ICT-PSP Project no. 297158

EUROPEANAPHOTOGRAPHY

EUROPEAN Ancient PHOTOgraphic vintaGe repositoRies of digitAized Pictures of Historical qualitY

Starting date: 1st February 2012

Ending date: 31st January 2015

Deliverable Number:	D 4.3.1
Title of the Deliverable:	Enrichment Report (first release)
Dissemination Level:	Public

Contractual Date of Delivery to EC:	Month 24
Actual Date of Delivery to EC:	January 2014

Project Coordinator

Company name : KU Leuven
Name of representative : Fred Truyen
Address : Blijde-Inkomststraat 21 B-3000 Leuven PB 3301
Phone number : +32 16 325005
E-mail : fred.truyen@kuleuven.be
Project WEB site address : http://www.europeana-photography.eu

Context

WP 4	Indexing and multilingual support
WP Leader	KMKG
Task 4.5	Metadata multilingual enhancement
Task Leader	Promoter
Dependencies	

Author(s)	Valentina Bachi, Elisa Debernardi, Antonella Fresa (Promoter)
Contributor(s)	All partners
Reviewers	Nacha Van Steen (KMKG), Nikos Simou (NTUA), Sofie Taes (KU Leuven)
Approved by:	All consortium

History

Version	Date	Author	Comments
0.1	15/10/2013	Valentina Bachi, Antonella Fresa (Promoter)	
0.2	14/01/2014	Valentina Bachi (Promoter)	Updated according to KMKG initial comments and further developments of the enrichment process
0.3	24/01/2014	Valentina Bachi (Promoter)	Updated according to NTUA's review and added statistics from MINT
1.0	29/01/2014	Valentina Bachi, Elisa Debernardi, Antonella Fresa (Promoter)	Final version including the latest comments from the reviewers

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

The enrichment process is the main task of WP4. The main activities for the content providers were the creation/update of local catalogues and the use of a multilingual vocabulary, expressly designed for EuropeanaPhotography, that allows partner to enrich the metadata with translation in several European languages.

Beside the local cataloguing activity, that is an activity to be carried on in the databases of each content provider, the role of the multilingual enrichment and the vocabulary is particularly important once the databases are imported into the software environment in use for the project (MINT), and it is supported by a specific Multilingual mapping tool..

So far, the metadata production process produced over 219.000 records new records available for Europeana, corresponding to 612.000 multilingual concepts, over 8.000 new items labelled with Public Domain Mark and over 38.000 labelled under CC-BY-NC-ND license.

The result of the ingestion periods December 2013 and January 2014 is currently over 210.000 new Europeana records, thus reaching and exceeding the project's Indicators n. 2, 3, 4 and 5:

2	No of items enriched	Enrichment	200,000
3	No of multilingual items	Multilingual support	200,000
4	No of Europeana compliant records	Transformation	150,000
5	No of items on Europeana	Ingestion process	100,000

Providing rich and multilingual metadata produces benefits for the whole digital community: for the project as a whole, for the single content providers, for Europeana, and certainly for the final end-users of Europeana's portal. How these enriched multilingual metadata look in Europeana is extremely important for the actual delivery of such benefits; therefore, the visual impact of the records in Europeana needs to be user-friendly, attractive and easy-to-understand.

Three elements are the basis of the enrichment process as intended in EuropeanaPhotography: the mandatory metadata fields that were set for the project, the local catalogue enrichment and the Multilingual vocabulary. It is interesting to note that all this work, that for the content providers can be intense and time-consuming, is not simply "lost" in the cyberspace once Europeana has ingested the metadata in its portal: the content provider is able to get back his metadata through a procedure that allows to enrich the local database too.

The content providers found sometimes this task of enrichment an hard job, especially when big efforts were necessary at the local catalogue level, but they are all happy with the experience and with the results, as emerges from the overall analysis of a questionnaire about the enrichment process that was circulated in the consortium for producing this deliverable.

The first batch of metadata was delivered to Europeana in November 2013: the records were processed according the initial set-up of the MINT mapping tool and Vocabulary.

The further batches of metadata will be processed according to an enhanced set-up of the tool, that includes an advanced version of the Vocabulary with 3 further languages: Russian, Ukrainian and Chinese. Moreover, at the moment we are exploring the integration of additional vocabularies like Geonames.

2   INTRODUCTION

2.1   THE ENRICHMENT PROCESS IN EUROPEANAPHOTOGRAPHY

This document intends to describe the rationale and the result of the enrichment process, which is the main task of WP4 in EuropeanaPhotography..

The enrichment as intended in EuropeanaPhotography project is a synonym for "indexing". This process adds structured information to the source metadata records. The information related to a photograph can be more or less wide, also depending on the procedures and the requirements of each content provider.

The content providers discussed a little about which of their activities can be considered as part of the enrichment, as there is not a strict definition of "enrichment" in informatics, and there is an "enrichment" phase in the project.

One thing, common for all partners in the project, is the mapping of the keywords. Everybody maps the keywords, used in the metadata of their institution, to the keywords of the EuropeanaPhotography vocabulary. Some partners do this using the MINT tool while others prefer doing this outside MINT environment by adding the vocabulary URI's directly in their metadata, but everybody does this. And it can be considered as "enrichment" - to add something to the metadata: in particular it becomes searchable in many languages, although partners not enhance the metadata to every single image.

In addition to this, in few cases the partners edited the metadata of every single image in the part of keywords in their own repository. For example, NALIS did not have a local controlled vocabulary in the beginning of the project, and they gradually built one during the first year. And then NALIS edited what was already locally ingested, so that to use as keywords only those from the local controlled vocabulary.

This allowed NALIS to map their local vocabulary to the EuropeanaPhotography vocabulary and to upload to MINT their metadata in a complete form - already containing links to the EuropeanaPhotography vocabulary. This enhancement of every single piece of metadata is actually another "enrichment" activity.

Pure academically - the term "enrichment" is often used when to an existing piece of metadata people add links to similar or related objects (often outside their own collection). For example, a photo of an actor, may be added with links to posters for plays where he acted etc.

Enrichment process in EuropeanaPhotography project takes place at two stages:

- At the time of the local cataloguing, to be done "manually" by each content provider in his own archive, before importing it into the MINT mapping tool

- At the time of the mapping, to be done through the MINT tool and in particular the Multilingual Vocabulary that allows to implement multilingual keywords;

Local cataloguing

The metadata related to each photograph are of different kinds: a first level includes the curatorial/cultural information related to the photograph, expressed in already existing catalogues in the data bases of the content holders or in a new catalogue if no one is available.

A second level is composed by the digitization metadata, such as file formats, file sizes, equipment used, standards complied with, colour depth, etc; this information is important as well as the cultural information, in the light of publishing and re-use of the digital content, and has to be added in the local catalogue records after the digitization is completed.

Another level of information that can exist in the local catalogue is represented by the use of authority files. These can refer to keywords and other particular categories of information, such as: events, locations and individuals.

Each content holder should have implemented its own repository with the metadata related to the three levels illustrated above. If, for example, one repository is not using a controlled vocabulary for its keywords, it will be necessary to implement this before going to MINT.

At this point, the local catalogue records are ready to be imported into MINT.

MINT mapping

The MINT mapping process supports further "enrichment" for the metadata: the multilingualism.

The mapping activity of local metadata to the intermediate schema includes a few steps that allow the content provider to apply the multilingual vocabulary to the metadata, so that multilingual metadata are added.

At this point, the enrichment process is completed and the metadata are ready for transformation to EDM, publication in NTUA's OAI-PHM server and then for harvesting by Europeana team.

2.2   POSITIVE EFFECTS OF THE ENRICHMENT PROCESS

From the point of view of EuropeanaPhotography project

The enrichment process and in particular the multilingual vocabulary have the positive effect to uniform the metadata provided to Europeana according to mandatory fields for all the content providers. The Vocabulary organizes the digital content among thematic lines; furthermore, the Vocabulary can be used to support dissemination purposes, by creating for example thematic views of the content on the EuropeanaPhotography website.

The main contribution of providing multilingual metadata to Europeana is clearly focused on the end-user experience. In detail, having multilingual metadata allows users to consult all collections at once with the use of keywords in their own language, making in that way the discovery of records much easier.

From the point of view of Europeana

The importance of enriched metadata for Europeana is highlighted in the following paragraph.

"Metadata is the only thing we have to make an object discoverable. But we still have quite a few objects in Europeana that have no free-format description at all (a huge 8.1 million objects), or have no title (nearly 1 million objects). As the primary means of searching in Europeana is a free text search, it is very unlikely that these objects will show up in any searches", writes Jan Molendijk, Europeana's Technical & Operations Director, on the EuropeanaPro blog[1].

Implementing a remote, multi-lingual, cross-domain searching as the Europeana rationale implies, requires high quality metadata also to guarantee the interoperability of the collections and the quality of all search results. The big effort of Europeana Team to engage the users in a rich and stimulating discovery experience is highly enhanced and fostered by the EuropeanaPhotography metadata.
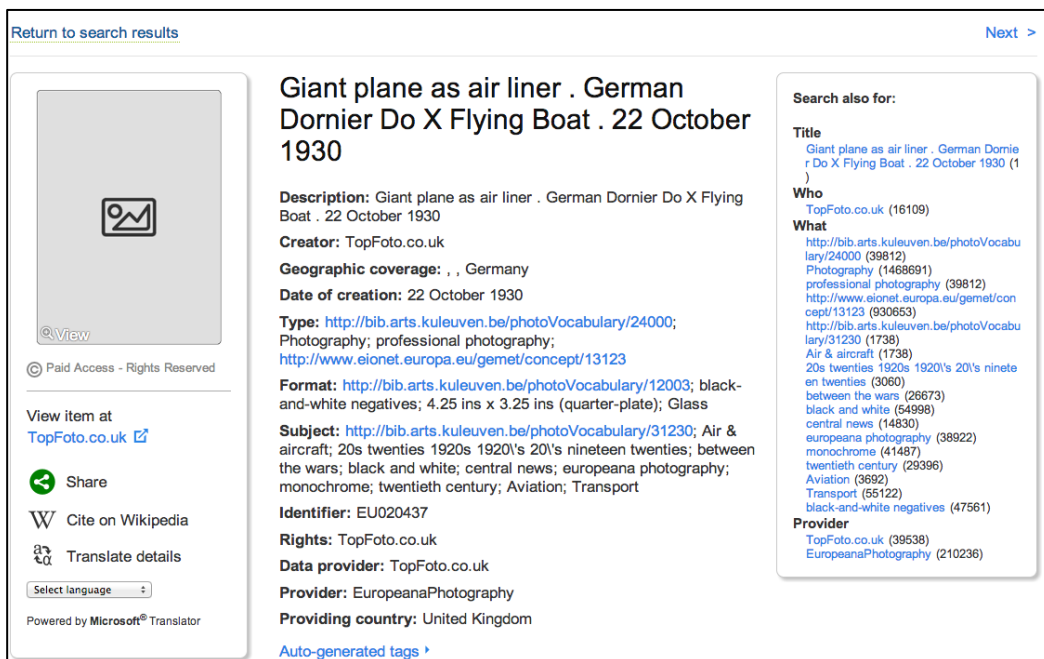
The result of EuropeanaPhotography enrichment process is to provide to Europeana very valuable metadata, describing each photograph with rich information available in 15 languages so that:

- The end-user will easily search and found items of his/her interest by using English language or his/her own language;

- the end-user will then enjoy the digital cultural content in his/her own language, gaining wide and easily-understandable information about the photograph he/she is looking at.

It is important however to note at this point that, despite the fact that EuropeanaPhotography provides rich and multilingual metadata to Europeana, the way that this information is displayed on Europeana's portal is unfortunately not very attractive. See image in the next page.

In detail the URIs of the vocabulary terms appear together with the metadata, fact that does not look nice because the observer is not interested in the URI but on its label. For seeing the labels of the term in all the languages, the viewer needs to expand to "Auto-generated tags" by clicking on the small link under the metadata
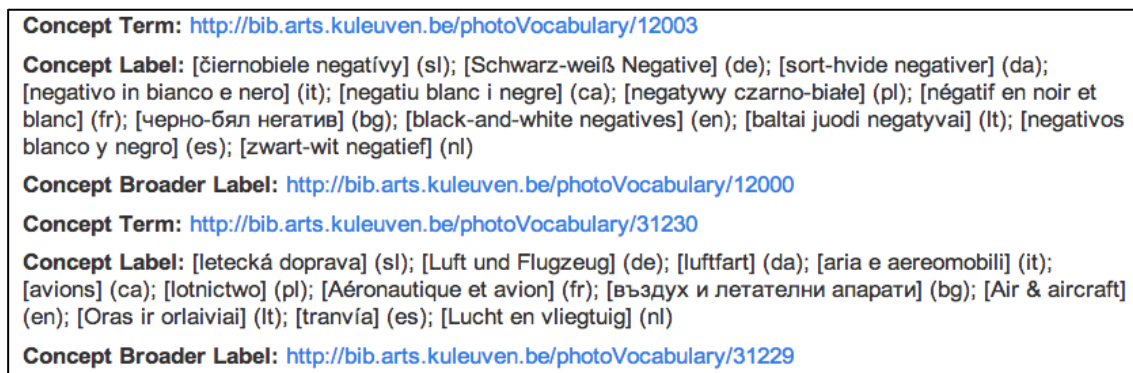
---

[1] Blogpost appeared on 20th December 2012 to explain Europeana's strategies for improvement to metadata quality in 2013 (http://pro.europeana.eu/pro-blog/-/blogs/1442749)

Figure 1: An EuropeanaPhotography record as it appears on Europeana's portal

By doing that, the following panes appear, that show again the URI together with all its labels in all languages and the broader term if any. As it can be easily noticed, however, the labels are not easily readable. What we think that it would be more appropriate for the vocabulary presentation is to replace the URIs on the top with internal links to the auto-generated tags of the page. There the URI would be available together with the labels but in a table of two columns (language, label) that would dramatically increase their readability.

Figure 2: The multilingual thesaurus as appears on Europeana's portal

From the point of view of EuropeanaPhotography partners

During this project, big discussion was carried on mostly on IPR issues and this discussion is also reflected on the labelling process of the metadata to be provided to Europeana, that is one of the mapping steps to be done through MINT. We consider it a part of the enrichment process as well, also because the Rights label is one of the key elements that Europeana itself consider of the utmost importance, as witnessed by the effort Europeana puts in its Rights Labelling campaign.

In this project, as in other ones of the same nature (e.g. EUROPhoto) emerged the tension that is normally felt by the content holders when dealing with digital holdings to be opened through the Internet.

This tension is in facts deriving from a paradox: on one side, repositories understand the positive effects of opening their contents on the Internet, but are afraid of allowing free access to them, because they wish to protect their collections as precious and income-generating assets.

What we in the end understood so far from the EuropeanaPhotography experience is that opening the access to the digital content is a true potential, underestimated so far.

In facts "metadata should be seen as advertisement for content", and the benefits which will derive from providing on line collections (as Europeana in facts is) with rich metadata are evident for the content holders, because this will:

- increase their relevance in the digital space,

- engage new users with their holdings,

- truly fulfil the specific mission of public cultural institutions to make cultural heritage more accessible to society.

From the point of view of the creative re-use of digital content available in Europeana (and other on-line collections)

Moreover, it is becoming acknowledged that one of the barriers to the re-use of digital cultural content for new creative services/products is represented by the lack of good discovery/research mechanisms inside online collections (Europeana and others), due to poor metadata.

It is not possible for a creative enterprise to re-use the digital content available online for new products /services, if the digital content is not retrievable, or does not show rich and interesting information to be re-used. For this reason, rich metadata and a dialogue among content holders, creative industry and technology providers are necessary to create and support new products for in-house or external markets, to eventually boost opportunities for employment and economic growth.

Beside the content holders, others will get benefits from this:

- **Educational institutions** will benefit from new creative products for teaching and learning, based on the use of digital cultural content and also, in particular, for example, the re-use of content accessible via Europeana

- **Research institutions** will be helped to engage in the development of innovative applications

- **General public (citizens)** will be encouraged to access to the digital cultural heritage in a rich variety of forms.

Now, it is evident that the above stated issue (i.e. the visualization of the metadata in Europeana) is more than crucial, because Europeana cannot be a good channel for content advertisement unless the visualization of this content is satisfactory, for both the content provider and the final user. It is therefore extremely important that technical issues at the basis of the current unattractive visualization are unlocked, in order to improve the "look" of EuropeanaPhotography images in the portal. The same consideration is valid also for the issue of thumbnails, that are currently (January 2014) not visible Europeana's technical team is working closely in order to provide a solution.

2.3 BACKGROUND

Preliminary work about metadata has started well before the digitization phase: in facts, being the consortium quite large and the content providers heterogeneous, a first step was undoubtedly to check the overall context, to understand where the work on metadata should start from.

For this reason, as reported in the D1.1.1 Annual Report first release, par.1.3.3 :

"a survey was created and sent to each partner to probe the use of controlled vocabularies in each institution, and the underlying technical possibilities of each database program. The survey results were analyzed by KMKG and presented to all partners at the content seminar in Leuven in April 2012, where a set of mandatory fields for the consortium was decided."

The decision over the EuropeanaPhotography mandatory fields was complemented by an analysis about the languages, preparatory to the realization of the EuropeanaPhotography multilingual vocabulary.

The language analysis after the survey is sum up in the same paragraph 1.3.3:

"Most partners' metadata is only available in the local cataloguing language, making a strong argument for the wide use of a common vocabulary within the consortium, to provide both the consortium and the public with the best possible means to make the metadata complimentary, provide interesting search possibilities and allow for the most significant search results."

2.4   APPROACH

Guidelines and standards were set at the very beginning of the project, to feed the local cataloguing activity to be carried on by the content providers on an individual basis. Next to that, proper tools were realized by the technical partners NTUA and KMKG to support metadata mapping according to the EuropeanaPhotography mandatory fields, and to provide metadata translation in different languages.

Content providers digitized their photos, indexed them, added the requested fields to their own local catalogues (if necessary), and then used the tools integrated in MINT to complete the whole process. A dedicated group for support, the so-called "metadata task force" guaranteed a proper monitoring and technical help. It is also worth to mention the big effort in training activities that allowed all the content provider, even the most un-experienced, to successfully accomplish the tasks. The result is a huge metadata amount, rich in information and available in different languages, delivered to Europeana.

2.5   STRUCTURE OF THE DOCUMENT

Chapter one is the publishable executive summary, i.e. a document in miniature that may be read in place of the larger document.

Chapter two is the present introduction.

Chapter three is dedicated to the rationale of the enrichment process: the selection of metadata mandatory fields, the local catalogue enrichment, the Vocabulary. In this chapter a sum up of the work done before the enrichment phase is provided and sum up once again, with proper reference to previous deliverables where needed.

Chapter four provides comments about the enrichment experience by the content providers that actually had to deal with it.

Chapter five shortly describes and sum up the results of the enrichment process, to support the enriched metadata delivered to Europeana at M24.

Chapter six is the brief conclusion of the document.

3 RATIONALE OF THE ENRICHMENT PROCESS

This chapter describes and sums up the three elements that are the basis of the enrichment process:

- The mandatory metadata fields that were set for EuropeanaPhotography

- The local catalogue enrichment

- The Multilingual vocabulary

According to the DoW, enrichment of EuropeanaPhotography metadata should focus on (but not limited to) three items: "events, locations and individuals", possibly by selecting them from structured lists.

It was discussed whether to select "individuals" from Authority files, but in the end it came out not to be technically possible. It would be difficult to incept the whole Authority Files into MINT and too a big effort to reduce it "manually", also because it will be almost impossible to know and select renowned authorities for all the countries represented in this project.

So the "individuals" have to be added by typing the names of the represented people in each photo at the time of indexing, by each content provider. The same goes for "events" and "locations"

Moreover, the Multilingual vocabulary includes three subcategories: "Subject", "Technique" and "Photographic Practice"

3.1 MANDATORY METADATA FIELDS

To understand which metadata fields should be considered mandatory for the content providers, two aspects were taken into consideration: the local catalogues and the Europeana Data Model (EDM).

As for the local catalogues, as reported in D2.1 Content Seminar Proceedings, par. 5.1, the situation was the following:

"What we found (thanks to the metadata survey conducted at the beginning of the project by KMKG, A/N) *was that:*

Record ID: most partners (11 of 14 who filled out the survey) already register some kind of record ID in their database.

Title / description: as a mandatory field for EDM (see 5.2), every partner needs to have either a title of a description of the work he will publish on Europeana. However, most partners (13/14) have already foreseen this in their database.

Keywords / subjects: not mandatory for EDM, but needed for the thematic Europeana Photography Vocabulary we will build, the keywords will be organised in a thesaurus structure to allow for maximum readability and interoperability between partners, and provide a basis to search within the thousands of photos we will make available. Already included in the databases of 11 partners.

Dimensions: *12/14 partners give some indication to the dimensions of the physical object to be digitised (not the subject) in their database. This is not a mandatory field, and in photography often standardized.*

Material / technique: *some, but not all partners differentiate between materials and techniques. However, 12/14 give some indication as to the process used to create the original photo, which is a second group of concepts we will be organising in a thesaurus structure*

Places: *information describing the place where the photo was taken is available for 9 out of 14 partners. This information will be linked to Geonames of another Geographical thesaurus that is available in multiple languages.*

Author: *for the most part the author of the photo has been registered in partners" databases (12/14). Additional information might be gotten from the collection name, or historical descriptions of sub-collections.*

Copyright information: *only 6/14 register some kind of copyright notice in their database. This is only necessary when a collection or sub-collection has more than one copyright holder. If not, copyright details will be provided by the consortium when publishing data on Europeana.*

These metadata were sometimes, but more often not, supported by controlled lists or thesauri, mainly when applied for names, keywords, places and techniques. When partners used controlled vocabularies, little to no international standards were in use; vocabularies are mainly developed in-house."

This quite heterogeneous context has to be combined with the Europeana Data Model (EDM) requirements, that include some mandatory fields. As most of them are automatically generated by the MINT mapping system or by Europeana, EDM actually requires very little information to be provided by the partners: basically a title or description.

Such small information is considered enough for Europeana, but during discussions of the whole consortium it came out that a strict application of EDM will not clearly express the quality, variety and richness of the photographic content provided in this project.

For this reason, and in the light of providing "enriched" metadata to Europeana, the consortium agreed with a larger mandatory set of fields, outlined in the table below (the highlighted box).

Furthermore, several useful fields are in use by some partners but not by all, and it is therefore strongly recommended, although not mandatory, that these are provided by all the partners for each photo.

ESE mandatory	Minimal EU Photo	EU Photo Intermediate	Examples of content	Remarks
edm:country		edm:country		country of the data provider
edm:isShownAt		edm:isShownAt		link to metadata on your website
edm:isShownBy		edm:isShownBy		link to photo on your website
edm:provider		edm:provider		name of the aggregator, if applicable
edm:dataProvider		edm:dataProvider		name of the metadata provider
edm:type		edm:type	IMAGE	type of content partner delivers to Europeana
edm:rights		edm:rights		link to the Europeana rights documentation
edm:uri		edm:uri		permanent link where the metadata is shown
edm:language		edm:language		language of the metadata provider
dc: type		dc: type	Photography	Always "Photography", as agreed with partners @ content seminar
dc:title	dc:title	dc:title	Two women, George Bernard Shaw, Sint Michielskerk te Leuven	title of your work, can be the caption
	date	dcterms:temporal	1900, 1910-1920, before 1870	date when the photo was taken, so between 1838-1939.
	author	author	Stanley Kubrick, Alinari	name of the photographer, or studio, who made the photo
	technique	technique	daguerreotype, negatif	photographic technique used to make the photo, to be linked to EuPhoto vocabulary
	location	dcterms:spatial	Poland, Leuven, Rome, Western Europe, Rhine	place where the photo was taken, to be linked to Geonames or other
	description OR keywords	dc:subject and dc:description	Woman, WW II, wedding	since not everyone has keywords, a description can be enough. Preferably keywords however
	copyright	copyright	© KMKG - MRAH	text string with the name of the copyrightholder. If empty, the institution's name will be shown
		photographic practice		Why was a photo taken? Preferably in keywords, not always available in partners' metadata (or not always useful)
		subjectConcept		keywords
		subjectActor		person IN the photo, or who made the object in the photo
		subjectPlace		place IN the photo, usually the same as production place, except for paintings
		dimensions		measurements of the original
		material	glass, paper	material of which the original is made
		related works		links between several records, from the same album, different pictures from the same object, ..

3.2 ENRICHING THE LOCAL CATALOGUES

Being a digitization project, the digitization metadata have to be created in any case (with just one exception, i.e. the case of Fondazione Alinari).

Whether a local catalogue was already existing or had to be created ex-novo, main activity for all the content providers in their own databases was to check and identify the fields to be mapped in MINT.

The content providers had very different backgrounds, to be sum up in three scenarios:

1. Few partners did not have a local catalogue (database), they just had the photos in their archives. This is for example the case of Divadelny Ustav and ICIMSS. In this case, they had to create a database, according to the EuropeanaPhotography requirements, to digitize each piece and to add to the database also the digitization metadata. And then to map the concepts to MINT..

2. Most providers had an existing local catalogue (database) related to the photographs in their holdings. In this case, they mostly had to identify the local fields that correspond to EuropeanaPhotography required fields, in order to study and understand the easiest way to map the concepts inside MINT.

3. In one single case (Fondazione Alinari) the database already existed and the images were already digitized. For Fondazione Alinari, therefore, the main activity was to study and understand the easiest way to map the concepts inside MINT.

3.3 THE EUROPEANAPHOTOGRAPHY VOCABULARY

The preliminary information derived from the survey has driven to the realization of the EuropeanaPhotography Vocabulary, as widely described in D4.1 EuropeanaPhotography Vocabulary definition:

"In this survey, all partners were requested to state, among others, what vocabularies they were using (…), in which ways these vocabularies were used and to share their vocabularies. This survey was completed in the early months of this project and the results communicated to the consortium at the content seminar in M3. (…).

At this same content seminar, 3 main facets of the vocabulary have been decided upon. (…) From that point onwards, several drafts of the vocabulary have been made, shared with the content partners, discussed and refined."

The "source" catalogue record metadata used by each content provider is expressed in the national language of the content provider. Developing a multilingual vocabulary was a team work including the task leader KMKG, who organized the vocabulary in English and then provided Dutch and French translations, and most of the content providers, that translated the vocabulary in their home languages.

In particular the translations were done as follows:

- United Archives for German

- ICCU for Italian

- ICMSS for Polish

- Polfoto for Danish

- NALIS for Bulgarian

- Divadelny Ustav for Slovak

- Lithuanian Museum for Lithuanian

- CRDI for Catalan

- GENCAT for Spanish

This vocabulary originally included 12 languages - English (as the pivotal language), French, Dutch, German, Italian, Polish, Danish, Bulgarian, Slovak, Lithuanian, Catalan and Spanish – and consists of nearly 500 concepts in three facets – photographic technique, photographic practice and keywords – that are structured in a multifaceted, hierarchical way.

In January 2014, three more languages were added: Ukrainian and Russian (translations done by Mrs. Olga Barkova of the affiliated partner SC BALI from Kiev) and Chinese (translation done by Promoter's company members XiaoChun Situ and Elisa Debernardi).

Discussions are currently ongoing with potential partners in order to provide, Finnish, Israel and Gaelic translations of the Vocabulary.
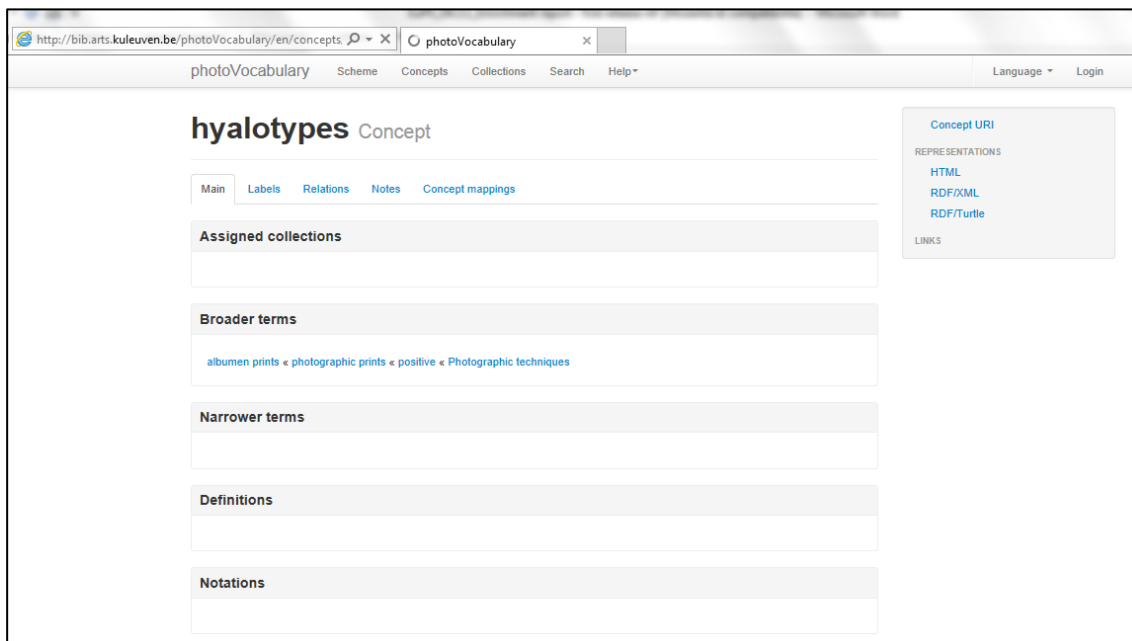
At the same time, EuropeanaPhotography is now collaboration with the Linked Heritage project and their successor, AthenaPlus, on aligning the developed vocabulary with vocabularies from other projects. In time, this will create a network of linked vocabularies allowing contextualization and further translation of keywords, and a linking of like metadata. The first tests have taken place in the summer of 2013, and are now being analysed. The further workflow and possibilities for collaboration will be decided on the basis of these test results.

3.4   GETTING BACK THE ENRICHED METADATA

Once the metadata are ready for harvesting by Europeana, here it ends the EuropeanaPhotography activities for the content providers, but another possibility is left for them, to get back at least a part of this enrichment work in their own databases.
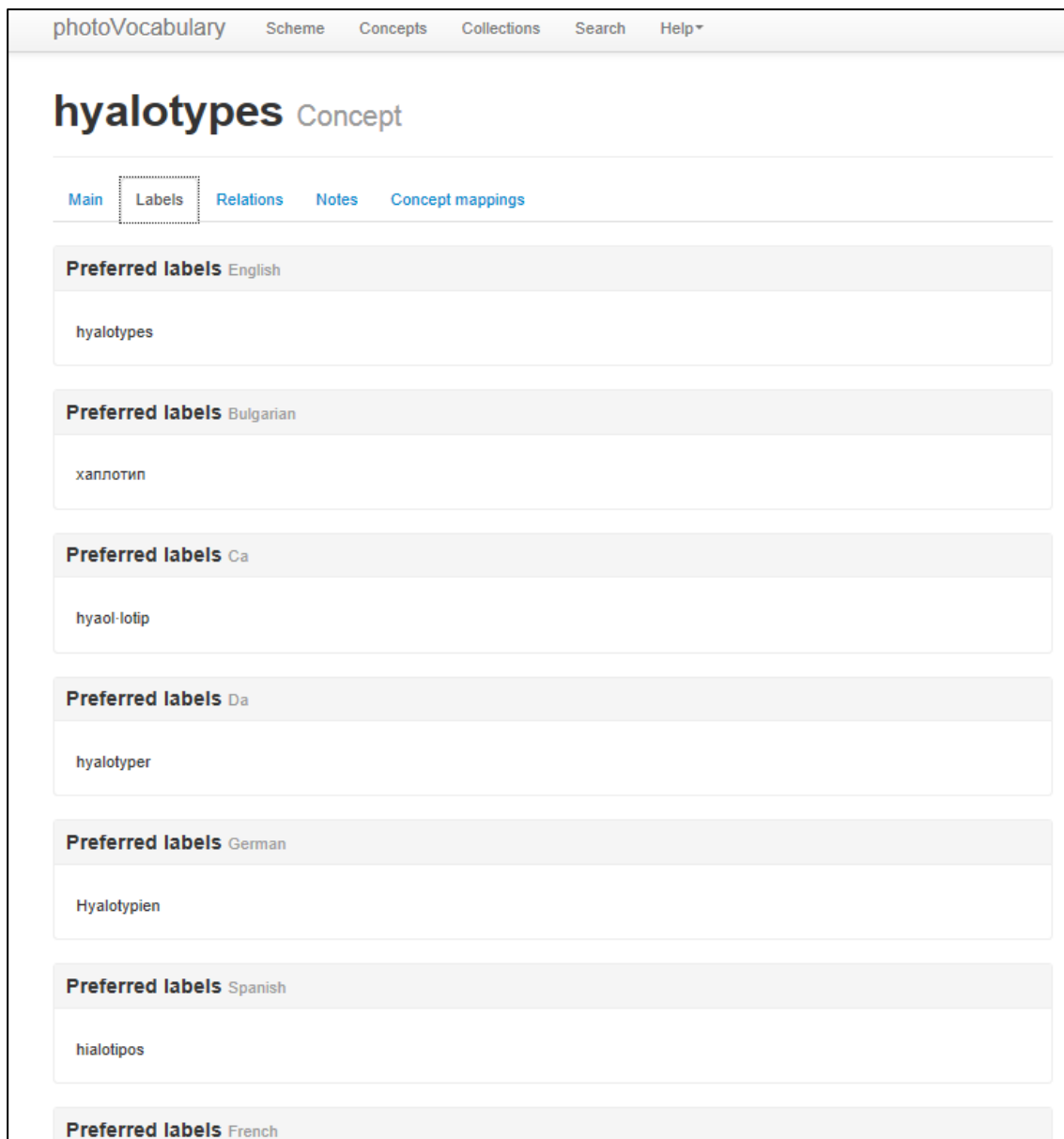
According to the MINT workflow, the provider imports the metadata of a record. An element called subject is in the metadata having the value "hyalotypes". After importing to MINT, the provider can map this value to the EuropeanaPhotography vocabulary. What will appear instead of this value in his resulting metadata (LIDO and EDM) using the MINT mapping is this URI:

http://bib.arts.kuleuven.be/photoVocabulary/11013

Selecting "Labels" on the top, the page will show the labels of the term in all the translated languages, see figure on next page:

The added value of using a SKOS vocabulary as we developed is that Europeana:

- displays on the portal the label of the term in the same language as the rest of the metadata (e.g. hyalotypes)

- extracts the remaining labels from this URI and uses them as indexing terms i.e. terms that can someone use for searching their portal (so if someone search through Europeana's portal for hyalotipi - the record will appear.)

A provider can download through MINT his metadata in LIDO and EDM that will have this URI. If he wishes to transfer this URI back to his original metadata, then he has to identify the terms he uses in his metadata and he has to replace them with the URIs of the vocabulary using his local metadata management system.

4 THE ENRICHMENT EXPERIENCE

After the first batch of metadata was processed and ingested, so as everybody had the personal experience of use of the tools and of the enrichment process, a questionnaire was circulated to collect comments and suggestions post-use. The questionnaire form is available as Annex I. The questionnaires provided by the partners are available in the project's repository[2].

All the partners consider the technical support for MINT and Vocabulary, and particularly the work of Nikos Simou and Nacha Van Steen, really precious and effective, with a true problem-solving approach that was really appreciated. This is certainly a prove of the fact that experienced technical partners and proven tools are extremely important in order to drive these kinds of project to success, and to orchestrate the work of many partners with different backgrounds and requirements.

CRDI probably sum up the matter in the clearest way, by saying:

"The hard job is in the local catalogue as you need to normalize the descriptions in order to fit with the standards. Then, when the local catalogue is fine, it is quite easy to do the mapping in MINT."

The MINT tool and the vocabulary mapping tool are generally considered user-friendly and intuitive, with very few exceptions and criticisms (e.g. Fondazione Alinari, ICIMSS). The weaknesses are useful to be known in order to allow improvement, and for this reason the coordination team was happy to hear different points of view.

For example, the comment of ICIMSS highlights one of the weaknesses in the structure of the vocabulary:

"We are sceptical about this vocabulary and find it difficult to use. As some thematic areas are covered at sufficient level of accuracy, the other seems to be treated too broadly. Too general terms means that a user searching for materials in Europeana let's say "horses", will have to search (if s/he can imagine the point of view of terminology creators) for "animals", and go through all pictures with no result or very limited result."

It was certainly difficult for the content providers, at the time of Vocabulary creation and designing, to predict possible problems like the one highlighted above: in facts many comments and suggestions came from the providers later on, while they were actually using the Vocabulary and facing actual issues. Following this comment, it is certainly possible to improve the Vocabulary for the next mapping and ingestion activities, by adding new terms and providing the related translations. A new "input period" is therefore going to be open for the partners to suggest terms additions.

The majority of the partners confirm that they have a wider and more advanced indexation system and taxonomy that the one developed to build the Vocabulary, but the whole group agrees that the Vocabulary needed to be created in order to be flexible and usable by all the content providers.

---

[2]   http://www.digitalmeetsculture.net/heritage-showcases/europeana-photography/europeana-photography-reserved-area/

REPOSITORY ❯ WPs ❯ WP4 ❯ questionnaires about enrichment

As explained by Arbejdermuseet:

"It was a necessary compromise. Maybe some of the content providers got more "own" terms in, than others. But our own vocabulary, for instance, is rather specialised, and it would have been without reason to insist on this and other providers' full vocabularies included. Of course we miss some terms in MINT, but I think we've managed it in a fair way"

But there were positive effects deriving from the need of adapting the existing databases and processes to the tools developed by the project. For example as Parisienne de Photographie says: *"The very positive aspect of the exercise was to force us to normalize the descriptions of photographic techniques and photographic practices throughout our corpus. An added – and more painful - benefit was to underline some weaknesses in our own indexation practices, therefore motivating us to improve it!"*

The whole "informatics" process was a challenge for the less experienced partners, as indicated by Divadelny Ustav: *"It is very difficult for us to provide the whole ingestion and enrichment work. There are very few people who are experts in this field in Slovakia and they enjoy their unique position. The positive effect is that we in the Theatre Institute slowly understand more and more the requirements and processes connected with this work."*

From the questionnaires, there were also suggestions and reflections related to a broader point of view that the project, for example:

MHF:

"We are hoping for some kind of solution on the IPR issue – maybe another metadata field with more specific description would be a good idea. We are hoping for more precise labelling – it might be a 'task' for lawyers and legislators from different countries. "

Topfoto:

"The vocabulary is very limited to our project; my recommendation to Europeana would be to create a mass vocabulary to cover all content providers to Europeana. it is a massive task but the inaccuracies created by a vast number of partial and overlapping vocabularies for all of the different projects going into Europeana simply does not make sense both in terms of replication of work and also the user friendliness to the end user.

Imagno*:*

"We are interested in learning more about famous or important people, one can see on the images and we are interested in the artists. Maybe there is a possibility to install a "who is who-tool", which helps users to find out more about the VIPs and artists of Europeans cultural heritage

5 RESULTS OF THE ENRICHMENT PROCESS

The following tables present statistics on the metadata published by the EuropeanaPhotography project to Europeana. In detail the EDM xpaths are presented together with the distinct values and the average length of the value.

Xpath	Count	Distinct	Length
⊟ rdf:RDF	219743		
⊟ edm:ProvidedCHO	219743		
@rdf:about	219743	219574	93.9
⊟ dc:creator	227598	5974	16.3
@xml:lang	229767	11	2
⊟ dc:description	172302	126686	129.6
@xml:lang	172302	11	2
⊟ dc:format	323565	1464	20.4
@rdf:resource	159078	29	49
@xml:lang	301907	8	2
dc:identifier	263888	226493	17.2
dc:rights	231091	241	39.2
⊟ dc:subject	1943770	93128	14.5
@rdf:resource	331273	423	49
@xml:lang	1901558	11	2
⊟ dc:title	230939	140396	62.1
@xml:lang	232100	12	2
⊟ dc:type	372189	49	15
@xml:lang	372189	8	2
@rdf:resource	115187	13	49
dcterms:created	170526	12668	11.4
dcterms:extent	161104	9029	16.3
⊟ dcterms:medium	110563	55	7.3

Figure 3: Statistics for the edm:ProvidedCHO resources

@xml:lang	112427	7	2
@rdf:resource	9456	2	49
dcterms:spatial	276634	8483	14.3
edm:type	219743	1	5
⊟ dcterms:alternative	7386	5948	54.2
@xml:lang	12332	2	2
⊟ dc:publisher	3337	30	23.3
@xml:lang	3337	1	2
⊟ dc:contributor	3781	898	15.3
@xml:lang	7844	1	2
dc:date	7844	1052	8
dcterms:provenance	18958	10	14
dcterms:isPartOf	4586	153	31.8

Figure 4: Statistics for the edm:ProvidedCHOs resources (continued)

Xpath	Count	Distinct	Length
⊟ rdf:RDF	219743		
⊞ edm:ProvidedCHO	219743		
⊟ edm:WebResource	424148		
@rdf:about	424148	404060	77.2
dc:rights	46584	28	52.8
⊟ skos:Concept	612058		
@rdf:about	612058	451	49
⊟ skos:prefLabel	7323597	4858	16.2
@xml:lang	7323597	12	2
⊟ skos:altLabel	653906	461	15.8
@xml:lang	653906	2	2
⊟ skos:broader	579994		
@rdf:resource	579994	82	49

Figure 5: Statistics for the edm:WebReousrce and skos:Concept resources

Xpath	Count	Distinct	Length
⊟ rdf:RDF	219743		
⊞ edm:ProvidedCHO	219743		
⊞ edm:WebResource	424148		
⊞ skos:Concept	612058		
⊟ ore:Aggregation	219743		
@rdf:about	219743	219574	93.9
⊟ edm:aggregatedCHO	219743		
@rdf:resource	219743	219574	93.9
edm:dataProvider	219743	46	22.7
⊟ edm:isShownAt	204407		
@rdf:resource	204407	204403	77.1
⊟ edm:isShownBy	210285		
@rdf:resource	210285	200244	77.8
⊟ edm:object	194821		
@rdf:resource	194821	184654	70
edm:provider	219743	2	20.1
⊟ edm:rights	219743		
@rdf:resource	219743	5	38.9
⊟ edm:hasView	10029		
@rdf:resource	10029	10027	103.8

Figure 6: Statistics for the ore:Aggregation resources

As we can see 612.058 skos:Concept resources have been used in total for 219.743 items. In other words ~3 skos:Concepts resources have been used per each item. In detail the edm elements to which the vocabularies resources are mapped are

- dc:format: 159.078 vocabulary resources have been mapped (29 distinct)
- dc:subject : 311273 vocabulary resources have been mapped (423 distinct)
- dcterms:medium: 9456 vocabulary resources have been mapped (2 distinct)

Of particular interest it is the statistics for the IPR label that the providers chose for this first batch of metadata:

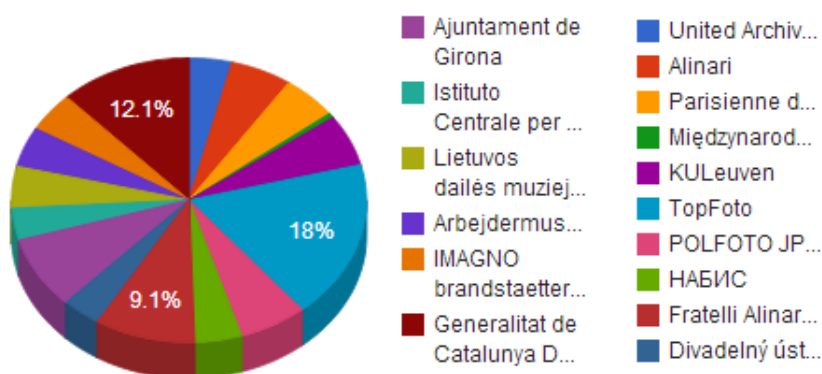| Rights Label	#items	% total
RR-P	95504	43,46%
RR-F	70126	31,91%
CC-BY-NC-ND	38427	17,49%
Public Domain Mark	8037	3,66%
Unknown	7649	3,48%
RR-R		0,00%
CC0		0,00%
CC-BY		0,00%
CC-BY-SA		0,00%
CC-BY-ND		0,00%
CC-BY-NC		0,00%
CC-BY-NC-SA		0,00%

It is worth to notice that the very big percentage of RR-P depends on the choice of most private partners. Not all the private partners chose RR-P for their metadata, some chose RR-F that is the second preferred label according to these statistics, but it is certainly notable that the 95.504 items that are labeled RR-P all belong to private content providers. who have a business-protective approach when applying rights labels to their digital content with IPR relevance and business relevance.

This is still an open issue, and discussions are ongoing with Europeana to identify a possible improvement of the Rights Label framework that meets such need while at the same time ensuring the enjoyment and access of the user to the digital material in Europeana.

A good performance is also done by the CC-BY-NC-ND: almost the 20% of the whole, while only a 3.5% records have "unknown" rights label,

Finally this is the graph which shows the composition of the whole batch:

Records per organization

Interesting note: the main "contributors" so far come from both sectors:

    18% corresponding to 39.538 by TopFoto (private company)
    12% corresponding to 26.595 records by GenCat (public archive)

It is a priority task for the "metadata task force" to ensure the best alignment of all the partners to the workflow. Given that most of the providers did not have previous familiarity with the MINT tool, it is evidence of accomplished good work the fact that everybody was able to deliver metadata at the first harvesting. Refinements can certainly be done and the look of records in Europeana can be improved, but the overall process was successfully completed by all the content providers.

6 CONCLUSION

The enrichment process was a crucial activity in EuropeanaPhotography during the year 2 and resulted in over 219.000 new items available in NUTA's servers, of which over 210.000 have already been ingested in Europeana. The job of the content providers is not finished, as during year 3 the rest of the images (total amount foreseen in the project: over 430.000) needs to be processed and then delivered to Europeana.
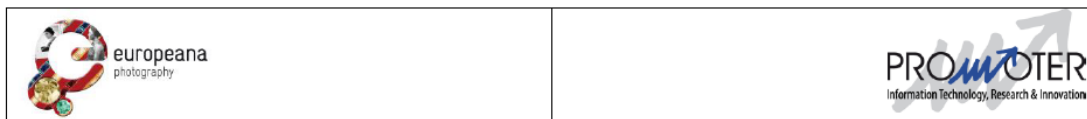
The experience gained in year 2 with local enrichment and use of MINT and Vocabulary mapping tools will drive the providers in a smoother and faster process towards the full completion of the tasks.

The continuous support by the metadata task force, and by NTUA through plenary training sessions, will keep the providers on track in order to make available on NTUA's servers the rest of metadata by M33, October 2014 corresponding to the deadline of D4.3.2 Enrichment Report second release.

In the upcoming months, efforts to improve the whole system and to possibly add new languages to the EuropeanaPhotography vocabulary will be provided, as well as the possibility for the consortium to add other terms.

At the same time, cooperation with Europeana,(that is progressing on many different fronts next to the metadata technical issue, e.g. IPR and rights labeling issue, dissemination, cooperation in the EuropeanaPhotography exhibition) will be targeted to improving the visualization of data and solving present and possible technical problems.

7 ANNEX I: THE "ENRICHMENT" QUESTIONNAIRE

EUROPEANA PHOTOGRAPHY – D4.3.1 ENRICHMENT REPORT

QUESTIONNAIRE FOR THE PARTNERS

This questionnaire aims to collect comments from the partners about the enrichment experience.

The enrichment was done at two levels: on the local catalogue (i.e. your own database) and inside MINT, during the mapping process.

To be completed by each content provider: NAME OF THE PARTNER……………………..

1. Mandatory metadata fields - In the Content Seminar held in Leuven on 12-13 April 2012, a set of mandatory fields for EuropeanaPhotography mapping was established. What do you think about this, were the mandatory fields narrower or wider than the local fields you normally use in your database?

Any other comment?

2. The EuropeanaPhotography Vocabulary developed in collaboration with KMKG – What do you think of the final result (about 500 terms)? Is it a good compromise between two requirements: from one side to be an easy/swift tool and on the other side to be complete?

Any other comment/suggestion?

3. The "hard job" on the local catalogue and into MINT - Was it difficult or easy to provide this "enrichment" work? What are the positive effects that you experienced?

Any other comment/suggestion?

4. EuropeanaPhotography through MINT focused mostly on multilingual enrichment. Which other kind of enrichments would you wish to be developed in the future?

Any other comment/suggestion?