



Oliver Pohl

Sustainable Metadata Management for Cultural Heritage Image Data using XMP

In:

Proceedings of the 8th Conference of Japanese Association for Digital Humanities.
Hitotsubashi Hall, Tokyo, September 9-11, 2018

S. 149-151

<https://conf2018.jadh.org>

Online edition: ISSN 2432-3144

Print edition: ISSN 2432-3187

Persistent Identifier: urn:nbn:de:kobv:b4-opus-29149

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.



Sustainable Metadata Management for Cultural Heritage Image Data using XMP

Oliver Pohl¹

When managing large quantities of data, it is a common solution to utilize a centralized data management software to forge a connection between metadata and the data objects themselves. In case of text-based objects without any attached metadata, it is easy for humans to contextualize these objects by recognizing patterns such as filenames, titles, authors etc. This task becomes a challenge when dealing with non-text-based objects like images in the cultural heritage domain. Without metadata or expert knowledge, it becomes difficult to estimate the creation date of a painting or tell the name of its painter. Thus, the ability to contextualize data depends on whether there is a working connection between the metadata store and the data object itself. This connection fails as soon as the file is moved on the file system without having these changes also applied in the corresponding data base, or when the file is shared without a reference to its original location. This paper presents an approach to overcome that type of co-dependency by utilizing XMP to embed cultural heritage metadata directly into image files to ensure their location-independent long-term preservation. The “Corpus Vitrearum Medii Aevi” Germany (CVMA) project serves as an example use-case.

CVMA Germany is a joint project by the Berlin-Brandenburg Academy of Sciences and Humanities and Academy of Sciences and Literature Mainz, with teams sitting in Berlin, Potsdam, Mainz and Freiburg. The focus of CVMA is to catalog and document stained glass windows situated in German churches built throughout the medieval and early modern period. Furthermore, the project publishes collections of stained glass photographs including metadata in an online image archive (CVMA Deutschland, 2018). The most fundamental data are uncompressed high-resolution photographs of the church windows. For proper documentation of these photos, the CVMA Germany developed its own metadata schema (CVMA Deutschland, 2016), building upon Dublin Core (Weibel et al., 1998) and IPTC (2014), and implementing its own CVMA namespace, which focuses on stained glass research in art history. When entering metadata, the project also makes use of authority files such as the GND (Deutsche Nationalbibliothek, 2018) for storing information about creators and donors, or geonames to avoid any unambiguous data and create compatibility with the linked open data cloud.

To maintain readability of its image files in the future, the CVMA embeds metadata directly into the files, thereby ensuring their readability across platforms without running into any co-dependency with an additional management layer, thus creating a union of the image and its metadata.

Regarding long-term preservation, the CVMA is guided by patterns employed in the domain of text-based document preservation of PDF files. When creating PDF documents, any fonts, images and additional metadata are embedded into the document itself so users can access the contents as intended without the use of third party software or data (Adobe Systems, 2006).

In both cases, CVMA and PDF, metadata are embedded using the eXtensible Metadata Platform (XMP) (Adobe Systems, 2012). When using XMP, it is possible to define your own metadata schemata and embed metadata as RDF/XML into any file without having any effect on its readability (Bright, 2006). For instance, when embedding metadata into a TIFF file, image viewers with XMP capabilities can show these additional data, while viewers without these capabilities remain unaffected and show the image as if it would not contain any additional data. XMP has seen a surge of usage in digital asset management systems (Regli, 2009) because of the easy way of jointly transporting files and metadata.

There are different workflows on how to embed and read XMP metadata throughout the different project branches in Germany. The Freiburg team utilizes

¹ Berlin-Brandenburg Academy of Sciences and Humanities

exiftoolGUI (2018), a graphical user interface wrapper for the EXIF- and XMP manipulation command line tool exiftool (Phil Harvey, 2018). Both are free open source software products and can be configured to make use of custom metadata standards. However, these tools are only helpful when editing small sets of data or when searching across a small pool of files, since they lack any management capabilities.

The team in Potsdam uses the proprietary application FotoStation (Fotoware, 2018). In contrast to exiftoolGUI, FotoStation comes with management and search features. Moreover, it is very easy to employ custom metadata standards and configure metadata input forms by drag and drop, making it very user friendly. Unfortunately, there are only few options when it comes to software that is capable of embedding XMP metadata into files with the option of utilizing custom metadata standards.

When it comes to online display, the images are sent to the CVMA digital online image archive. On ingest, the back-end extracts metadata into a relational data base for efficiency purposes.

On the one hand the utilization of XMP comes with some challenges and obstacles. For instance, modifying just a single metadata field in a large TIFF file can be very resource intense, since the whole image needs to be re-written. Additionally, although the metadata is now embedded into the images, it takes strict file management guidelines that determine how to name and where to store files in order to avoid creating a messy file collection. Moreover, it is still a challenge to integrate XMP-image-files into research data stores that handle heterogeneous types of data from varying domains of the sciences and humanities (Grunzke et al., 2016).

Utilizing XMP has proven to be resourceful with the CVMA project for multiple reasons. First, ensuring long-term preservation has become a matter of just backing up the files. Secondly, due to its RDF/XML based nature, XMP makes it easy to extract and map metadata into different metadata schemata for easier re-use in other online image archives. Additionally, it also becomes easily feasible to expose the metadata to the linked open data cloud. As for now, XMP had a very positive effect on the workflows within the CVMA project by consolidating the (re-)use of existing metadata vocabularies, by using authority files and creating a new metadata standard for church windows for other related projects to re-use.

Bibliography

- Adobe Systems (ed).** (2006). PDF Reference: sixth edition. https://www.images2.adobe.com/content/dam/acom/en/devnet/pdf/pdf_reference_archive/pdf_reference_1-7.pdf (accessed 4 May 2018).
- Adobe Systems (ed).** (2012). XMP Specification Part 1: Data Model, Serialization, and Core Properties. <https://www.images2.adobe.com/content/dam/acom/en/devnet/xmp/pdfs/XMP%20SDK%20Release%20cc-2016-08/XMPSpecificationPart1.pdf>.
- Bright, J.** (2006). First steps: XMP. *Journal of Digital Asset Management*, 2(3–4): 198–202 doi:10.1057/palgrave.dam.3650025.
- CVMA Deutschland** (2016). Corpus Vitrearum Deutschland: XMP Metadatenpezifikation Version 1.1 <http://www.corpusvitrearum.de/cvma-digital/spezifikationen/cvma-xmp/11.html> (accessed 7 May 2018).
- CVMA Deutschland** (2018). Digitales Bildarchiv - Öffentliche Betaversion: Corpus Vitrearum Deutschland <http://www.corpusvitrearum.de/cvma-digital/bildarchiv.html> (accessed 7 May 2018).
- Deutsche Nationalbibliothek** (2018). Gemeinsame Normdatei http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html (accessed 7 May 2018).
- exiftoolGUI** (2018). exifToolGUI <http://u88.n24.queensu.ca/~bogdan/> (accessed 7 May 2018).
- Fotoware** (2018). FotoStation | Blazing fast photo management software <https://www.fotostation.com/> (accessed 7 May 2018).
- Grunzke, R., Hartmann, V., Jejkal, T., Prabhune, A., Herold, H., Deicke, A., Hoffmann, A., Schrade, T., Meinel, G. and Herres-Pawlis, S.** (2016). Towards a metadata-driven multi-community research data management service. Rome, Italy.

- IPTC** (2014). *IPTC - NAA Information Interchange Model Version 4*. International Press Telecommunications Council <http://www.iptc.org/std/IIM/4.2/specification/IIMV4.2.pdf> (accessed 7 May 2018).
- Phil Harvey** (2018). ExifTool <https://www.sno.phy.queensu.ca/~phil/exiftool/> (accessed 7 May 2018).
- Regli, T.** (2009). The state of digital asset management: An executive summary of CMS Watch's Digital Asset Management Report. *Journal of Digital Asset Management*, 5(1): 21–26 doi:10.1057/dam.2008.49.
- Weibel, S., Kunze, J., Lagoze, C. and Wolf, M.** (1998). Dublin Core Metadata for Resource Discovery. <http://www.rfc-editor.org/info/rfc2413> (accessed 4 May 2018).