The Language Quality Game by Microsoft lets users find issues in interface translations.	http://social.technet.microsoft.com/wiki/contents/articles/9301.language-quality-game-player-instructions.aspx

Best Practices:
- Design interfaces, which let users determine the language of their contribution.
- Keep in mind that users' cultural diversity is also reflected in their contributions.
- Clearly mark user-suggested translations as such and differentiate them from already curated/approved translations.
- Decrease the number of steps a user must take to contribute annotations or translations.
- Help the user to choose a term via autocompletion based on vocabularies - this minimizes the effort for the user and provides unambiguous or semantic tags.
- Gamifying the experience can motivate users to contribute more and better content.

Example:

Figure 14: 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.

5.1.   Levels of Multilinguality of Cultural Heritage Objects

There are several levels of multilinguality in digital libraries which need to be taken into account when trying to homogenize the language of the interface - there is the language of the object and the language of the metadata-providing institution. Both might result in the presentation of several languages in the metadata. Automatic translation of metadata elements might be good for accessing the content, but could obfuscate the original language of the object, which still needs to be translated for certain use cases. The metadata should enable users to infer the language of the object in a clear manner. Translations for titles, for example, can be offered to the users but should not be automatically imposed.

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 (An example is figure 15 which shows a Europeana result page with static elements in German and metadata in Dutch). When thinking about a translation of the content and structure of the site, providers should consider all the different elements on the pages which could be translated. This could be applicable to:
- menu items,
- automatically pulled content coming from blogs, news feeds, etc. also dynamic content,
- metadata of objects,
- query suggestions and auto-completion features.

It is not feasible to present the user with a site where all elements are always translated to the user's preferred languages. This would also defeat the digital libraries' aim of presenting users with the wealth of cultural heritage, which is naturally embedded in language use, e.g. some users want the objects in its original language. Translating all metadata makes these objects hard to identify. There is also some literature on the problems which arise with translations in the cultural heritage sectors. Guillot, 2014, addresses the translations of text in the museum context as cultural representations.

Nantel and Glaser (2008) argue that a perfect translation of the website is not enough and users may still find the website hard to use if it is not adapted to their specific culture. In their study, they showed that users' perceived usability of a website, depends on the quality of the translation and increases if the linguistic background of the website's designer corresponds to the one of the users. Similar results were obtained by Vyncke

and Brengmann (2010) who reviewed a decade of studies on the impact of localisation on the effectiveness of websites.

Daniel et al. (2011) offer advice on the use of design elements, such as language, colour, metaphors and page layout, for different cultures. Bilal et al. (2006, 2007) conducted user tests with children for the different icons used in the international Children's Digital Library.
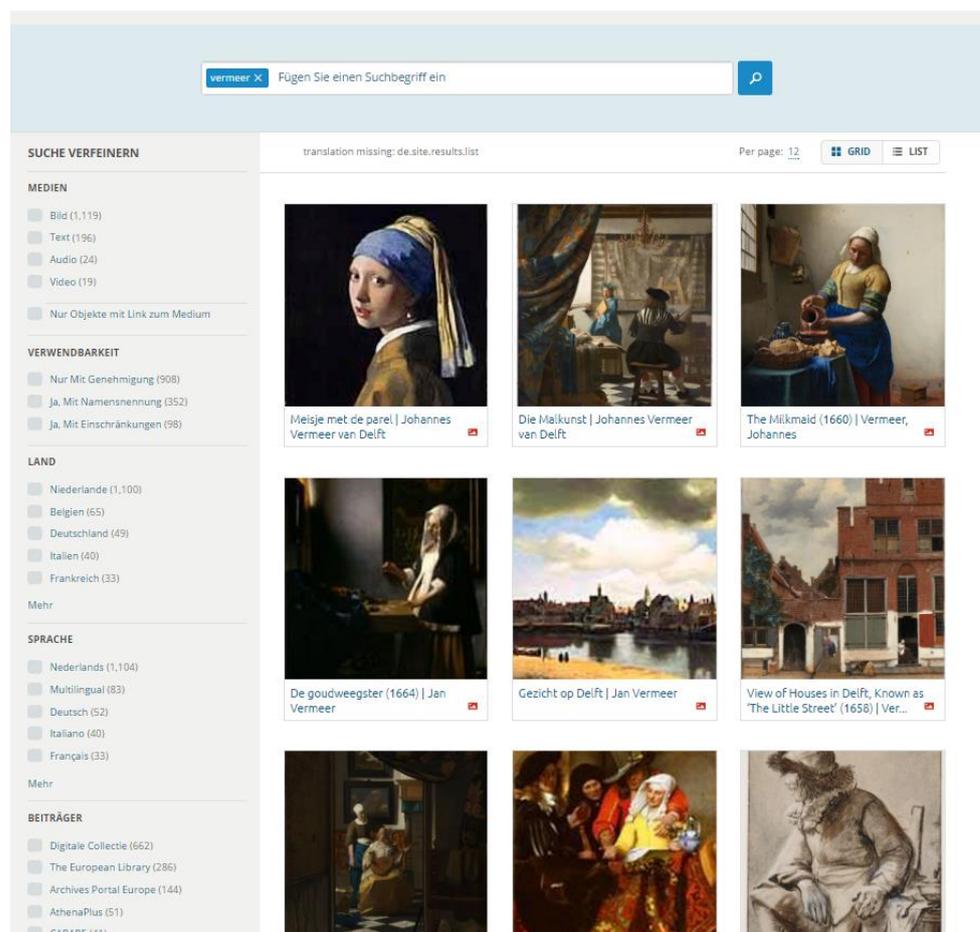
Figure 15: Europeana search result page in German with objects catalogued in Dutch.

Users often do not understand institutional conventions regarding the language of objects and the language of the 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. It is therefore not advised to use this as signal for the language of the content or the cataloguing language.

Especially if this information is used in filtering results, it should be made very clear what it refers to.

5.2.   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.

5.3.   Providing Sustainable Multilinguality

Offering multilingual solutions, which bridge the language gap in an information system, is not a decision made only 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, 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 Translations

To evaluate automatic translations, several metrics can be used. The general approach is the comparison to a humanly created benchmark of translations. Commonly, two different metrics are used to automatically compare the translation with the benchmark: (1) BLEU's main feature is the comparisons of n-grams in the machine translation and the human translation (Papineni et al., 2002) and (2) NIST is similar to BLEU with a few adaptations (Doddington et al., 2002). Further research has developed both scores to correlate even more with human judgement (e.g. Lavie, Sagae &Jayaraman, 2004). For an overview on the common practices for translation quality assessment, please refer to (House, 2014). Recently, the Multidimensional Quality Metrics were developed. It is a framework developed by the EU-funded project QTLaunchPad[16] for assessing the quality of translations. It is adaptable to specific tasks and assesses quality on several dimensions based on the requirements of a given project[17].

Another approach are the human Machine Translation evaluation metrics: Fluency and Adequacy (LDC, 2005). Adequacy is measured by comparing translations to a gold standard assessing the degree of information which is represented in the translated version compared to the original one. The assessment is done with an ordinal 5-point scale. Similarly, fluency is assessed with the difference that no gold standard is used for the evaluation but the standard grammar rules of the given language. Additional metrics can be the count of the words which were not translated as well as the incorrectly translated words (Chen et al., 2012). Not many studies exist which translate metadata records or records from the cultural heritage domain. The preliminary study by Chen et al. (2012) evaluated the performance of current machine translation systems translating metadata records from the Portal to Texas History. In another paper, she reports on a research project to improve performance of automatic translation services for metadata records (Chen et al., 2014), where she also applied human MT evaluation metrics, namely, fluency and adequacy. Matusiak et al. (2015) present a case study where bilingual metadata records were created combining human translations and the mapping of vocabularies. In their discussion they point out that the mere machine translation might not be sufficient to transfer cultural objects and their meaning into different languages. They advocate for human translation and professional indexing.

---

[16] http://www.qt21.eu/

[17] http://www.qt21.eu/launchpad/content/multidimensional-quality-metrics

6.2. 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[18] and the EuropeanaTech Meeting 2015[19] dedicated a whole session with round table discussion to this issue[20]. 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[21]) 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 highlight the fact 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).

6.3. 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. Methods and processes are not much different from the ones used for usability and user interaction evaluation (e.g. Lazar et al., 2010).

In the cultural heritage domain, not many studies have been conducted with regard to multilinguality and interface design, but some results from other domains can be used to improve usability of multilingual digital libraries. In the online retail business, many studies have been conducted to evaluate the impact of language and culture on website use.

---

[18]   http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-insight/insight-issue2-datamodelingdataquality

[19] http://www.europeanatech2015.eu/

[20] http://pro.europeana.eu/blogpost/we-want-good-quality-data-and-we-want-it-now

[21] http://oaei.ontologymatching.org/2014/library/results.html

6.4. Evaluating your User Interactions

For evaluating usability, user interfaces and user interactions in multilingual environments, the same basic usability heuristics can be applied as for monolingual interfaces or content. The heuristics by Nielsen[22] or the Eight Golden Rules of Interface Design by Ben Shneiderman[23] provide valid guidelines for performing evaluations in multilingual digital libraries. Qualitative methods ranging from interviews to observation as well as quantitative methods such as logfile analysis can be employed to evaluate usability in digital libraries with multilingual content and audiences. An overview on these methods is for example given by Lazar et al., 2010. When it comes to testing multilingual interfaces, translations and search performance in particular, more specific evaluation procedures have to be employed

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 of studying 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 with the intention of analyzing and classifying user queries in order to understand search behavior in multilingual contexts and to improve search systems (Mandl et al., 2009, 2010).

Between 2011 and 2013 the Cultural Heritage in CLEF (CHiC) lab used Europeana data 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.

---

[22] http://www.nngroup.com/articles/ten-usability-heuristics/
[23] https://www.cs.umd.edu/users/ben/goldenrules.html

7. References

Agosti, M., G. Angelaki, T. Coppotelli, and G. M. Di Nunzio. 2007. "Analysing HTTP Logs of a European DL Initiative to Maximize Usage and Usability." In *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers: 10th International Conference on Asian Digital Libraries, ICADL 2007, Hanoi, Vietnam, December 10-13, 2007. Proceedings*, edited by Dion Hoe-Lian Goh, Tru Hoang Cao, Ingeborg Torvik Sølvberg, and Edie Rasmussen, 35–44. Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-77094-7_9.

Angelaki, Georgia. 2007. "M1.4. Interim Report on Usability Developments in The European Library: -." EDL-Project.

Aula, Anne, and Melanie Kellar. 2009. "Multilingual Search Strategies." In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, 3865–70. New York, NY, USA: ACM.

Baierer, Konstantin, Evelyn Dröge, Vivien Petras, and Violeta Trkulja. 2014. "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*, 1–11.

Bates, Marcia J. 2002. "The Cascade of Interactions in the Digital Library Interface." *Inf. Process. Manage.* 38 (3): 381–400. doi:10.1016/S0306-4573(01)00041-3.

Bilal, Dania. 2006. "Measuring the Usability of an International User Interface: Culture and Design Representations." In *Human-Computer Interaction Symposium, American Society for Information Science & Technology*. http://web.utk.edu/~dania/HCI%20Symposium.doc.

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

Calhoun, Karen S, and Glenn Patton. 2011. *WorldCat Quality: An OCLC Report*. OCLC.

Chen, Jiangping, Olajumoke Azogu, and Ryan Knudson. 2014. "Enabling Multilingual Information Access to Digital Collections: An Investigation of Metadata Records Translation." In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, 467–68. IEEE Press.

Chen, Jianping. 2016. Multilingual Access and services for digital Collections. ABC-CLIO

Clough, Paul, and Mark Sanderson. 2006. "User Experiments with the Eurovision Cross-Language Image Retrieval System." *Journal of the American Society for Information Science and Technology* 57: 697–708.

"D 1.2 User Requirements Analysis." 2006. Multimatch-Project. http://www.multimatch.eu/docs/publicdels/D1.2Final.pdf.

"D1.3 Ontologies for Sound." 2014. Europeana Sound. http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D1.3-Ontologies-for-sound-v1.2.pdf.

"D2.2.1 Specifications of Post Querying Processing Functionalities." 2012. Assets. http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/ASSETS

/Deliverables/D2.2.1%20Specification%20of%20post%20querying%20processing%2
0functionalities.pdf.

"D2.3.1 Multilingual Mapping of Schemes and Vocabularies." 2011. EuropeanaConnect.
http://www.europeanaconnect.eu/documents/D2.3.1%20Multilingual%20mapping
%20of%20schemes%20and%20vocabularies-Draft.pdf.

"D3.1.3 Publishable Report on Best Practice and How Users Are 151 Using the Europeana
Service." 2011. EuropeanaConnect.
http://www.europeanaconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport
_v1.0.pdf.

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

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

"D4.4 Specialist Training Material : TMP User Manuals and Semantic Mapping Procedures."
2013. AthenaPlus. http://www.athenaplus.eu/getFile.php?id=301.

"D5.2 Report on Search Engines and HTTP Log Analysis." 2009. TELplus-Project.

"D7.4 User Requirements for Advanced Features." 2009. Cacao-Project.

Dangerfield, and Karlshoven. 2015. "Report and Recommendations from the Task Force
on Metadata Quality."
http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20
Quality%20Report.pdf.

Daniel, Alao Olujimi, Awodele Oludele, Rehema Baguma, and Theo Van Der Weide. 2011.
"Cultural Issues and Their Relevance in Designing Usable Websites." *Nternational
Journal of Innovative Technology & Creative Engineering* 1 (2): 20–29.

Diekema, Anne R. 2012. "Design and Development Considerations for a Multilingual
Digital Library." In *Design, Development, and Management of Resources for Digital
Library Services*, 1–12. IGI Global, 1-12.

Dobreva, Milena, and Sudatta Chowdhury. 2010. "A User-Centric Evaluation of the
Europeana Digital Library." In *The Role of Digital Libraries in a Time of Global Change.
ICADL 2010*, edited by Gobinda Chowdhury, Chris Koo, and Jane Hunter, 6102:148–
57. Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Doddington, George. 2002. "Automatic Evaluation of Machine Translation Quality Using N-
Gram Co-Occurrence Statistics." In *Proceedings of the Second International
Conference on Human Language Technology Research*, 138–45. HLT '02. San Francisco,
CA, USA: Morgan Kaufmann Publishers Inc.
http://dl.acm.org/citation.cfm?id=1289189.1289273.

Dragisic, Zlatan, Kai Eckert, Jérôme Euzenat, Daniel Faria, Alfio Ferrara, Roger Granada,
Valentina Ivanova, et al. 2014. "Results of the Ontology Alignment Evaluation
Initiative 2014." In *Proceedings of the 9th International Conference on Ontology
Matching-Volume 1317*, 61–104.
http://oaei.ontologymatching.org/2014/results/oaei2014.pdf.

Eleta, I., and J. Golbeck. 2012. "A Study of Multilingual Social Tagging of Art Images: Cultural Bridges and Diversity." In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 695–704. ACM.

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

Freire, Nuno. 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*, 283–88. Würzburg: Ergon Verlag.

Gäde, Maria. 2014. "Country and Language Level Differences in Multilingual Digital Libraries." Dissertation, Humboldt-Universität zu Berlin, Philosophische Fakultät I. http://edoc.hu-berlin.de/docviews/abstract.php?id=40595.

Gäde, Maria, and Vivien Petras. 2014. "Multilingual Interface Preferences." In *Proceedings of the 5th Information Interaction in Context Symposium*, 231–34. ACM.

Ghorab, M Rami, Johannes Leveling, Dong Zhou, Gareth JF Jones, and Vincent Wade. 2009. "Identifying Common User Behaviour in Multilingual Search Logs." In *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, 518–25. Springer.

Glinka, Katrin, Sebastian Meier, and Marian Dörk. n.d. "Visualising The» Un-Seen «: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces." In *Kultur Und Informatik. Cross Media*, 105–18. http://mariandoerk.de/papers/kui2015.pdf.

Gonzalo, Julio, Anselmo Peñas, Felisa Verdejo, and Carol Peters. 2008. "Workshop on Best Practices for the Development of Multilingual Information Access Systems: The User Perspective: D3.2 - TrebleCLEF-Project."

Guillot, Marie-Noëlle. 2014. "Cross-Cultural Pragmatics and Translation: The Case of Museum Texts as Interlingual Representation." In *Translation: A Multidisciplinary Approach*, 73–95. Palgrave Advances in Language and Linguistics. Heidelberg: Springer.

Hall, Mark, Hugo Huurdemann, Marijn Koolen, Mette Skov, and David Walsh. 2014. "Overview of the INEX 2014 Interactive Social Book Search Track." In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction. 5th International Conference of the CLEF Initiative, CLEF 2014, Sheffield, UK, September 15-18, 2014. Proceedings*, 212–28. http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-HallEt2014.pdf.

Hearst, Marti. 2009. *Search User Interfaces*. Cambridge: Cambridge University Press.

House, Juliane. 2014. "Translation Quality Assessment: Past and Present." In *Translation: A Multidisciplinary Approach*, 241–64. Palgrave Advances in Language and Linguistics. Heidelberg: Springer.

Isaac, Antoine, Hugo Manguinhas, Juliane Stiller, and Valentine Charles (eds.). 2015. "Report on Enrichment and Evaluation." EuropeanaTech Task Force on Evaluations and Enrichments. http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation//FinalReport_EnrichmentEvaluation_102015.pdf.

Isaac, Antoine, Shenghui Wang, Claus Zinn, Henk Matthezing, Lourens Van der Meij, and Stefan Schlobach. 2009. "Evaluating Thesaurus Alignments for Semantic Interoperability in the Library Domain." *IEEE Intelligent Systems*, no. 2: 76–86.

Jiangping Chen, Ren Ding, Shan Jiang, and Ryan Knudson. 2012. "A Preliminary Evaluation of Metadata Records Machine Translation." *The Electronic Library* 30 (2): 264–77.

Keegan, Te Taka, and Sally Jo Cunningham. 2005. "Language Preference in a Bi-Language Digital Library." In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2005, Denver, CO, USA, June 7-11, 2005, Proceedings*, 174–75. doi:10.1145/1065385.1065427.

———. 2008. "What a Difference a Default Setting Makes." In *Research and Advanced Technology for Digital Libraries, 12th European Conference, ECDL 2008, Aarhus, Denmark, September 14-19, 2008. Proceedings*, 264–67. doi:10.1007/978-3-540-87599-4_28.

Király, Péter. 2015. "Query Translation in Europeana." *Code4Lib*, no. 27. http://journal.code4lib.org/articles/10285.

Lamm, Katrin, Thomas Mandl, and Ralph Koelle. 2009. "Search Path Visualization and Session Performance Evaluation with Log Files." In *10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, September 30 - October 2, 2009, Revised Selected Papers*, 538–43. Springer.

Landry, Patrice. 2009. "Multilingualism and Subject Heading Languages: How the MACS Project Is Providing Multilingual Subject Access in Europe." *Catalogue & Index: Periodical of CILIP Cataloguing and Indexing Group* 157.

Lavie, Alon, Kenji Sagae, and Shyamsundar Jayaraman. 2004. "The Significance of Recall in Automatic Metrics for MT Evaluation." In *Machine Translation: From Real Users to Research*, 134–43. Springer.

Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Chichester, West Sussex U.K.: Wiley.

Lemmens, Bert, Joris Janssens, Ruth Van Dyck, Alessia Bardi, Paolo Manghi, Eric Beving, Kathryn Máthé, Katalin Dobó, and Armin Straube. 2011. "D2.2 The Common HOPE Metadata Structure, Including the Harmonisation Specification." Deliverable 2.2. HOPE - Heritage of the People's Europe. http://www.peoplesheritage.eu/pdf/D2_2_Metadata%20Structure.pdf.

Leveling, Johannes, M Rami Ghorab, Walid Magdy, Gareth JF Jones, and Vincent Wade. 2010. "DCU-TCD@ LogCLEF 2010: Re-Ranking Document Collections and Query Performance Estimation." In *CLEF 2010 LABs and Workshops, Logfile Analysis (LogCLEF)*. http://doras.dcu.ie/15835/1/DCU-TCD%40LogCLEF_2010.pdf.

"Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations." 2005. Revision 1.5. LDC (Linguistic Data Consortium). www.ldc.upenn.edu/-Projects/TIDES/Translation/TransAssess04.pdf.

Mandl, Thomas, Maristella Agosti, Giorgio Maria Di Nunzio, Alexander Yeh, Inderjeet Mani, Christine Doran, and Julia Maria Schulz. 2009. "LogCLEF 2009: The CLEF 2009 Multilingual Logfile Analysis Track Overview." In *10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, September 30 - October 2, 2009, Revised Selected Papers*, 508–17. Springer.

Mandl, Thomas, Giorgio Maria Di Nunzio, and Julia Maria Schulz. 2010. "LogCLEF 2010: The CLEF 2010 Multilingual Logfile Analysis Track Overview." In *Working Notes 11th Workshop of the Cross-Language Evaluation Forum, CLEF 2010*. http://clef2010.org/resources/proceedings/LogCLEF-2010-Overview.pdf.

Manguinhas, Hugo. 2014. "Semantic Enrichment Framework." https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y.

Marlow, J., P. Clough, J. Cigarrán Recuero, and J. Artiles. 2008. "Exploring the Effects of Language Skills on Multilingual Web Search." In *30th European Conference on IR Research, ECIR 2008, Glasgow, UK, March 30-April 3, 2008. Proceedings*, 126–37. Springer.

Matusiak, Krystyna K, Ling Meng, Ewa Barczyk, and Chia-Jung Shih. 2015. "Multilingual Metadata for Cultural Heritage Materials: The Case of the Tse-Tsung Chow Collection of Chinese Scrolls and Fan Paintings." *The Electronic Library* 33 (1): 136–51.

"Multilingual Terminology Of Art Nouveau 'Micro' Object Types." n.d. D3.1. PartagePlus. http://www.partage-plus.eu/en/servefile?id=34.

Nantel, Jacques, and Evelyne Glaser. 2008. "The Impact of Language and Culture on Perceived Website Usability." *Research on Management of Technology and Innovation in a Global Context* 25 (1–2): 112–22. doi:10.1016/j.jengtecman.2008.01.005.

Oakes, Michael P., and Yan Xu. 2010. "A Search Engine Based on Query Logs, and Search Log Analysis by Automatic Language Identification." In *10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, September 30 - October 2, 2009, Revised Selected Papers*, 6241:526–33. LNCS. Springer.

Oard, Douglas W., Julio Gonzalo, Mark Sanderson, Fernando López-Ostenero, and Jianqiang Wang. 2004. "Interactive Cross-Language Document Selection." *Information Retrieval* 7 (1): 205–28. doi:10.1023/B:INRT.0000009446.22036.e3.

Olensky, Marlies, Juliane Stiller, and Evelyn Dröge. 2012. "Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy." In *6th Research Conference, MTSR 2012, Cádiz, Spain, November 28-30, 2012. Proceedings*, 252–63. Berlin: Springer.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. "BLEU: A Method for Automatic Evaluation of Machine Translation." In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311–18. Association for Computational Linguistics.

Peinado, Victor, Javier Artiles, Julio Gonzalo, Emma Barker, and Fernando López-Ostenero. 2008. "FlickLing: A Multilingual Search Interface for Flickr." In *In Working Notes for the CLEF 2008 Workshop*.

Peters, Carol, Martin Braschler, and Paul Clough. 2012. *Multilingual Information Retrieval: From Research to Practice*. Springer: Heidelberg, Germany.

Petras, Vivien, Nicola Ferro, Maria Gäde, Antoine Isaac, Michael Kleineberg, Ivano Masiero, Mattia Nicchio, and Juliane Stiller. 2012. "Cultural Heritage in CLEF (CHiC) Overview 2012." In *CLEF (Online Working Notes/Labs/Workshop)*. http://www.clef-initiative.eu/documents/71612/0cadb163-3e32-4f16-a659-b457480c2a29.

Petras, Vivien, Juliane Stiller, and Maria 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*, edited by Colleen Cool and Kwong Bor Ng, 141–63. doi:10.4018/978-1-4666-2991-2.ch009.

Petrelli, Daniela, Micheline Beaulieu, and Mark Sanderson. 2002. "User Requirement Elicitation for Cross-Language Information Retrieval." *The New Review of Information Behavior Research* 3: 17–35.

Ruecker, Stan, Milena Radzikowska, and Stéfan Sinclair. 2011. *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing*. Ashgate Publishing, Ltd.

Srinivasarao, Vundavalli. 2008. "Mining the Behavior of Users in a Multilingual Information Access Task." presented at the CLEF 2008 Working Notes. http://clef.isti.cnr.it/2008/working_notes/vundavalli-paperCLEF2008.pdf.

Steichen, Ben, and Luanne Freund. 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*, 3483–92. CHI '15. New York, NY, USA: ACM. doi:10.1145/2702123.2702541.

Stiller, J., M. Gäde, and V. Petras. 2013. "Multilingual Access to Digital Libraries: The Europeana Use Case." *Information-Wissenschaft Und Praxis* 64: 86–95.

Stiller, Juliane, Maria Gäde, and Vivien Petras. 2011. "Is Tagging Multilingual?: A Case Study with BibSonomy." In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, 421–22. JCDL '11. New York, NY, USA: ACM. doi:10.1145/1998076.1998165.

Stiller, Juliane, Antoine Isaac, and Vivien Petras (eds). 2014. "EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: Final Report." http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/MultilingualSemanticEnrichment//Multilingual Semantic Enrichment report.pdf.

Stiller, Juliane, Marlies Olensky, and Vivien Petras. 2014. "A Framework for the Evaluation of Automatic Metadata Enrichments." In *Metadata and Semantics Research: 8th Research Conference, MTSR 2014, Karlsruhe, Germany, November 27-29, 2014. Proceedings*, edited by Sissi Closs, Rudi Studer, Emmanouel Garoufallou, and Miguel-Angel Sicilia, 238–49. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-13674-5_23.

Stiller, Juliane, Vivien Petras, Maria Gäde, and Antoine Isaac. 2014. "Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences." In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, 238–47.

Toms, Elaine, and Mark Hall. 2013. "The CHiC Interactive Task (CHiCi) at clef2013." In *CLEF 2013 Evaluation Labs and Workshops, Online Working Notes*. http://ceur-ws.org/Vol-1179/CLEF2013wn-CHiC-TomsEt2013.pdf.

Vassilakaki, Evgenia, Frances Johnson, and Richard J Hartley. 2012. "Image Seeking in Multilingual Environments: A Study of the User Experience." *Information Research* 17 (4).

Vyncke, Femke, and Malaika Brengmann. 2010. "Are Culturally Congruent Websites More Effective? An Overview of a Decade of Empirical Evidence." *Journal of Electronic Commerce Research* 11 (1).

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

Wu, Dan, Daqing He, and Bo Luo. 2012. "Multilingual Needs and Expectations in Digital Libraries: A Survey of Academic Users with Different Languages." *The Electronic Library,* 30 (2): 182–97.