Report on Preparing Europeana for IIIF involvement Task Force

Appendix D: Implementation Examples D.3: The University of Heidelberg Case Study

Pierre-Edouard Barrault   2017

Content

Content

Appendix D.3: Implementation of IIIF collections in Europeana - The Universitätsbibliothek Heidelberg study case

Pierre-Edouard Barrault, Operations Officer, Europeana Foundation

1. Introduction

We have worked on improving the results of mapping process from the METS to EDM schemas, for metadata records associated with cultural heritage objects. We chose to present the case of the Universitätsbibliothek Heidelberg, which was founded in 1386 and is Germany's oldest university and one of the world's oldest surviving universities. Its magnificent collection of about 25000 records contains parchments and early printed books from the 14th century until Modern Age, or books, magazines and newspapers from the 19th and onward, in various languages including French, German, Italian or Spanish. It is without any doubt a solid accomplishment for an old book digitization project, demonstrating the value added from respecting both content integrity thanks to high digitization standards coupled with the IIIF framework, and informational quality through rich, highly-structured, open data. In addition, the institution proposes its collection under the Creative Commons - Attribution, ShareAlike (BY-SA) open license, allowing for free re-use.

Previously to this experiment, the collection of the Universitätsbibliothek Heidelberg in Europeana was based on harvests of the OAI-PMH server of the institution exposing metadata under the ESE schema. We used to receive limited metadata records in which multiple values for a given field were mapped in only one instance of this field. Fields such as dc:date, dc:type and dc:subject were biased. Having single strings introduced in a single metadata field with separators prevents the Europeana automatic semantic enrichment from detecting the appropriate string and enriching the record based on the matching string. Other shortcomings were based on the lack of language attributes or relevant hierarchical data.

2. IIIF implementation

We focused our work on this specific provider with the hope for improving its collections, which were already available in Europeana Collections, with the IIIF features they had implemented on their side. This open technological framework can be implemented within content management systems to enable deep visualisation features (zoom, crop, effects), and to make image sharing easier on the Web.

The main target of this experiment was about implementing IIIF metadata elements, which were not present in previously submitted data from this institution to the Europeana Collections database. After investigating the available data on the institution's side, we decided to harvest METS records as this was a much richer metadata source, regarding both IIIF core elements and metadata range and quality.

3. Data quality

Further work was done in the mapping, starting with the refinement and improvement of the overall data quality by relying on Linked Open Data resources from the GND authority vocabulary maintained by the German National Library, which were available in the original METS records. Agents roles and attributions were also implemented based on MARC Relators codes originally embedded in the METS records, such as *"aut"* that represents *"Author"*. The codes were used to identify the agents as creators or contributors, and then were normalized into strings to be directly incorporated into the resulting EDM records as additional metadata.

Finally, hierarchical relationships that were not made available in the original conversion were represented in the new metadata. This enabled a better experience for end users thanks to the display of a widget dedicated to browse hierarchical resources by following their cardinality or their appartenance.

4. Results

The first outcome of this work is an extensive report presenting this study case, standing as data guidelines available in the Pro section of Europeana Collections. However, our results rely on both qualitative and quantitative achievements.

The overall data improvement empowers the Europeana users - creatives, searchers, curious - with higher quality results, allowing them to tailor their experience even further from the main public access. Specific data reuse or data mining scenarios also benefit from such experiment, thanks to the Europeana's REST API. In addition, the compatibility with the IIIF framework ensure a seamless user experience carried out through extended visualisation features. This can be transposed into more advanced applications by directly reusing the aggregated IIIF metadata from Europeana, e.g. within Digital Humanities visualisation projects.

Finally, the updated datasets didn't necessarily grow in size, records wise. But instead of the former 1 thumbnail per record rule (for about 25K records), the newly added IIIF metadata enables the Europeana's viewer to fetch now more than 3.5M high-resolution pictures (+1600px wide) from all the connected JSON manifests.