SCIRES it
SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 4, Issue 2 (2014), 127-132
e-ISSN 2239-4303, DOI 10.2423/i22394303v4n2p127

METADATA MANAGEMENT IN EUROPEANA PHOTOGRAPHY

*Nacha Van Steen**

*Royal Museums of Art and History – Brussels, Belgium

Abstract

This paper describes the metadata management experience of the Europeana Photography project, a digitization action with final aim of making available in the internet over 430.000 items of early photographs with historical, cultural and artistic value, belonging to the first 100 years of the art of photography. The metadata converged in Europeana, the European Digital Library which collects about 31 million digitized items of European cultural heritage. It illustrates the way the Europeana Photography project has created and enriched metadata of images being delivered to Europeana through the project, and the context in which it has done so.

Keywords

Europeana Photography, semantic web, vocabulary

Whether you choose to see the current possibilities as a distinct evolution of the internet, web 3.0, or rather share the opinion that those possibilities were present form the first, the semantic web is the future of online life, for cultural heritage and their guardian institutions just as well. While we're evolving to a place where your smartphone, -watch or tablet becomes your personal butler, knowing what you want before you even realize it yourself, memory institutions stay rather slow on the uptake for various reasons – financial, a matter of priorities, or simply because the cultural heritage sector is often not the envisioned target audience of technical and technological developements.

However, as the Europeana Photography project[1] is coming to a close at the end of January 2015, having delivered over 400.000 photos from the period 1839-1939 to the Europeana[2] portal, we'd like to look back at the metadata management within the project and share with you how we tried to play a part in delivering data that is not only interesting, but also rich in metadata and description, and searchable in a large variety of languages. For your enjoyment, and as a teaser, Fig. 1-4 give you an idea of the content as delivered by Europeana Photography.

Fig. 1: *Gerona. Paseo Central de la Dehesa*. Photo by Desconegut, 1900, courtesy of Ajuntament de Girona. Public Domain marked

For us as a project, delivering content from 19 partners from 13 different European member states, metadata management and quality control was a key issue to guarantee high quality metadata on Europeana. While the images may speak for themselves, and include masterpieces from eminent private and public photo agencies, institutions and museums, they still are in danger of getting lost on a platform that operates on the scale of Europeana. Good metadata is key when you want your information found, discussed and shared, so from the beginning, good metadata was the rule within our consortium.

[1] Project under the 2007-2013 CIP framework by the European Commission
[2] www.europeana.eu

This started with defining the metadata elements which we considered to be elementary, necessary, and mandatory for all project partners.[3] Luckily for us, there was a shared sensibility within the consortium, but even at that time, the differences in approach between partners became visible. As a consortium grouping both private partners, who depend on high quality data and metadata in order to supply whatever demands of their clients, and public partners, who tend to breach the subject of metadata traditionally from a scientific point of view, this was both a very interesting contrast as well as a point of attention in dealing with the different datasets.

Fig 2: *Portrait of Count Mišelio de Boré*. Photo by Fotoateljé "Kaufmann & Kesseler", 1868, courtesy of Kretingos muziejus. Public Domain marked

[3] Europeana Photography deliverable 2.1: Content seminar proceedings, www.europeana-photography.eu/index.php?en/115/deliverables

The Europeana Photography vocabulary

After setting out the minimal requirements, it was possible to provide the partners with a software platform which would allow them to upload, map and transform their own in-house databases into a metadata schema compatible with the Europeana portal. This is the MINT online metadata mapping tool provided by the National Technical University of Athens (NTUA), in use for many other Europeana-feeder projects, which was customized to suit our needs, and came complete with bookmarks and a user manual. Then, a second issue had to be addressed: there were not only differences in approach when creating metadata, but also 12 different languages in use. To us as a consortium, it was not interesting enough to simply provide our partners' data as small photo albums within a larger one, we wanted to reach our public together, and across language barriers. So, we created a vocabulary.

The vocabulary was to be a thesaurus, as it provided us with structure and control, while not being as time consuming to create as an ontology – we needed it ready for use within the year. It allowed us to add synonyms and a variety of languages, and could be created rather easily in a spreadsheet. Because we wanted to achieve the highest quality of vocabulary possible, we used the sources available to us: the *Art and Architecture Thesaurus* from the Getty Institute[4] for the techniques, the headers from IPTC[5] for keywords, the information gathered throughout the Sepia[6] project , but mainly, we drew on the experience and expertise of our partners, all experts in their field, and with great knowledge of the needs of their collections.

Creating the shared Europeana Photography vocabulary meant creating a tool that allowed us to tag our metadata in the languages of all partners (since its conception, it has become available in 16 languages, including Chinese and Hebrew), aligning our collections to improve

[4] www.getty.edu/research/tools/vocabularies/aat/. The translations of the technical concepts will also be delivered to the AAT itself, enriching their vocabulary and adding to the sustainability of ours.

[5] International Press Telecommunications Council, www.iptc.org

[6] Safeguarding European Photographic Images for Access. Some documentation is available here: www.ica.org/5671/paag-resources/publications-for-archivists-managing-photograph-and-film-collections.html

searchability on Europeana, to more easily identify underlying themes in the collections, and to present the collections as a whole. It meant, from a more technical point of view, identifying those metadata elements who would most benefit from the multilinguality and uniformisation, without losing the individual characters from the datasets, and implementing the vocabulary in the delivery workflow.

Therefore, we decided to focus our vocabulary on three main topics: the photographic techniques used in the creation of the original, the keywords describing the image, and the reason for which the image was created, which we dubbed *photographic practice*[7]. We organized the concepts in a thesaurus structure, finding middle ground between high detail and rich descriptions, high quality of both concepts, translations and structure, and usability for partners as well as the public. Using a spreadsheet for the first drafts allowed to drag and drop, add new concepts easily, and share fast and simple between partners. Using the thesaurus structure allowed us to make use of synonyms, broader and narrower terms, but also to add scope notes and sources, again improving the quality of the vocabulary. The vocabulary was then added, either in our partners' local databases, or in the mapping stages before delivery, to the existing metadata. Further enrichment was done by Europeana itself, who, upon delivery of the content from all partners, matched agents names, place names, time periods and keywords to databases as DBPedia and Geonames, further improving searchability.

Throughout the whole process, quality control was a key issue for us, from the digitization of the images – supported by fact sheets[8] summarizing the available knowledge on the subject into an easily accessible format -, over the development of the vocabulary by experts, to the extensive use of the available resouces by the partners. We needed our partners to deliver rich, multilingual data to Europeana to do justice to the images they keep.

Creating a vocabulary on behalf of a project of course also has downsides: the timeframe in which to work is limited, therefore not all wishes can be granted; it is dependent on the input of the partners, who may have their own priorities; and it is project-driven, so not all themes are covered as they are/were not pertinent to the project.

Enriched metadata as part of the semantic web: a web of opportunities

While enriching metadata with vocabularies, either specifically created or existing ones, improves searchability and multilinguality dramatically and immediately, it's important to see it in the larger context of the semantic web, and understand and discover its potential for the cultural heritage world.

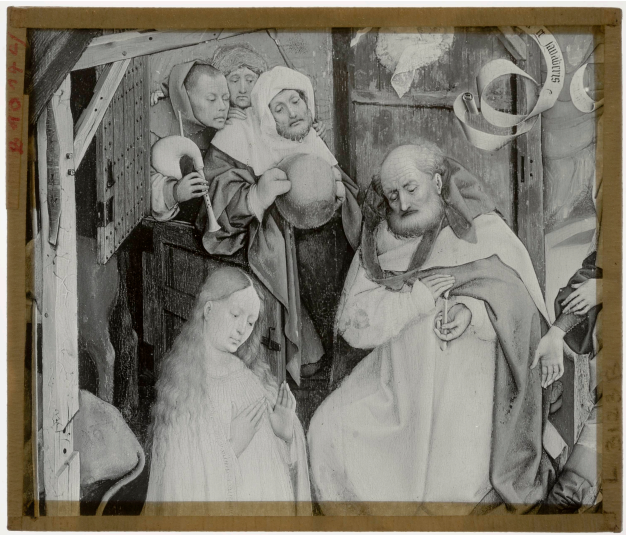

Fig. 3: *Detail from The Birth of Christ*. Anonymous, courtesy of KU Leuven. Public Domain marked

The semantic web is not a new concept: it has been around for over a decade, but is slowly finding footing in the cultural heritage community. It is, depending on the sources you wish to follow, either a third phase in the evolution of the internet, and therefore often considered synonymous to web 3.0, or a fulfillment of potential of things already there at the invention of the world wide web.[9] Either way, to be able to consider its potential, a basic understanding of the concept is useful. The semantic web harvests all the knowledge available on the internet to deliver you highly personalized content. To do so, it needs to understand the available information, so content creators (publishers, memory institutions, or even people on social media) need to identify their content: which strings of letters are names,

[7] The entire vocabulary, in all 16 languages, is available online at bib.arts.kuleuven.be/photoVocabulary/en.html

[8] www.europeana-photography.eu/index.php?en/117/documents

[9] www.w3.org/2001/sw

which are place names, and how do they relate to each other? These relationships between concepts then create a second layer of information in the background, and make it possible to link data to each other: you create a web of semantically[10] identified information. This allows you for instance to compare flight prices, get personalized publicity, of get introduced to the right people on social media. A new world of interaction, exploration and possibilities opens up, at a price.[11]

When contributing to the semantic web, whether it's information on the object in your memory institution or highly specialized vocabularies, you need to think first of the license under which you will publish your content, and find a balance between protecting what is your intellectual property and what you wish to share. In practice, an open license is virtually necessary.

Fig. 4: *Balkan Wars (1912-1913). Commander of 51th Infantry Regiment gives the order to attack*. Photo by Karastoyanov, D., 1912, courtesy of NALIS Foundation. Public Domain marked

Next, you identify your content with URIs[12], and you link your data through existing or bespoke software[13], being careful to use standards every step of the way. For the Europeana Photography project, this meant adding URIs in our spreadsheet, and SKOSifying[14] the vocabulary, before publishing it online under an open license, and adding the links in our metadata.

Making use of, and contributing to, the possibilities of the semantic web means sharing your stories beyond traditional channels, making them available to interest groups you hadn't thought of, or whose languages you do not speak. It means showing them in context, and making them more sustainable through sharing of the information. And it means engaging a larger audience in new ways: from here, it's a very short step to virtual exhibitions – Europeana Photography has created the All our Yesterdays virtual exhibition[15], based on our partners images – crowdsourcing or augmented reality applications. While Europeana Photography may be coming to the end of its life as a project, the group of partners will stay together, and for this reason founded an association: the expertise and drive developed thanks to Europeana Photography will continue on in the Photoconsortium[16], exploring the different roads to take with our content, providing innovative services related to the photographic heritage and sharing our knowledge.

[10] Information that has been identified by its meaning, that has been defined

[11] Generally speaking, the infringement on privacy and the question on the trustworthiness of the links are considered the greatest threats to the system.

[12] Uniform Resource Identifiers. URLs – Uniform Resource Locators – can be considered a subtype of URIs

[13] The Europeana Photography project works with MINT, developed by NTUA, for both metadata and vocabulary mapping. We also collaborate with the projects Linked Heritage and Athena Plus, who have developed a Terminology Management Platform precisely for this kind of vocabulary linking.

[14] Simple Knowledge Organisation System, standard for organizing concepts; classifies resources in terms of broader or narrower, preferred and alternate labels and allows for quick publication of thesauri and glossaries to the internet.

[15] www.earlyphotography.eu and in the App Store

[16] www.photoconsortium.net

Fig. 5: *Europeana Photography consortium during the Europeana Photography plenary meeting at Vilnius, Latvia, 10 September 2013*. Photo by Vaidotas Aukštaitis (Lithuanian Art Museum), 2013. This photo has been made with an ICA photocamera (Dresden, ca. 1920), slightly modified for contemporary use. CC-BY-NC-ND.

REFERENCES

Bachi, V., Debernardi, E. & A. Fresa (2014). *D4.3.2 Enrichment Report (second release) – preliminary version*. To be retrieved from http://www.europeana-photography.eu/index.php?en/115/deliverables

Bruce, T.R., & Hillmann, D.I. (2004). *The continuum of metadata quality: defining, expressing, exploiting*. Retrieved from http://www.ecommons.cornell.edu/handle/1813/7895

Feigenbaum, L., Herman, I. et al. (2007). *The semantic web in action*. Retrieved from http://www.thefigtrees.net/lee/sw/sciam/semantic-web-in-action

Harping, P. (2014). *The Getty vocabularies and Linked Open Data: introduction and editorial perspective*. Retrieved from http://www.getty.edu/research/tools/vocabularies/Linked_Data_Getty_Vocabularies.pdf

Hassanzadeh, O. (2011). Introduction to Semantic Web Technologies & Linded Data. Retrieved from http://www.cs.toronto.edu/~oktie/slides/web-of-data-intro.pdf

Herman, I. (2009). Introduction to the Semantic Web (tutorial). Retrieved from http://www.w3.org/2009/Talks/0615-SanJose-tutorial-IH/Slides.pdf

Margaritopoulos, T., Margaritopoulos, M. et al. (2008). *A conceptual framework for metadata quality assessment*. Retrieved from http://dcpapers.dublincore.org/pubs/article/viewFile/923/919

Truyen, F. (2012). *D2.1 Content seminar proceedings*. Retrieved from http://www.europeana-photography.eu/index.php?en/115/deliverables

Van Steen, N. (2013). *D4.1 EuropeanaPhotography Vocabulary Definition*. Retrieved from http://www.europeana-photography.eu/index.php?en/115/deliverables