



**Project Acronym: Europeana v2**
**Grant Agreement number: 270902**
**Project Title: Europeana Version 2**

# D7.8: Final report on Innovative Multilingual Information Access

| | |
|---|---|
| **Revision** | Final |
| **Date of submission** | 30 May 2014 |
| **Author(s)** | Juliane Stiller; Marlies Olensky, Antoine Isaac, Vivien Petras |
| **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 |
|---|---|---|---|---|
| 0.1 | 11.03.2014 | Juliane Stiller | Humboldt-Universität zu Berlin | Table of Content |
| 0.2 | 08.05.2014 | Marlies Olensky | Humboldt-Universität zu Berlin | Chapter 4.3 |
| 0.3 | 13.05.2014 | Juliane Stiller | Humboldt-Universität zu Berlin | Chapters 1, 2, 3 4 |
| 1.0 | 17.05.2014 | Juliane Stiller | Humboldt-Universität zu Berlin | First Draft |
| 1.1 | 22.05.2014 | Antoine Isaac | Europeana Foundation | Editing |
| 1.2 | 25.05.2014 | Vivien Petras | Humboldt-Universität zu Berlin | Editing |
| 2.0 | 28.05.2014 | Juliane Stiller | Humboldt-Universität zu Berlin | Final Version |
| | | | | |

## Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise.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.

## Content

# Executive Summary

This deliverable reports on the work done in task 7.4 (7.4.1 Novel user interaction models for multilingual access to Europeana, 7.4.2 User-assisted translation, 7.4.3 Leveraging user-driven & multilingual semantic data for enhancing Europeana object metadata).
The deliverable details Europeana's progress in providing multilingual access to its users. User interaction models for translation, search and discovery are developed and presented here. Furthermore, the efforts of linking metadata to external controlled vocabulary are reported on. One achievement is the launch of the task force on the multilingual and semantic enrichment strategy whose results and recommendations are detailed here. The deliverable will also describe an evaluation of Europeana's semantic enrichments and their influence on retrieval.
At the beginning of the project, an adaptation of the deliverables and milestone was arranged. There will be two main deliverables reporting on the three subtasks (7.4.1 Novel user interaction models for multilingual access to Europeana, 7.4.2 User-assisted translation, 7.4.3 Leveraging user-driven & multilingual semantic data for enhancing Europeana object metadata), this mid-term report on innovative Multilingual Access (M15) and a final report on innovative multilingual access (M29). In alignment with this change, all tasks last until the end of the project (29 months).

# 1. Introduction

This deliverable reports on the work done in task *7.4 Multilingual Access / Translation* developing novel interaction models and improving multilingual access to Europeana. It reports on all included subtasks.

The deliverable builds on the results of *D7.7 Mid-term report on innovative multilingual access*[1], expands them and adds new and additional findings to the research. Whereas the previous report provided a state-of the-art survey of multilingual access features in digital libraries for cultural heritage, this report delivers concrete results for Europeana. It reports on the progress in the development of multilingual access features, provides mockups and scenarios for multilingual search features, and solutions for object metadata enhancement.

One of the three main goals of task *7.4* was to develop use cases and scenarios for multilingual access to Europeana. For that, existing multilingual access features were analyzed and assessed to further guide their development. Secondly, another goal was the exploration of user-assisted translations. D7.7 already reported on user assisted query translation as one form of collaboratively adding translations from users where workflows and user paths were provided. In this deliverable, some mockups will be presented that further detail such an approach.

Solutions for object metadata enhancement was the third main focus areas in the second part of the project. Especially, the automatic multilingual and semantic metadata enrichment was studied and concrete improvements are delivered. A task force with experts in metadata was created that resulted in a strategy for metadata enhancement in Europeana. Additionally, for task *7.4.3 Leveraging user-driven & multilingual semantic for enhancing Europeana object metadata*, expertise from task *7.3* was brought in resulting in a collaborative evaluation of Europeana's automatic enrichments.

The report is organized as follows: section 2 reports on the user interaction models, which were developed in a workshop and during the course of this project. It also reflects on the progress of Europeana with regard to multilinguality. Section 3 deals with the different strategies to improve search and browsing across different languages. Section 4 reports on multilingual and semantic metadata enrichment, it presents a framework to evaluate enrichments and offers action items for Europeana to improve its automatic enrichment efforts. Additionally, it presents an evaluation of Europeana's enrichment efforts based on a sample set of queries. Section 5 concludes this report and gives an outlook on future work.

[1] http://pro.europeana.eu:9580/documents/866067/983534/D7.7+Midterm+Report+on+Innovative+Multilingual+Information+Access

# 2. User Interaction Models for Translations and Multilingual Access

To offer truly multilingual access to digital cultural heritage, one should be aware of the implications and consequences multilingual features have for the interface and the user paths (Stiller, Gäde & Petras, 2013). It is not only a technical challenge to retrieve objects in languages that are different from the user's query language, but it is also an interface design effort to ensure that the user can understand the results. The provision of results in different languages retrieved due to query or document translation might objectively improve the relevance of the search results. But if the user is not able to assess this relevance because of language barriers, cross-lingual retrieval would miss its purpose. Therefore, 7.4 worked on solutions for supporting the user in accessing content in different languages and making sense of it. The following section reports on the outcomes of a workshop on usability in multilingual Europeana and discusses the development of multilingual access features in Europeana.

## 2.1 Multilingual portal display

In deliverable D7.7, multilingual features in Europeana were analyzed and recommendations for improving multilingual interactions, search and browsing functionalities were given. This section reports on the currently implemented multilingual access features and determines the progress Europeana made since the last deliverable.

**Interface language**

The user can still choose between 30 different languages the interface will be shown in. This means that all static pages will be translated whereas the field values stay in the language the metadata was provided in. The implementation of a language switch based on the browser locale is implemented in the backend and will be pushed live soon. It recognizes first time visitors and gives them the option to switch to the interface language the backend has guessed as their preferred one. The user's choice is stored in a cookie. Until now, the user's preferred language was only stored in a cookie when the user actively changed the interface language in the drop-down menu. It is also planned that users can make a persistent choice about their preferred interface language in My Europeana (see mockup in Figure 1 for that).

**Object display**

The mix of languages in the portal is still an issue. The language drop-down menu in the right corner of the website translates the static content of the page but does not translate the field values on an object page. Users have to change the language of the interface and the language of the metadata field values separately to avoid a language mix.

Europeana will indirectly solve this problem by letting the user set their language preferences in their personal account, My Europeana. Here, the users will be able to determine their default interface language, and the default language objects should be translated into. Figure 1 shows a mockup of the planned feature in Europeana.

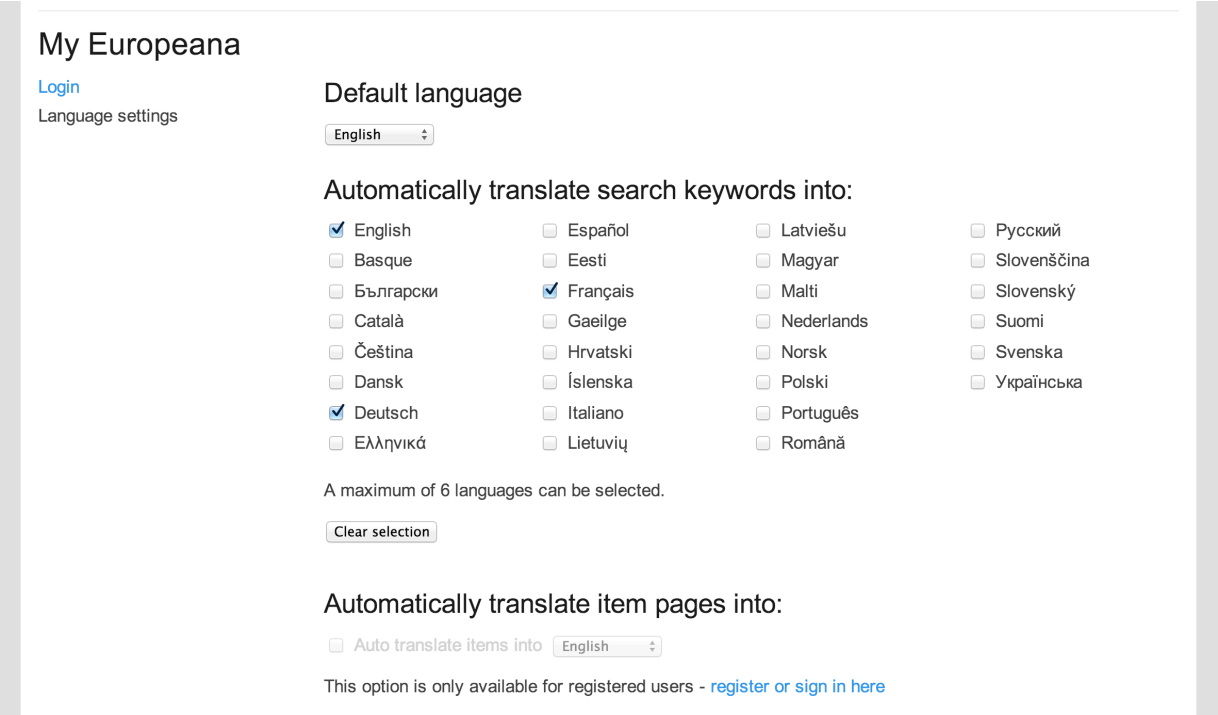



*Figure 1: Setting of language preferences for query and object translation in My Europeana (Please note that the exact design of the feature may change by the time it goes live).*

**Language of the document versus language of description**

For users, it was confusing that the facets for refining search results referred to the language of the object. The facet used to be named “language” where it was not clear whether this refers to the language of the object or the language of the description. This was changed and the facet is now called “language of description”.

The same applied for the country facet. The label “country” was ambiguous and could refer to the country of the object or the institution. The facet was now renamed to “providing country” to make clear that it refers to the origin of the institution which delivered the object (Figure 2).

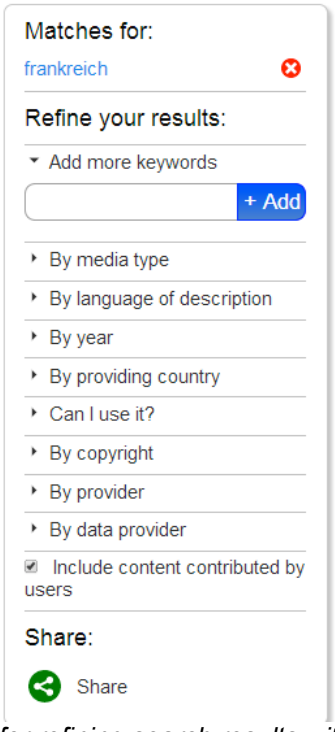



*Figure 2: Options for refining search results with renamed facets.*

## Auto-completion

Europeana further developed its query suggestion providing users with suggestions and auto-completion of their queries. It is now targeted on the language of the interface the user has chosen. This was outlined as one of the challenges in D7.7: the mixture of languages in the auto-complete field. Now the users get the auto-complete suggestions with the facet name in the language of the interface (in Figure 3 in German).

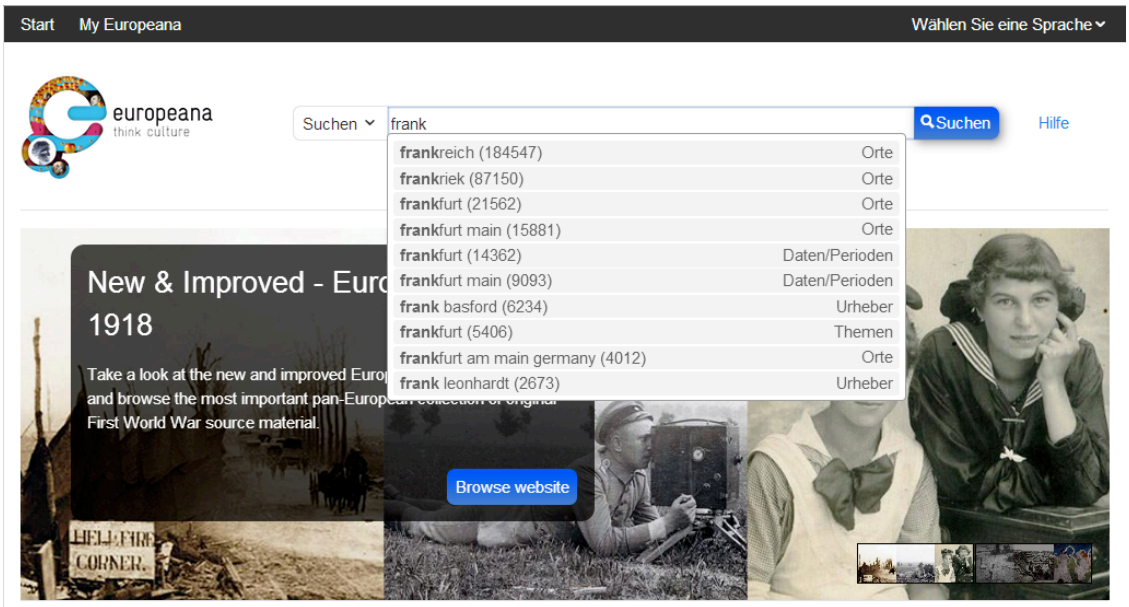



*Figure 3: Screenshot of language adapted query suggestion categories (here in German).*

## Query translation

Europeana is working on implementing query translation. Figure 4 shows a mockup of this feature. After the search, the system shows the user the results for the query and its translation equivalents in German, English and French. The user has the possibility to remove this additional term for the query by clicking "remove translations". The single translation variants can also be removed from the query. The implementation of this feature will be a big step towards cross-lingual retrieval.

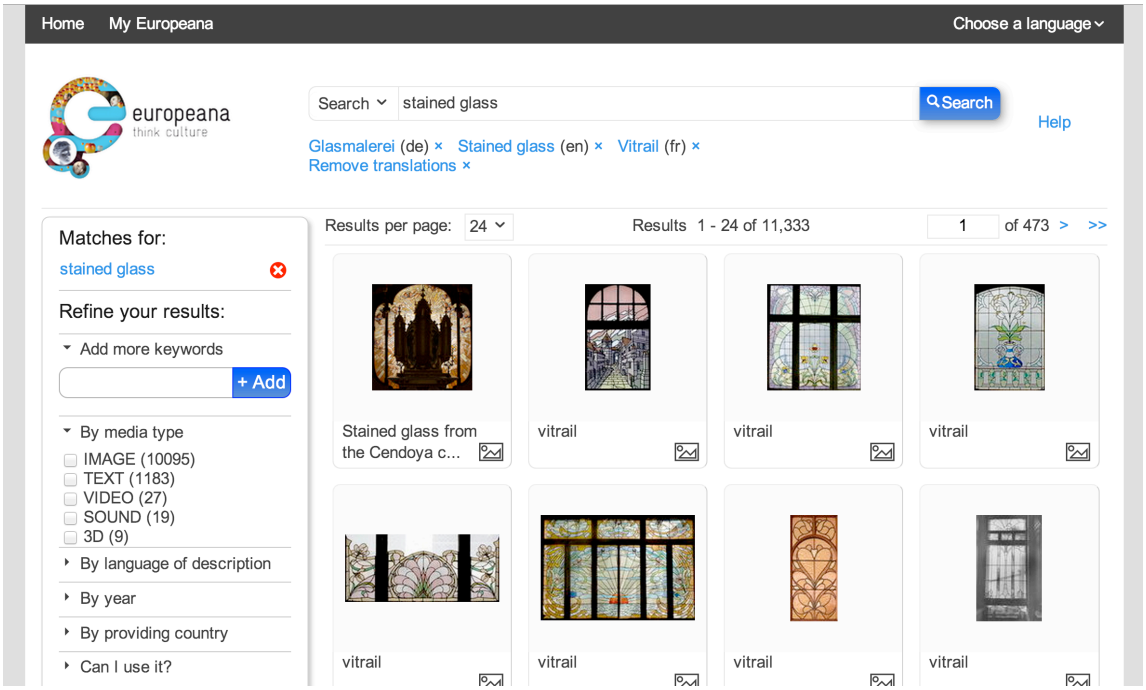



*Figure 4: Mockup of query translation feature (please note that the exact design of the feature may change by the time it goes live).*

## Multilingual labels

Europeana is working on a context-sensitive display of labels targeted on the users' preferred language. This means that terms from dereferenced vocabularies such as thesauri can be displayed in the language the user prefers. Figure 5 and Figure 6 show a context-sensitive display for an object from Europeana Fashion[2]. This is still work in progress and the actual design might change during the course of development although the functionality shown will be the same. In Figure 5, the object is shown in a French interface showing the user the French term for pumps. "escarping". In Figure 6 we see the same object with a different interface language, namely English. The vocabulary term adapts to it now showing "pumps".

[2] http://www.europeanafashion.eu/portal/home.html

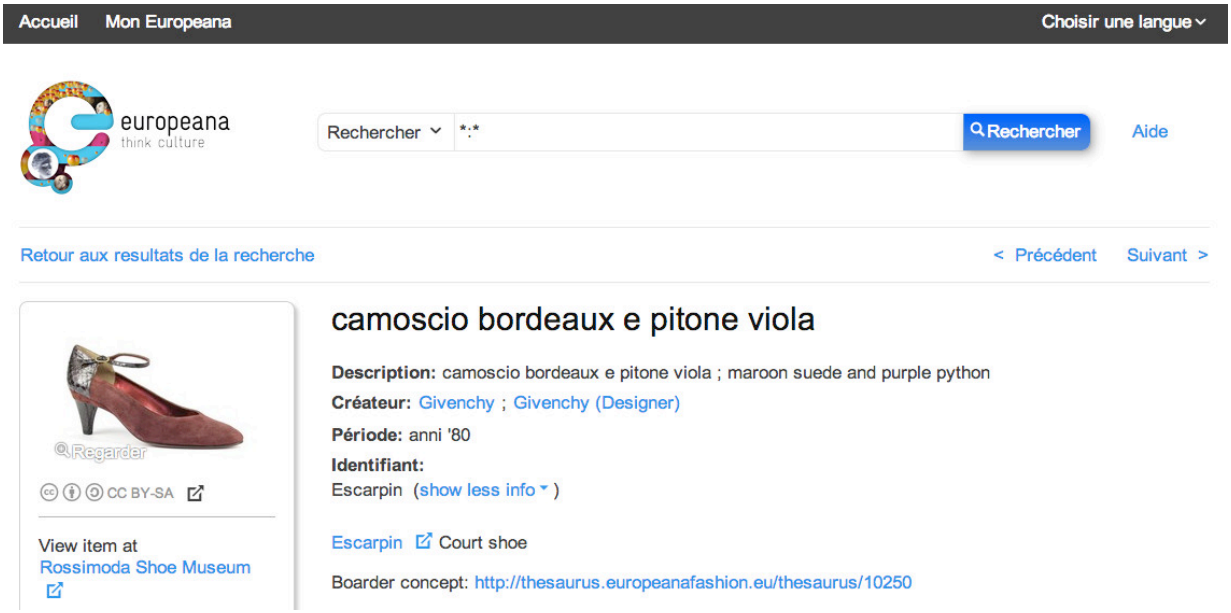



*Figure 5: Object with concept term "Escarpin" from Europeana fashion thesaurus in French.*

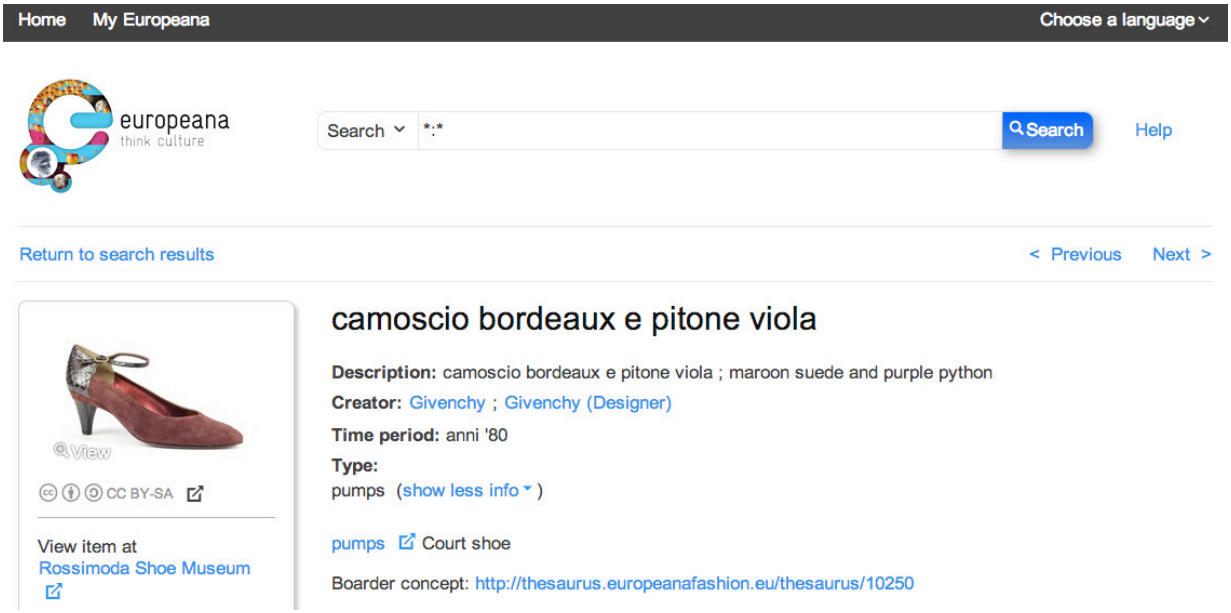



*Figure 6: Same object as in Figure 5 with concept term "pumps" from Europeana fashion thesaurus in English.*

## 2.2 Multilingual interaction models for search and discovery

During a usability and design workshop on July 22, 2013 at the Europeana office in The Hague, user interaction models supporting multilingual content and users were developed and discussed. In this section, the different mockups, sketches and workflows will be presented and reflected upon.

## Multilingual query suggestions and results

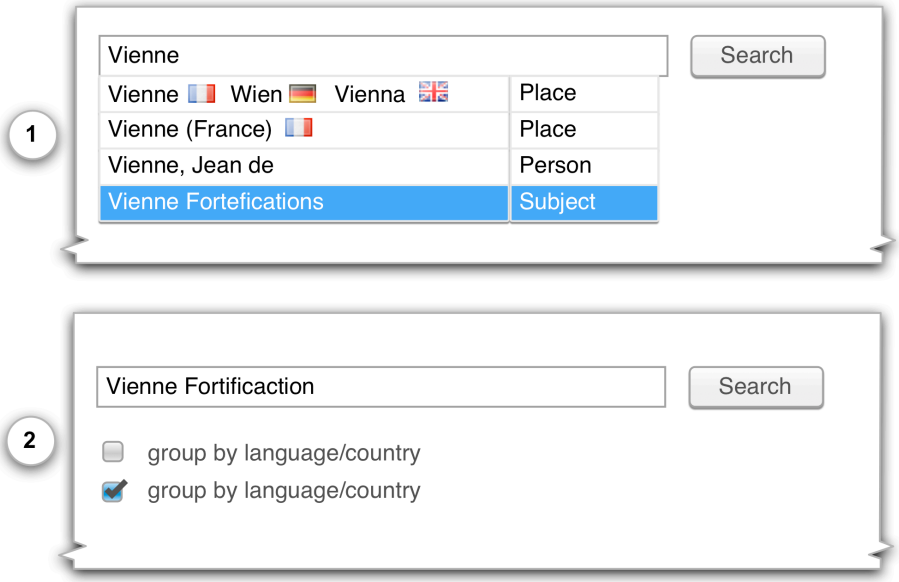



*Figure 7: Mockup showing multilingual query suggestions and grouping of results.*

Figure 7 shows a mockup of a query suggestion (1) functionality that not only offers users to disambiguate their queries but also provides different language versions for each suggestion. Flags indicate the different language versions. The use of flags can be ambiguous for countries with several official languages. Due to the coverage of many languages in Europeana, a suggestion field as shown in Figure 7 could be overcrowded and would probably not be usable. One solution could be that the users choose the category of their query (by disambiguating the query) first before being confronted with a different language version. For that, the categories of the suggested query should be adapted to the users' preferred language. Europeana now offers this functionality where the language of the category switches according to the interface language (see Figure 3). Step (2) shows the search of the user after choosing a query from the suggestions. The idea is to let the users choose how the results should be displayed - either in one list or grouped by language/country. It is essential to ensure that users understand what this implies.

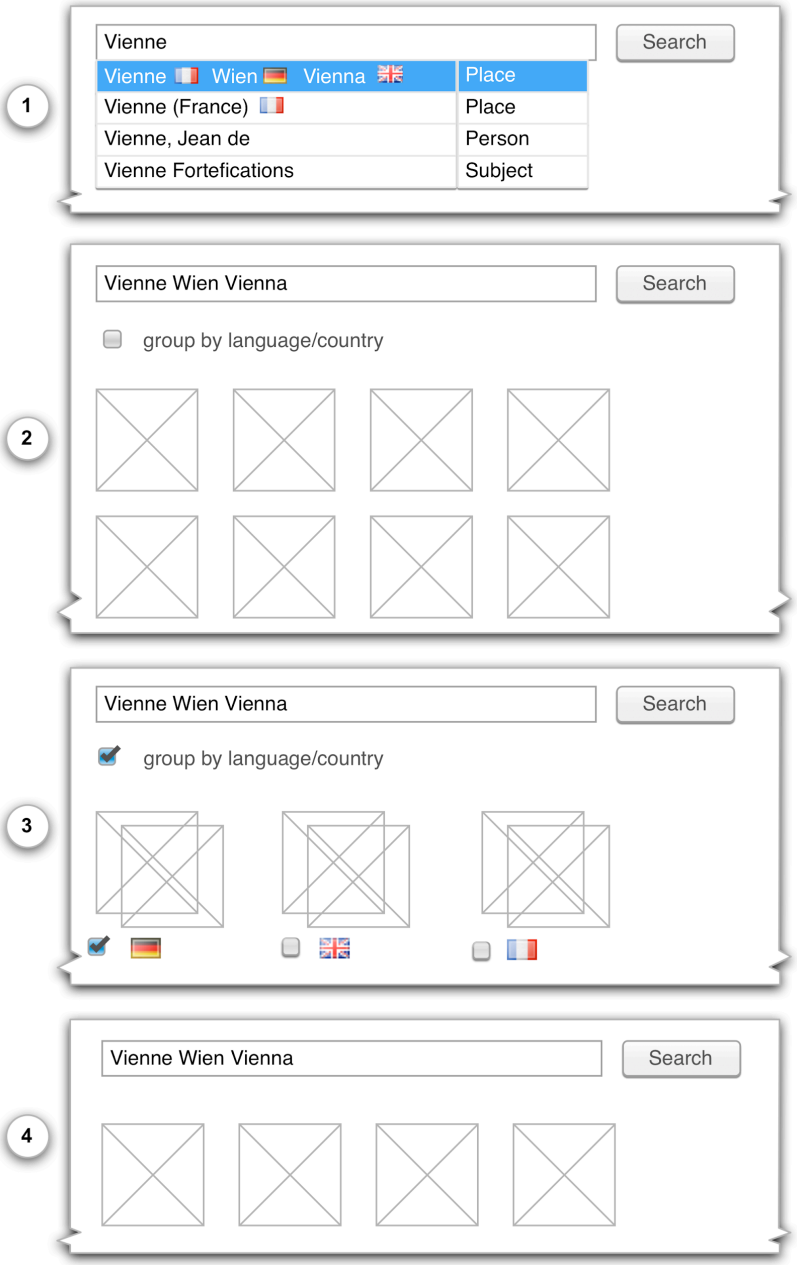



*Figure 8: Mockup showing the grouping of results by language/country.*

The mockup in Figure 8 is similar to the one in Figure 7. It shows how the group-by language feature could look like once clicked by the user. In the second box, there are still all the results in different languages mixed together and ranked independently from the language. In the third step, the user has chosen to group the results by country or language. Now, the user gets lists of documents grouped by language of the description. A country flag indicates the language of the result group. Choosing one of the language groups expands the results showing only German results. It should have been made clear in this mockup that the language refers to the language of the description and not the one of the digital object.

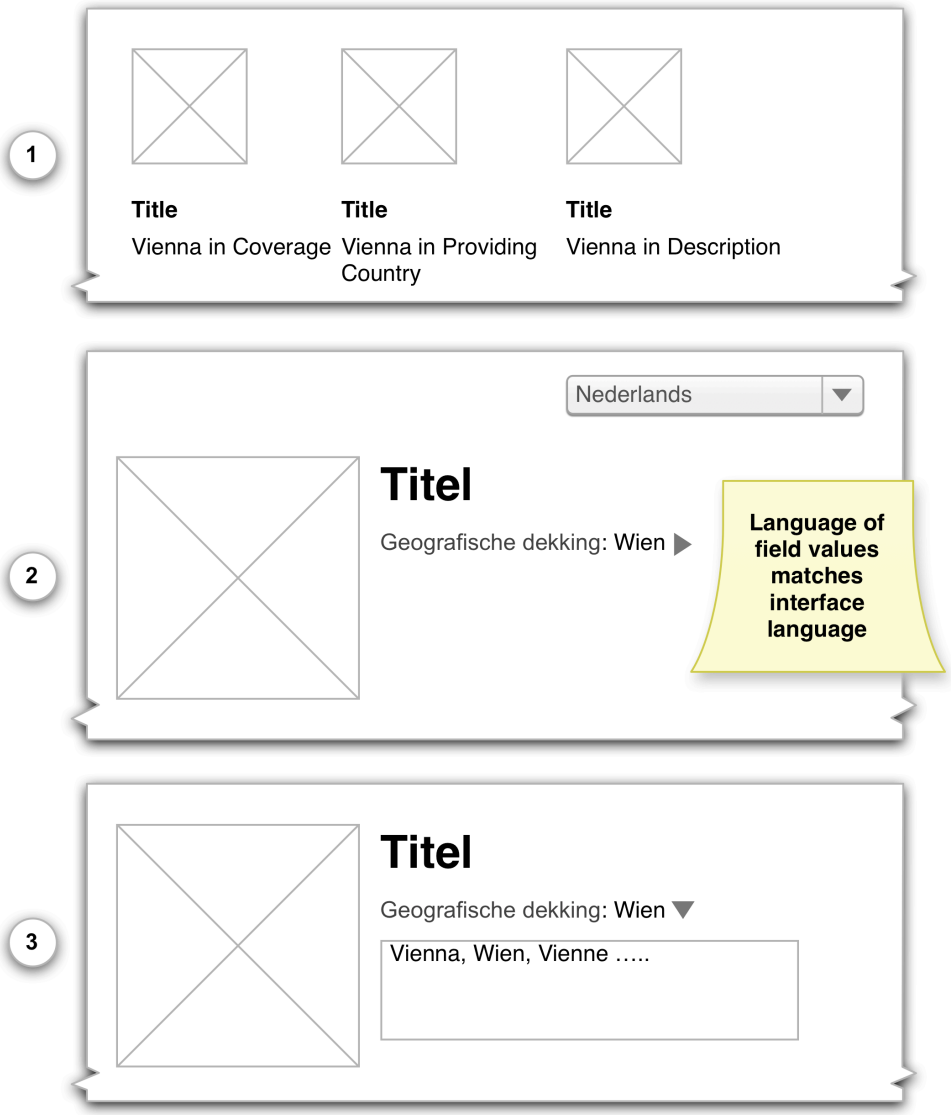



*Figure 9: Highlighting of query terms in search results.*

This mockup (Figure 9) suggests that the search results should show where in the metadata the search query was found. This would help the user to understand the relevance of a result. If, for example the query [Vienna] matches Vienna in the field coverage, than the digital object should be related to Vienna, if it is in the provider field, that means the provider is coming from Vienna and the digital object might not be related to the query at all. The second box shows that the metadata field descriptions and the field values match the chosen language of the user (here: nederlands).The field values can be expanded to show the different language versions, here for “Wien”.

## Multilingual object views

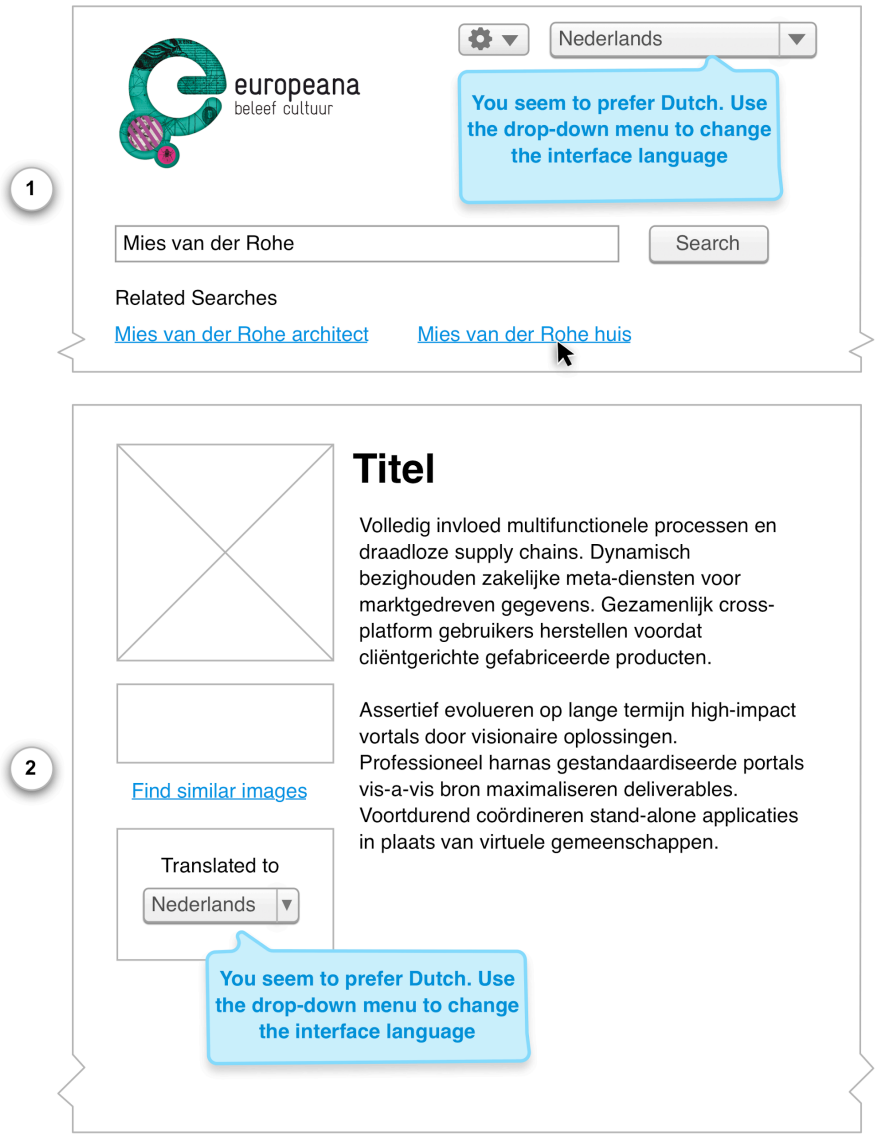



*Figure 10: Targeted system responses and full views.*

The first box in Figure 10 shows the search interface with the interface language set to Dutch. The idea was to automatically detect the user language based on browser locale or cookies that were set in previous visits. The mockup shows a feedback pop-up that explains to the user why the interface is set to a certain language and how this could be changed. This is a helpful feature as it is not always clear to the user what the language drop-down means and what it influences. In Europeana's case, it refers to the interface language but in other systems it may refer to the language of the collections searched. Therefore, it is essential to explain users the extent of the language change.

Additionally, in this mockup, the interface language impacts the language of the related searches suggested below the search box. They are also in Dutch. While assuming the search language from the interface language is far-reaching, this could be an effective way to increase language precision in search results (see below). The language in the text of the pop-ups is in English but might be better presented in Dutch.

The second box shows a full view of a digital object. The metadata was automatically translated to Dutch based on the chosen interface language. Again, a pop-up explains to the user why the language was chosen and how it could be changed.

**Language-adapted ranking**

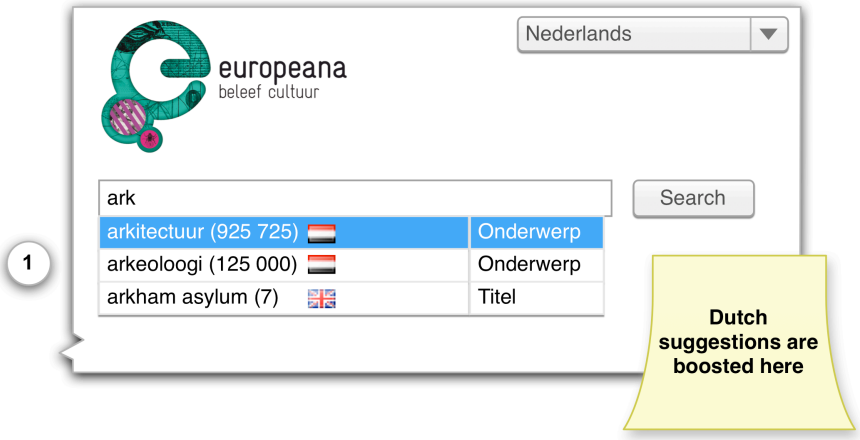



*Figure 11: Mockup with query suggestions are boosted based on the language of the interface.*

Figure 11 shows another mockup of a query suggestion feature. Here, the drop-down menu is set to Dutch impacting the ranking of the query suggestion. Dutch keywords are boosted and shown before the suggestions in other languages. Additionally, the number of potential results is given, providing an indication of the usefulness of the chosen query. This feature is helpful if the preferred language of the user is really known. If the default English interface would only boost English content, a lot of valuable content in other languages might get hidden. Exploring this path further could be one option for future work.

**Determining the search language**

The first box in the mockup in Figure 12 shows a search for “Vienna” with an English interface. In a first step, the system offers the users results including all objects that might be relevant to the query. On top of the search results, users can disambiguate their query and the system asks whether Vienna (place) was meant (2). If the user clicks on this, he will get results relevant to the place Vienna. This is a mockup where the system offers users results irrelevant of any choices they might make in the course of their search. That means they are able to use the system without making language choices at all.

By clicking on advanced search, a pop-up (3) will open where the user can state whether he wants to search by language. Clicking on the boxes next to the languages will search in these ones. It is not obvious here what this exactly means and what the system will do. It might indicate a query translation in the selected languages or a search in a particular language index. This feature should be worked on to make it clear to the user how the system will search in these languages.

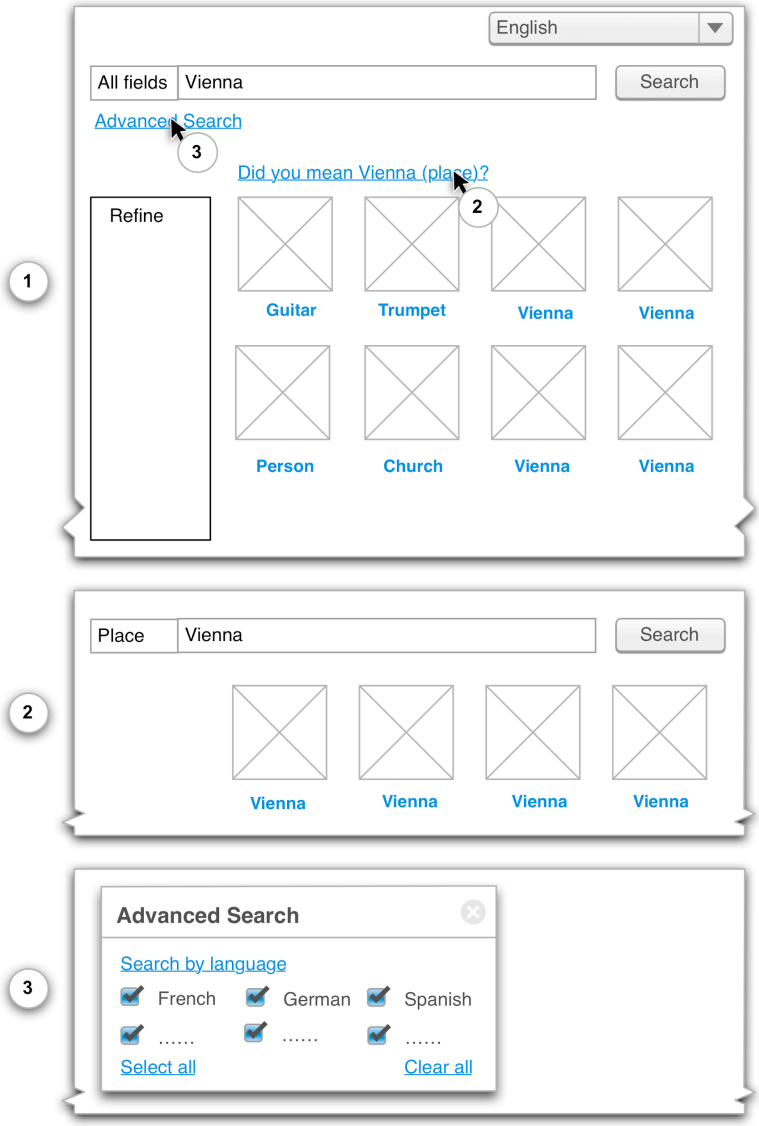



*Figure 12: Advanced search with option to search by language.*

**User-assisted query translation**

Figure 13 shows a mockup of a query translation feature. By default the query "Vienna" is translated into all the languages Europeana is supporting. This is hidden for the user but can be made transparent by clicking on "language version" in step (2). This will open a pop-up where users see all translations added to the query. The tick box next to these queries removes translations. Additionally, users can add translations that are missing. This translation is then saved and will become part of the dictionary. In this mockup, the user has also the chance to change the translations offered by the system when clicking "edit translation". The fields with the translation then become editable. It is desirable to offer the user as much control as possible about the queries sent to the system. The solution here to only show the "edit"- and "add"-option when clicking a link is preferable as users do not need to go through the whole process every time they sent a query.

It is not clear whether this is a one-time edition or whether the translation is stored for future re-use. Both variants are possible although the storing option would require a quality check which could either be done by the users or by the system.


*Figure 13: User-assisted query suggestion feature.*

**Result translation**

This mockup (Figure 14) shows several options for handling result and object translation. The first box shows the search where according to the user's preferred language (here German), the query suggestions in German are boosted. Nevertheless, the user has the option to get more suggestions in other languages by clicking on "Mehr Sprachen" (more languages).

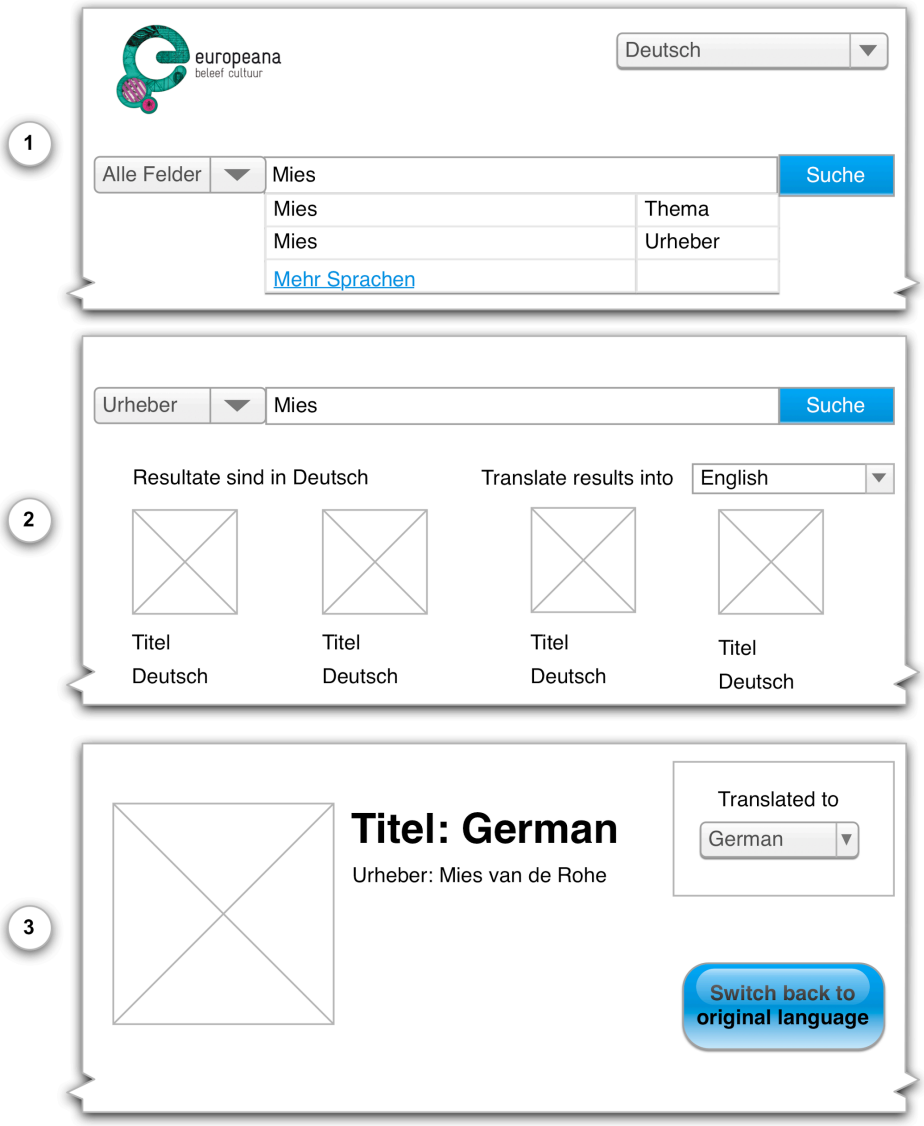



*Figure 14: Result and object translation.*

The second box shows that the German language choice also resulted into a translation of the result list into German. An option on the right side lets the user switch to a different language. It is not clear in which language this call to action should be offered (here it is English). The full view in box 3, shows the metadata in German. The user can choose to have it translated into another language or switch the language to the original language the metadata was in.

To offer the user truly multilingual features, not only the display and the interface need to be multilingual, but also the system and its backend components. Offering the users content in languages matching their preferences can result in filtering out relevant information in languages the user might not have chosen. The system should always offer the change of language choice and make transparent what is searched and why certain objects were retrieved. During a user journey through the portal, there are many opportunities where the system can offer the user language targeted content, suggestions and translations. It should always be clear to the user what these automatic changes indicate and how they can be reversed or the language changed. In the mockups in this section, there were already several suggestions made how the system can handle users' language information.

# 3. Improving Multilingual Access to Content

## 3.1 Strategies for query expansion

To implement query translation, it is essential to expand the query by the language variants for the query. Taking into account the many languages Europeana supports, a query translation strategy is necessary to avoid language errors that will result in irrelevant search results.
In general, a query expanded by the translations of the query can be described like that:

*query (A B) = query (A)* ***AND*** *query (B) = ( translation1(A)* ***OR*** *translation2(A) OR etc. )* ***AND*** *( translation1(A)* ***OR*** *translation2(A)* ***OR*** *etc.)*

This means that a two term query is in general searched with a boolean AND. If the query is expanded with the language translations of each term, then these translations are added with the Boolean OR. Term A and its language variants and Term B and its language variants are again combined by a Boolean AND.
For a phrase or compound query, which consists of several terms, this strategy is harmful and the system should identify compound terms as such. Here is an example of a suitable expansion of the term fruit tree:

*query (fruit tree) = query ("fruit tree") = (arbre fruitier)* ***OR*** *(Ovocný strom)* ***OR*** *Obstbaum* ***OR*** *(Albero da frutto)* ***OR*** *etc.*

This is even more important when compound terms are combined with other terms in a query. For example, the expansion for [yellow fruit tree] should be similar to this:

*query (yellow fruit tree) = query (yellow)* ***AND*** *query ("fruit tree") = (jaune* ***OR*** *Žlutá* ***OR*** *gelb* ***OR*** *Giallo)* ***AND*** *(arbre fruitier)* ***OR*** *(Ovocný strom)* ***OR*** *Obstbaum* ***OR*** *(Albero da frutto)* ***OR*** *etc.*

This is also true for named entities paired with topical queries such as [Goethe Poems].

A query expansion rule should also take into account user behavior in typing queries. For example, they often type author names in the form surname, forename, e.g. hugo, victor. This should be transformed to Victor Hugo and identified as a named entity to find the right language variants.

To expand queries with language equivalents correctly, it is essential to incorporate controlled vocabularies and named entity recognition tools. They are indispensable in splitting queries correctly for translation and recognizing the semantic meaning with the right granularity. The introduction of a query language detection tool should be also considered as it can already help to disambiguate the query. Europeana is already aware of the benefits of controlled vocabularies and is planning to further embed them into the platform. The next sections give an overview of these efforts.

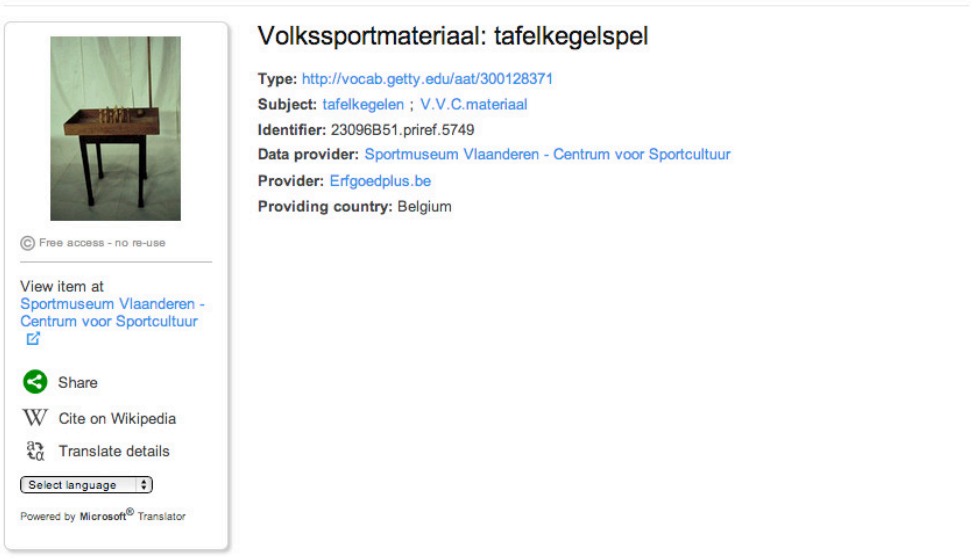
## 3.2 OpenSKOS

Waisda? is a crowdsourcing video tagging game which is one of the key applications of task *7.2 Development of innovative applications for user interaction and User-generated content*. As part of its prototypical development for use with Europeana data, an OpenSKOS[3] connector was implemented that enables the use of SKOS (Simple Knowledge Organization System) vocabularies as a basis for the tagging. User terms can now be matched to vocabulary terms that are in the OpenSKOS repository. The exploitation of these matches enables multilingual expansion of the user tags which improves retrieval across languages. One OpenSKOS instance[4] is deployed by Europeana 1914-1918[5], containing terms in seven languages which are linked to the Library of Congress Subject Headings[6]. In future, this will be used for Waisda? to leverage multilingual terms for tagging (D7.3 Report on innovative applications). Europeana is also planning to use the OpenSKOS repository for matching metadata values to controlled vocabularies. One result of this would be the exploitation of translation equivalents.

## 3.3 Adding multilingual vocabularies from providers

To enable cross-lingual retrieval, Europeana encourages providers to submit their targeted controlled vocabularies. For now, this mainly happens within projects that explicitly state that vocabularies are delivered to Europeana, but it should be one of the goals of Europeana to have all providers submit their thesauri, authority lists and classifications.

One of the main achievements is the exploitation of submitted links of the AAT (Getty Art & Architecture Thesaurus)[7]. For Europeana, it is now possible to fetch additional data from the vocabulary terms such as their translations. In the past, links to AAT could be only indexed and displayed as strings. Figure 15 shows such an object with a link to an AAT concept in the type field. The link is not dereferenced so the subject term as such cannot be used for retrieval.




[3] http://openskos.org/
[4] http://skos.europeana.eu/
[5] http://www.europeana1914-1918.eu/de
[6] http://id.loc.gov/authorities/subjects.html
[7] http://www.getty.edu/research/tools/vocabularies/aat/

*Figure 15: Before the change, the Link to AAT subject is in the metadata but not dereferenced so the term could not be used for retrieval.*

Now the AAT URIs are dereferenced at ingestion time enabling the indexing of additional multilingual and semantic data coming from AAT. This is also displayed in the platform and visible for all users. Figure 16 shows an object with dereferenced URIs from AAT. The subject “hourglass” is now retrievable in several languages more than the one provided.


*Figure 16: AAT terms and their translation displayed in the portal.*

For this to work, the providers need to resubmit their collections replacing the old AAT identifiers with the new URIs. Going forward, Europeana also wants to dereference links coming from providers to vocabularies such as GND[8], Iconclass[9] and VIAF[10].

[8] http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html
[9] http://www.iconclass.org/
[10] http://viaf.org/

# 4. Multilingual & Semantic Metadata Enrichment

Task 7.4.3, studied solutions on how to leverage multilingual semantic data and use it to improve Europeana object metadata and cross-lingual retrieval. Europeana enriches its object metadata fields (source) automatically with different vocabularies (target). Table 1 shows an overview of the source metadata fields and the target vocabulary for the enrichment types used in Europeana, namely places, agents, concepts and time periods.

| Enrichment type | Target vocabulary | Source metadata fields |
|---|---|---|
| Places | GeoNames[11] | dcterms:spatial, dc:coverage |
| Concepts | GEMET[12], DBpedia[13] | dc:subject, dc:type |
| Agents | DBpedia | dc:creator, dc:contributor |
| Time periods | Semium Time[14] | dc:date, dc:coverage, dcterms:temporal, edm:year |

*Table 1: Enrichments type, their target vocabulary and source fields.*

In 2012, we manually analyzed 200 enrichments by Europeana and categorized the different flaws and problems we encountered (Olensky et al., 2012). We identified three dimensions that influence the quality of enrichment: metadata level, vocabulary level, workflow level. Based on these findings, task 7.4.3. decided to create a task force on defining a strategy for multilingual and semantic enrichments in Europeana[15]. The task force ran from October 2013 to March 2014 and motivation and scope were determined in the task force charter[16]. On November 8, 2013, the task force members met for a workshop in Berlin to analyze six different datasets from Europeana and determine why enrichments did not work and how they can be improved. The following recommendations and findings were taken from the final task force report[17]:

Analyzing the datasets, it was found that enrichment flaws can be caused during one of the three stages that the metadata undergoes until it is displayed in Europeana:

1. Creation of the metadata by the provider.
2. Mapping to EDM.
3. Ingestion into Europeana.

During the process of enrichment itself, two choices are key to success or failure:

- the vocabulary used for enrichment, and
- the rules established for using the right terms for enrichment (on the target and the source side).

[11] http://www.geonames.org/
[12] http://www.eionet.europa.eu/gemet/
[13] http://dbpedia.org/About
[14] http://semium.org/time.html
[15] http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+force+multilingual+semantic+enrichment
[16] http://pro.europeana.eu/documents/468623/87545c7f-0f3c-4432-9a9b-c329e8b57ddd
[17] http://pro.europeana.eu/documents/468623/8b75b054-712e-432b-a0f7-761898e6f60e

The task force revealed the common obstacles in automatic semantic enrichment and listed its findings based on the 5 levels: metadata quality, mapping to EDM, ingestion, vocabulary choice and enrichment process. Table 2 lists the findings of the task force for the different levels.

| Issue | Findings |
|---|---|
| **Metadata quality** | • close collaboration with providers and institutions would improve metadata quality<br>• encourage the use of persistent, linked data URIs for vocabularies<br>• establish rules for field formatting<br>• feedback for flagging wrong metadata |
| **Mapping to EDM** | • more specific and targeted documentation highlighting common issues<br>• supporting tools for mapping that combines display of data, indexed fields and enriched fields<br>• metadata clinics for aggregators |
| **Ingestion** | • quality score at ingestion time to identify low quality metadata<br>• validation reports for providers to show them metadata quality issues<br>• metadata quality score threshold for executing enrichments, e.g. to ensure that fields are formatted correctly |
| **Vocabulary** | • encourage the delivery of vocabulary fitting the collection's context by the data provider<br>• exploit classifications of providers<br>• explore alignment of vocabularies and the exploitation thereof<br>• skip the broader terms in GEMET and do not use them for enrichments |
| **Enrichment process** | • establish enrichment rules for every field, e.g. pursuing basic splitting of values and document them well<br>• enrich all keywords within a field and do not stop enrichment after the first match in a field<br>• match the language of the metadata field (often, the language of the country of origin is sufficient) with the language of vocabulary |

*Table 2: Findings and outcomes of the task force on multilingual and semantic enrichment taken from the task force report.*

To better prevent flaws in the enrichment process, a staged model was developed that enables the identification of recurring problems in enrichments. Enrichment challenges occur at different phases in the workflow and should be identified and alleviated then. The next section details this model.

## 4.1 Staged model of multilingual metadata enrichment

Based on the findings of the task force, a step-by-step guide for adding enrichments to metadata was developed. It is crucial to ensure at each step that the resulting

enrichment is as beneficial as possible. Figure 17 shows this basic three-step model. We distinguish between three different stages in the enrichment process:

1. *Analysis*: the pre-enrichment phase focuses on the analysis of the metadata fields in the original resource descriptions, the selection of potential resources to be linked to and derives rules to match and link the original fields to the contextual resource.
2. *Linking*: the process of automatically matching the values of the metadata fields to values of the contextual resources and adding contextual links (whose values are most often based on equivalent relationships) to the dataset.
3. *Augmentation*: the process of selecting the values from the contextual resource to be added to the original object description. This might not only include (multilingual) synonyms of concepts to be enriched but also further information, for example broader or narrower terms.

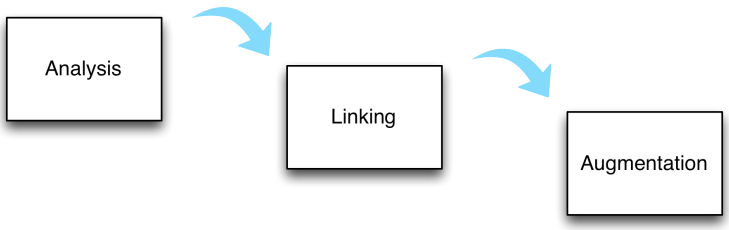



*Figure 17: Basic model of enrichment stages.*

The staged model of enrichment is developed to ensure that at each step, quality controls ensure that automatic enrichments are executed correctly.  shows the model augmented by the different steps that need to be taken into account to ensure a smooth process.

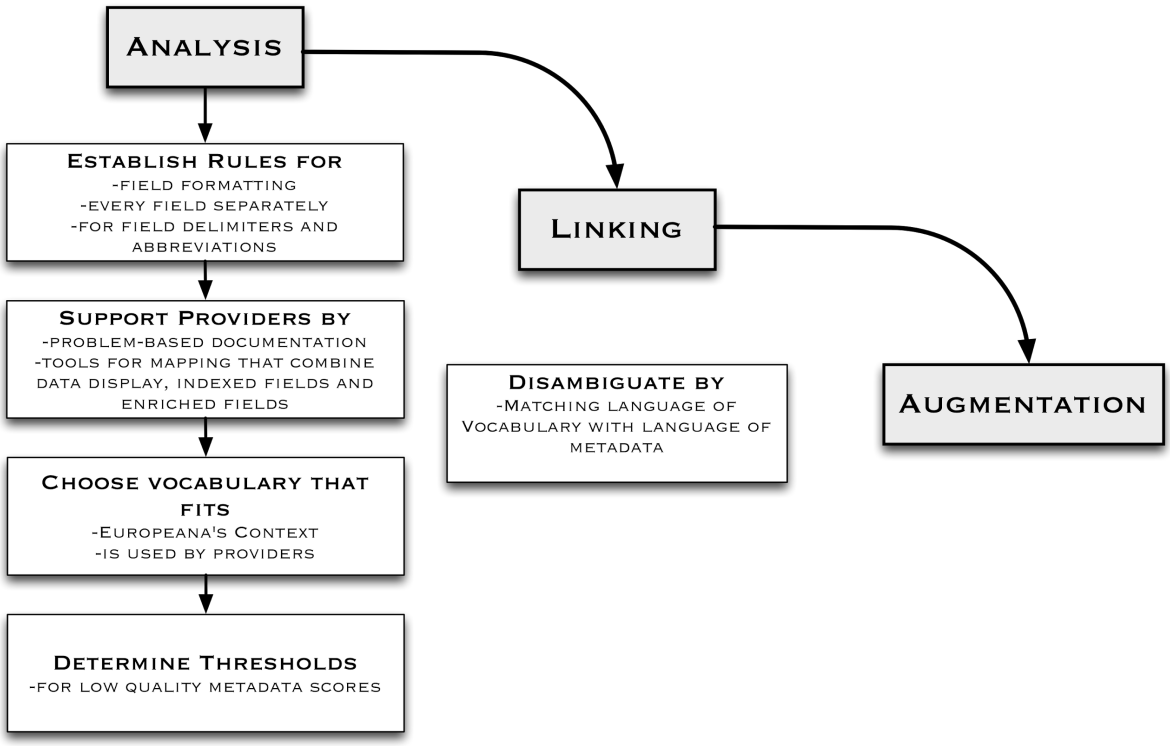



*Figure 18: Model of enrichment stages.*

This model shows that the most important part of the enrichment process is in executing a thorough analysis of the source data and target fields. The more effort is put into the first stage the better the enrichments become. Similarly, any wrong decision made in the analysis stage will affect the Linking and augmentation stages as they are built upon another.

## 4.2 Framework and measures to evaluate enrichments and their effectiveness

Evaluating and measuring the effects of automatically enriching metadata is important to ensure that highly curated metadata is represented correctly. The goal of enrichments is not only to link datasets and resources of Europeana but also to enable high-performance retrieval across languages and contexts. The evaluation of enrichments targets the measurement of their effectiveness or their potential harm when enrichments were added incorrectly.

Once the enrichment process is fully executed going through the stages analysis, linking and augmentation, the effectiveness of enrichments can be evaluated. Here, the focus can be either on the enrichments or their influence on the retrieval performance. The latter is similar to an end-to-end evaluation that focuses on the outcomes of enrichments from an objective point of view. Not the process itself is important but how and if it impacts user satisfaction, so the entire process - from one end, the query, to another, the result list - is evaluated (van Hage, 2007).

For the enrichments, one could determine how often objects were enriched and how good these enrichments were. The retrieval performance measures the impact of enrichments on the quality of the search results.

To evaluate enrichments and their effectiveness, the objects (a set of objects) and a list of queries are the basis to measure the frequency, coverage, quality or relevance of the enrichment. Table 3 gives an overview of the measures one can take to evaluate enrichments. Looking at the objects or the queries, several different measures can be carried out that target different dimensions of the enrichment process. The frequency provides information about the quantitative aspects of the enrichments, the distribution and coverage gives a fuller account of these quantitative numbers. The quality looks at the accuracy of the enrichments counting the wrong enrichments, whereas the relevance measures has several degrees and looks at the influence of enrichments on retrieval and findability.

| Evaluation based on type of measurement | Objects | Queries |
|---|---|---|
| **Frequency** | Number of enriched objects, enrichments per objects | Number of queries that retrieve enriched objects |
| **Distribution/Coverage** | Proportion of enriched objects<br>Distribution of enrichments across facets | Proportion of enriched objects (set)<br>Percentage of queries retrieving enrichment across facets |
| **Quality** | Percentage of wrong enrichments and percentage of objects with wrong enrichments | Percentage of wrong enrichments in result set per query |
| **Relevance** | Relevance of enrichment to the object | Relevance of enrichments to queries |

*Table 3. Framework to guide enrichment evaluations.*

To look at relevance of enrichments related to queries, the categories described in Table 4 can be used.

| Relevance categories | Description |
|---|---|
| **Retrieved query match** | The record was found due to the enrichment, i.e. the query term is only present in the auto-generated tags (enrichments) but not in the original metadata. In these cases the query is a translation, transliteration or a broader term of an original metadata term. |
| **Query match** | Enrichment is relevant to the query, but the object has not necessarily been found because of the enrichment, as the query is also part of the original metadata. |
| **Broader term match** | Enrichment is relevant to the query, because it is a narrower term of the query, but had no influence on the retrieval of the record. |
| **Partly query match** | Enrichment is relevant to only a part of the query or only part of the enrichment is relevant to the query. |
| **Query independence** | Enrichment is independent from the query, but a correct enrichment for the record. |
| **Wrong enrichment** | Enrichment is independent from the query and is incorrect. |

*Table 4: Relevance categories and their descriptions.*

In the next section, an evaluation was conducted based on a list of queries and their results lists.

## 4.3 Evaluation

In this section, we report on an evaluation of enrichments in Europeana and its results.

For the enrichment evaluation, we extracted the 1,000 most frequent queries in Europeana from Google Analytics[18] for the first quarter of 2014. After cleaning the queries[19], we randomly chose 100 queries for the evaluation. The 100 selected queries were then searched in Europeana and the top 12 results were documented for each query. A total number of 1,121 records was assessed for their enrichments. We manually checked these records for enrichments that were produced by Europeana coming from GEMET Thesaurus (WHAT), Geonames (WHERE), Semium

[18] http://www.google.com/analytics/

[19] Queries with Boolean operators and wildcards were removed.

(WHEN) and DBpedia (WHO and WHAT). Enrichments to these vocabularies that were done by the data provider and are therefore part of the original metadata were not evaluated.
In total, these 1,121 records were enriched 1,083 times: 53% of the enrichments come from the WHEN-facet, 28% from the WHAT-facet, 16% from the WHERE-facet and 3% from the WHO-facet (Figure 19).

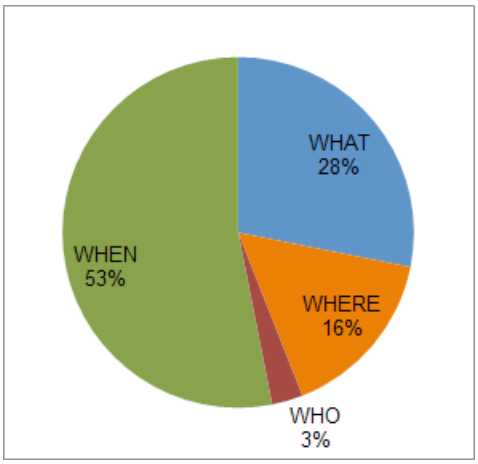



*Figure 19: Distribution of enrichments over facets.*

38% (424 records) of those 1,121 records were enriched with at least one enrichment. 73% of all queries had at least one record in the first 12 results that contained at least one enrichment.
Of the 424 enriched records, 72% were enriched with terms in only one facet, 26% in two facets and 2% in three facets. 51% of all enriched records were enriched with two terms, followed by 19% with four enrichment terms and 17% with one.
The relevance assessment was carried out only for the WHAT-, WHO- and WHERE-facet as we had decided to exclude numbers from the queries and no textual query was categorized as time period. For those three facets, a total number of 508 enrichments was assessed for their relevance to the query.
In Table 5 the relevance categories are listed with examples and the codes used for this evaluation.

| Relevance category | Example | Code |
|---|---|---|
| **Retrieved query match** | The query *primera guerra mundial* (First World War in Spanish), which retrieved a number of records that only had *I Guerra Mundial* in their original metadata and were retrieved because of the enrichment with the concept World War I (and the respective translations) from DBpedia[20]. | R+++ |
| **Query match** | The query *Lilien, Ephraim* retrieved enriched records for result records 6 to 12. All of these seven records were enriched with the person Ephraim Moses Lilien[21] which makes the enrichments relevant to the query. Yet, all of them also had the person Ephraim Moses Lilien as creator | R++ |

[20] http://dbpedia.org/page/World_War_I
[21] http://dbpedia.org/resource/Ephraim_Moses_Lilien

| | | |
|---|---|---|
| | in their original metadata. Therefore, it seems unlikely that those records were retrieved because of the enrichments. Nevertheless, these enrichments are very useful and most importantly correct. | |
| **Broader term match** | Most examples come from WHERE-enrichments. In these cases, the records have been enriched with a geographic location and as the enrichment process also adds broader terms, a record enriched with *Bratislava* was also enriched with *Slovakia*, or a record enriched with *Melilla* was also enriched with *Spain*. Again, the enrichment is correct and definitely useful but it did not influence the retrieval of the specific records. | R+ |
| **Partly query match** | This category was applied to result records from two queries. One was the German query *sport zeitung* where all of the retrieved 12 result records were enriched with the concept *sports* (and as broader term *recreation*). Therefore, the enrichment is only relevant to part of the query. The inverted case where only part of the enrichment is relevant to the query is attributed to results retrieved by the query *Napoleon*. Result records 5 to 12 were all enriched with *Napoleon Orda*, who is a Polish artist. Therefore, only the first name is relevant to the query. In this particular case, it seems most likely that the users were looking for records associated with *Napoleon Bonaparte*, as he is usually referred to as solely *Napoleon*. Nevertheless, the retrieved records were enriched with the correct person. In both cases, the records also contained the query term in the original metadata. | R |
| **Query independence** | These are correct enrichments but not relevant to the query, i.e. the enrichment-facet does not correspond with the category of the query. For instance, a query for a person retrieves result records that were enriched with a time period, concept or geographic location: the query *Lerski, Helmar* (who was a Swiss photographer) retrieved records that were enriched with *documentation* and *photograph* or *industrial process* and *photography*. Even though it is debatable how meaningful and adequate these enrichments are (as they are quite generic), they are correct enrichments. However, they are only indirectly related to the query and have definitely not influenced the ranking and retrieval of the result records. | R- |
| **Wrong enrichment** | These enrichments are not relevant to the query, analogously to the previous category R-. Yet, in addition, they are incorrect. These enrichment flaws can have different reasons. All incorrect enrichments are separately listed in the subsection *Enrichment errors discovered during the evaluation*. | R-- |

*Table 5: Relevance categories with examples and codes.*

### *4.3.1 Results of the relevance assessment*

The overall results across facets show that enrichments from category R- (correct, but not relevant to the query) are the most frequent ones with 46%, followed by enrichments from category R++ (correctly enriched but the query also occurs in the original metadata) with 20%. Unfortunately, incorrect enrichments (R--) come in third place with 16%. The enrichments that match the query account for 22% (R+++ and R++) of the cases and potential improve multilingual retrieval. 2% of the cases belong to the relevance category “retrieved query match” - the objects that were solely found due to the enrichments (Figure 20).
When summarizing the categories according to their relevance to the query, then 38% of all enrichments were in some way relevant to the query term(s) and 62% were irrelevant to the query and therefore to the retrieval of results.

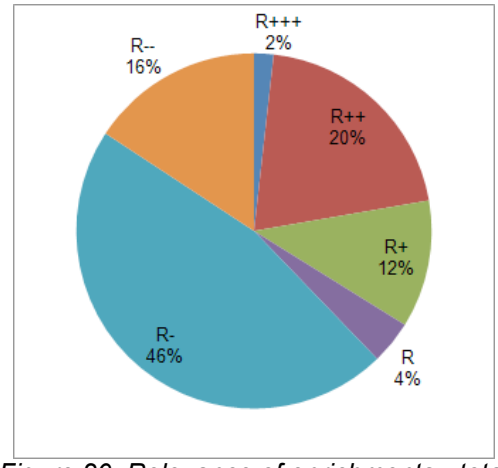



*Figure 20: Relevance of enrichments - total*

Evaluating the results according to the different facets, it can be observed that the enrichments truly relevant to the query only occur in the WHERE and the WHAT facet with a small share (2% each within the facet) compared to the other enrichment categories (cf. Fig). In the WHO facet, the enrichments, if relevant to the query, also were present in the metadata. The main reason for this is that most persons (or person names) are named entities and therefore no translations exist. Consequently, enrichments of persons can be more interesting in a semantic way (if the link to the external resource also provides information about related persons) but not so much in a multilingual way. In our case, the persons were enriched with persons from Dbpedia which do not provide (machine-readable) relations to other persons (such as: ex:wasTeacherof or ex:wasMarriedto).
In the WHERE-facet, 65% of all enrichments were correct and in some way related to the query (R+++, R++, R+), whereas this share is much lower in the WHAT-facet (17%). The share of irrelevant enrichments (R- and R--: 83%) is the highest in the WHAT-facet. Also, incorrect enrichments are more frequent in the WHAT-facet (22%) than in the WHERE-facet (8%) (Figure 21).

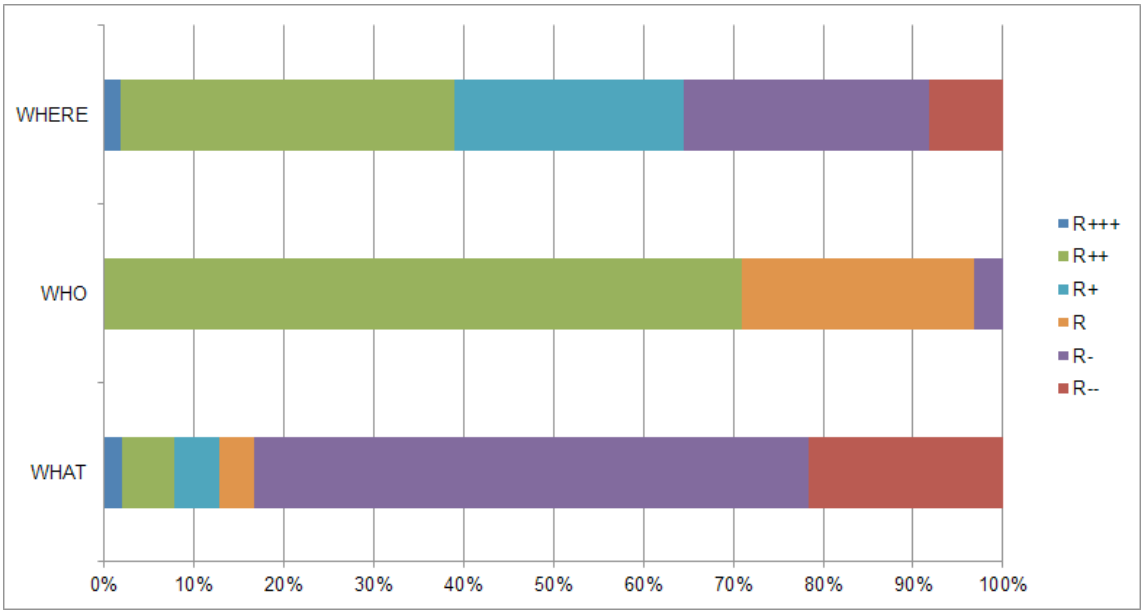



*Figure 21. Enrichment relevance per facet*

The results of the relevance assessment show that Europeana should focus the enrichment evaluations on the WHERE- and WHAT-facet. As mentioned before, the enrichments in category R- are correct enrichments that are irrelevant to the query. In the majority of cases, these enrich generic terms from the subject or type field (e.g. *paper* plus broader term *industrial product* or *photograph* plus broader term *documentation*). These enrichments are correct and definitely contribute to the multilingual enhancement of metadata. Yet, they can also create a lot of noise and increase recall at the expense of reduced precision. For instance, if a user searches for the term *documentation* in Europeana, they will have different ideas of what they would expect as result. For one user, it might be useful to get every object of the type *document*, *photograph*, *CD-ROM*, *film*, *video* (all narrower terms in the GEMET) as a result for this query. For another user, this increased recall might be annoying. It is a difficult balance between correct and useful, correct as well as incorrect enrichments. However, as these enrichments do not retrieve completely irrelevant records, they are not as harmful as incorrect enrichments.

**Analysis of wrong enrichments**

The category "wrong enrichment" only looked for the relevancy of the enrichment with regard to the query but did not take into account whether the object found was relevant to the query. Table 6 shows the distribution of wrong enrichment across objects and queries for the facet.

| | # of wrong enrichments | # of objects with wrong enrichments | # of queries that retrieved wrong enrichment |
|---|---|---|---|
| **WHAT-facet** | 66 | 32 | 6 |
| **WHERE-facet** | 14 | 7 | 1 |

*Table 6: distribution of wrong enrichments (R--) over objects and queries.*

Table 7 shows whether the wrong enrichment influenced the retrieval of this particular object. The abbreviations in the table mean the following:
MA – Multilingual ambiguity, the number behind it indicates the error listed in 4.3.2.

SA – Semantic ambiguity
GA – Geographic ambiguity

| Query | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Arikha, Avigdor** | | MA2 | MA2 | MA2 | MA2 | MA2 | | | | | | |
| **Gersht, Ori** | | SA | | | | | | | | | | |
| **Internet** | | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 | MA4 |
| **Mauclerc** | | | | | | | | | | | | MA |
| **Nes, Adi** | | | MA2 | MA2 | MA2 | | | | | | | |
| **Rubin, Reuven** | | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 | MA2 |
| **Lilien, Ephraim** | | | | | | GA | GA | GA | GA | GA | GA | GA |

*Table 7: Queries and their 12 results indicating the type of error and whether it influenced retrieval (red) or not (green).*

The green fields indicate whether the object was still relevant to the query. Red field indicate that the object was not relevant to the query, meaning it was ranking incorrectly due to the wrong enrichment. This is an indication on how much influence wrong enrichments have on retrieval. It should be noted that only a small fraction of queries was used on this analysis.

### *4.3.2 Enrichment errors discovered during the evaluation*

This subsection lists the enrichment errors discovered during the evaluation process. All of these issues were reported to the Europeana Office. One issue (No 5) was already known to Europeana. All of these enrichment errors can be traced back to cross-lingual ambiguities.

1. *Kiri* means *letter* in Estonian but also *lease* in Maltese. Therefore, records that contain the term *kiri* in the subject field have been enriched with *housing legislation* and *lease*. This error does not seem to affect many records, but illustrates multilingual disambiguation challenges. Example record: http://europeana.eu/portal/record/92097/1195ECF7CAA361BA8DE3A6B3628458E9202CCF9D.html
2. German records with the term *Tür* (door) in their metadata have been enriched with the concepts *ecological parameter* and *species* from the GEMET. The term *Tür* is a cross-lingual homonym and means *species* in Turkish. This enrichment error is caused by cross-lingual ambiguity and the use of a non-domain specific vocabulary. Example record: http://europeana.eu/portal/record/08535/ED150C17BEB8579DDA9CD7A7EF592A5965070BD2.html The same happens as “art” is in german a Homonym meaning art and species. As the language is not identified as English the subject “art” is matched to the German “art (spezies)”
3. The term *art* does not only refer to artistry in English but is also a cross-lingual homonym and means *species* in German, Norwegian, Danish. Analogously to example 2, records that have *art* as subject have also been enriched with *ecological parameter* and *species*. When we shortly tested those two concepts as queries, the query *species* seems to retrieve truly relevant records on the first result pages, whereas the query *ecological parameter* ranks example records from the faulty enrichments first. Example record:

http://europeana.eu/portal/record/08533/1A93A228E13E1DCAF04B163042E9A5306DB53EF9.html

4. Records that have the term *Forum* as type in their metadata were enriched with *Internet* and *newsgroup* because of the French translation of *newsgroup* to *forum* in the GEMET. Again, this is a problem of cross-lingual homonyms. Therefore, the query *newsgroup* not only ranks records that are truly related to *newsgroups* and *internet* but also many records related to *Forum Romanum*, *Trajan's Forum* or any other architectural forum. A browse through the result records showed that not all of the enrichments are incorrect, but still the term *forum* can have different meanings in different languages (e.g. also *panel* is translated to *Forum* in German) and is not necessarily related to *newsgroups* and *internet*. Example record: http://europeana.eu/portal/record/08501/508B948646942BAA52C6BA56A8EE1F0CE16CE48C.html
5. All records with the subject term *drawing* (with the meaning *painting*) have been enriched with the concept *drawing* from the GEMET[22], which actually means *sampling* (i.e. to draw a sample). This enrichment error is a good example for problems that arise when a non-domain specific vocabulary without any restrictions is used for the enrichment (Olensky, Stiller & Dröge, 2012). We consider it rather unlikely that users will look for translations of this enrichment (e.g. German: Probenahme or French: prélèvement) and as a result get lots of records that are drawings. Still, this should be corrected. Example record: http://europeana.eu/portal/record/08502/EE19616B7AEF6A9073EAAB4E9FB8CCDA7664C0FE.html

### *4.3.3 Summary*

To summarize the results of this evaluation, the framework is used. Table 8 gives the numbers for each evaluation point for the quantitative assessment on all enrichments.

| Type of measurement | 1121 objects | 100 queries |
|---|---|---|
| **Frequency** | 38% of objects enriched | 75% of queries retrieve enriched objects |
| | 2.5 enrichments per enriched objects | In average 50% of the objects are enriched per query (from queries that have enriched objects) |
| **Distribution/Coverage** | Distribution of enrichments across facets: 53% When, 28% What, 16% Where, 3% Who facet | Percentage of queries retrieving enrichment facets: 62% When, 33% What, 21% Where, 6% Who |

*Table 8: Descriptive statistics on all objects and their enrichments.*

Although the number of objects that have an enrichment seems to be low (38%), a user is still likely to encounter enriched objects. 75% of the queries retrieve objects that were enriched. On average, 6 objects out of the 12 results are enriched per query. Looking at the different facets, most of the enrichments come from the WHEN-

[22] http://www.eionet.europa.eu/gemet/concept/11812

facets followed by the WHAT-facet. This distribution is similarly also reflected in the queries that retrieve the facets.

As the evaluation was only conducted for the WHAT-, WHO- and WHERE-facet, the number of enrichments decreases to 508 evaluated enrichments. Table 9 shows the values for quality and enrichment.

| Type of Measurement | 1121 objects | 100 queries |
|---|---|---|
| **Quality** | 16% of enrichments are incorrect | 8% of queries with wrong enrichments in results set |
| | 3.6% of objects with wrong enrichments | |
| **Relevance** | 84% of enrichments are relevant to the object | 34% of enrichments can be mapped to queries |

*Table 9: Quality and relevance of enrichments of the WHAT-, WHO-, WHEN-Facet.*

16% of the enrichments were wrong, resulting in 3.6% of the objects being enriched incorrectly. This seems little but it needs to be noted that it still applies to 8% of the queries. Given that these queries are the most used queries of Europeana, the exposure of wrong enrichments might still result in many unsatisfactory results for users if the enrichment was relevant to the query. The queries that retrieve objects with incorrect enrichments often retrieve many of them. These 8 queries retrieved in average 10 results with wrong enrichments. This means that on the first result page, the user might not find a single relevant result. This is the case for the query [Internet] were 11 out of 12 results were retrieved due to a semantic ambiguity ("forum" referring to newsgroup in GEMET). On the other hand, there are wrong enrichments that did not influence the quality of the search results. For example, the query [Rubin, Reuven] retrieved 11 out of 12 objects with wrong enrichments. Although the enrichments are wrong, the objects were still relevant to the query as the query matched the value of the creator field. Solving this problem will lead to a considerable improvement of enrichment quality and relevance.

Looking at the relevance of enrichments, we see that in the analyzed facets the majority of enrichments is relevant to the object (84%). 34% of the enrichments are evaluated as relevant to the query. This means that although enrichments are mainly correct, their influence on the retrieval performance for the evaluation sample are small. It should be noted that queries and datasets often occur in frequent languages such as German, English and French and that these queries also retrieve datasets in these languages. Objects which were solely found because of enrichment often tend to be from datasets in underrepresented languages. That means for example that it is more likely to retrieve an English object with an English query, than a Lithuanian object with an English query based on enrichment. The nature of this sample features more queries in frequent languages, so it is less likely to retrieve objects that were solely found because of the enrichment. To fully understand the bias introduced by frequent query languages and the proportion of datasets in these languages, more studies are needed.
Additionally, enrichments are not only improving retrieval but also link objects across languages and collections which can then be used for browsing these relations.

## 4.4 Results and Recommendations

The task force on multilingual and semantic enrichment already produced many recommendations that can be implemented by Europeana. Most of these recommendations target solely the improvement of enrichments and might be less feasible for Europeana to implement. Therefore, this section details some actionable items for Europeana.

**Tackle multilingual ambiguities**

For the automatic enrichments, one of the most pressing issues is the match of the vocabulary language with the field value language in Europeana. Disregarding the language while matching leads to wrong enrichments. In the past, Europeana removed these critical terms from the enrichment process but as the evaluation has shown, this is still an issue and cannot be solved case by case but should be avoided preventively. One of the measures is either to determine the language of the to-be-matched field or to determine the language by the field “language of description”.

**Improve documentation**

Although documentation is thorough, it is still criticized by providers that mapping guidelines are hard to understand and the process of automatic enrichments is often not clear and transparent. Especially in this regard, Europeana should answer a couple of questions which make it clearer to the provider how their data might be influenced by enrichments in the portal. The task force on metadata quality determined some simple questions which Europeana should provide documentation on[23]:

- How to populate required EDM fields (e.g. dc:title) in the most meaningful way, when they do not exist in the original metadata for individual objects?
- How should the Dublin Core fields of an original 'one-size-fits-all' record be distributed among various EDM classes (e.g. ProvidedCHO, WebResource)?
- When/how should two separate fields in the original metadata be mapped into one value in EDM?
- When/how should multi-valued, single fields be separated into distinct fields in EDM?
- How should date intervals be mapped to proper time spans and not individual dates?
- How should persistent identifiers, e.g. identifying vocabulary terms, be handled during mappings?
- How to provide explicit and persistent links/URIs as metadata values (cf. previous section)?
- Is it appropriate for mappings to assign one value for a field (e.g. subject) over an entire dataset?
- How to adapt the degree of granularity of the metadata through the mapping (i.e. dcterms:created over dc:date)?

**Define quality thresholds for metadata**

The quality of enrichments is greatly determined by the quality of metadata. Therefore automatic quality checks for metadata sets should be determined. This does not only improve the enrichments but also the overall quality of the data taking

[23] Taken from task force report: http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+force+multilingual+semantic+enrichment

away the burden of Europeana to make manual spot checks. To determine this threshold and come up with good guidelines could be part of a new task force.

**Exploit provider's classifications and controlled vocabularies**
One of the main findings of the task force was that none of the automatic enrichments Europeana provides is as good as the targeted vocabularies and classification of the providers. Europeana needs a strategy that can better exploit this rich data and enables providers to submit it.

# 5. Conclusion and Future Work

This deliverable reported on the work of task 7.4 Multilingual Access and on the collaborative work with regard to evaluating enrichment with task 7.3. It delivered mockups for multilingual search and retrieval scenarios that can be used by Europeana.
During the course of this project, Europeana made a lot of progress with regards to multilinguality. This is not only true for multilingual display and features that let users retrieve content in languages they might not know but also technical solutions that improve cross-lingual retrieval. Europeana is a trailblazer in enriching cultural collections with controlled vocabularies to enhance multilingual retrieval. The work done in this task enforced this position and enhanced multilingual access and display for users with different language skills.

Building on this progress, Europeana will continue to work on multilingual enrichments, display and the improvement of retrieval across languages. Especially, the language-aware ranking and provision of content based on user-preferences should be explored more. The risk is that relevant content gets hidden just because it is not in the preferred language. A balance needs to be found striving for content users can understand on the one hand and the possibility to stumble across the unexpected and allow for serendipity on the other hand.
Additionally, with regard to embedding external multilingual vocabulary, Europeana is still in an early stage of development given how much more coverage and granularity these efforts could gain. One of the requirements for further developments in this direction is the quantitative and qualitative analysis of the metadata Europeana is ingesting. Finding meaningful measure to determine metadata quality will result in better multilingual services which are augmented on this data. This is not only limited to enrichments but also relates to ranking factors and the display of multilingual data.

# References

M. Olensky, J. Stiller, and E. Dröge (2012): Poisonous India or the Importance of a Semantic and Multilingual Enrichment Strategy In: Metadata and Semantics Research, ed. by Dodero, J., Palomo-Duarte, M., Karampiperis, P., Springer, Berlin.

J. Stiller, M. Gäde, V. Petras (2013): Multilingual Access to Digital Libraries: The Europeana Use Case/Mehrsprachiger Zugang zu Digitalen Bibliotheken: Europeana/Accès multilingue aux bibliothèques numériques: Le cas d'Europeana. Information-Wissenschaft & Praxis 64 (2-3), 86-95

Van Hage, W. R., Isaac, A., & Aleksovski, Z. (2007, November). Sample Evaluation of Ontology-Matching Systems. In *EON* (Vol. 2007, pp. 41-50).