

Nachhaltiges Management von Bildmetadaten mit XMP, exiftool und Fotostation

Pohl, Oliver

opohl@bbaw.de

Berlin-Brandenburgische Akademie der Wissenschaften,
Deutschland

Schrade, Torsten

torsten.schrade@adwmainz.de

Akademie der Wissenschaften und Literatur Mainz,
Deutschland

Das Corpus Vitrearum Deutschland als Use Case

Das interakademische Vorhaben zur mittelalterlichen und frühneuzeitlichen Glasmalereiforschung „Corpus Vitrearum Medii Aevi“ (CVMA) in den Arbeitsstellen Potsdam und Freiburg steht derzeit der Herausforderung gegenüber, die im neu eingerichteten CVMA-Online-Bildarchiv hinterlegten Bilddateien samt Metadaten für die Langzeitarchivierung und -lesbarkeit vorzubereiten.

Das Hauptziel des CVMA ist es, alle Glasmalereien des Mittelalters fotografisch zu erfassen, zu dokumentieren und zu edieren, um nicht nur das kulturelle Erbe zu bewahren und ins digitale Zeitalter zu überführen, sondern auch räumliche Distanzen zu überbrücken und die Glasmalereifotografien samt ihrer Dokumentation der Öffentlichkeit zugänglich zu machen.

Für die zielgerichtete Dokumentation hat das CVMA ein eigenes Metadatenschema verfasst, welches auf etablierte Schemata wie Dublin Core (Weibel et al. 1998) und IPTC (IPTC 2014) aus der professionellen Fotografie aufbaut und diese mit einem speziell auf die wissenschaftlichen Bedarfe der Glasmalerei ausgerichteten eigenen Namensraum erweitert. Dabei werden die Metadaten in das zugehörige Bild integriert, um so zusätzliche Abhängigkeiten von Datenbank- oder anderer Verwaltungssoftware zu vermeiden. Dadurch wird die plattformabhängige Nutzung und maschinelle Interpretierbarkeit des Datenbestands gefördert, was nötig ist, damit die Daten auch in Zukunft von Mensch und Maschine nachgenutzt werden können (Library of Congress und National Science Foundation 2003).

XMP

Zur Lösung der Herausforderungen im Bereich Langzeitarchivierung orientiert sich das CVMA an den bereits bestehenden Lösungen aus der digitalen Dokument-Langzeitarchivierung von PDF-Dateien (Braun et al. 2010; ISO 2011), bei welcher sämtliche genutzte Abbildungen, Schriftarten und Metadaten mit in das Dokument eingebettet werden, sodass diese Datei letztendlich alle Inhalte originalgetreu ohne Zuhilfenahme dritter Anwendungen anzeigen kann.

Der PDF-Standard (Adobe Systems 2006) nutzt für die Integration von Metadaten die eXtensible Metadata Platform (XMP), mit welcher die zum Dokument zugehörigen Metadaten als RDF/XML kodiert in das Dokument eingebettet werden (Adobe Systems 2005; Bright 2006) ohne die digitale Interpretierbarkeit der zu beschreibenden Datei zu beeinflussen. Wird die Datei von einem Programm geöffnet, welches den XMP-Standard nicht unterstützt, wird diese Datei dennoch angezeigt. Umgekehrt ist es möglich, dieselbe Datei mit einem Texteditor zu öffnen, um die XMP Daten in Reintextform anzuzeigen.

Die Nutzung von XMP hat sich bereits auf mehreren Ebenen etabliert. Zu einem werden beim Abspeichern von Digitalfotos EXIF- und IPTC-Daten von Digitalkameras als XMP in die erzeugten Bilddateien geschrieben (Tescic 2005) und können so durch Bildverwaltungssoftware wie Adobe Bridge oder FotoWare Fotostation gelesen und verwaltet werden. Des Weiteren wird XMP in der Digital Asset Management Community für die einfache Verwaltung von Dateibeständen bevorzugt genutzt (Regli 2009). Durch die Einheit von Dokument und Metadaten verringert sich der Aufwand bei einem Datentransfer sämtlicher Abhängigkeiten auf das Kopieren des Datenbestands (Binder 2006) und hält gleichzeitig die Möglichkeit offen, Datenbanken aus den vorliegenden XMP-Daten zu erstellen und auf die XMP-Daten aufbauende Tools zu implementieren (Abdillah 2013). Der Kern von XMP liegt inzwischen auch als ISO-Standard vor. Das CVMA-Onlinebildarchiv dient hierfür als Beispiel, da sich die relationale Datenbank der Online-Plattform aus den XMP-Daten des verfügbaren Bildbestandes speist.

Ein weiterer Faktor für die internationale Verbreitung des XMP-Standards ist seine Erweiterbarkeit. Demnach ist es möglich, eigene Metadatenschemata als XMP an Dateien anzuhängen. Zwar ist dieser Anhang als RDF/XML kodiert, allerdings besteht die Restriktion kein RDFS und OWL nutzen zu können (Eriksson 2007). Das einzig gültige Subjekt in den RDF-Tripeln in XMP ist die zu beschreibende Ressource, also die Datei selbst. Es ist also nicht möglich Ontologien wie im herkömmlichen Semantic-Web-Kontext zu verwenden. Für das CVMA-Bildarchiv genügt es jedoch, die beschreibenden Metadaten für die einzelnen Bildressourcen anzulegen und so den Funktionsumfang des XMP-Standards voll auszuschöpfen.

XMP Workflow beim CVMA

In Deutschland haben die zwei CVMA-Arbeitsstellen in Potsdam und Freiburg unterschiedliche Workflows implementiert, um XMP-Daten in die Bilddateien einzupflegen. Beide benutzen zwar dasselbe Metadatenschema, allerdings unterscheiden sich die Tools zur Eingabe der Metadaten und die damit verbundenen Arbeitsabläufe.

Um sich für das Vorhaben der deutschen CVMA-Arbeitsstellen zu eignen, muss eine entsprechende Metadatenbearbeitungssoftware folgende Kriterien erfüllen: a) Die Software muss fähig sein, XMP-Metadaten zu schreiben und auszulesen. b) Die Software muss die Anzeige und Metadatenmanipulation von gängigen Bilddateiformaten wie TIFF und JPG unterstützen. c) Die Software muss die Möglichkeit zur Konfiguration eines eigenen Metadatenschemas anbieten. d) Optional: Die Software bietet die Möglichkeit zur lokalen Recherche mit dem im c) angelegten Metadatenschema.

Das Team in Freiburg nutzt die Freeware *exiftoolGUI*, welche eine konfigurierbare graphische Oberfläche für das Kommandozeilentool *exiftool* bietet. *exiftool* selbst ist ein Programm zum Auslesen und zur Manipulation von Bildmetadaten und unterstützt das Lesen und Schreiben von XMP-Daten. Über eine Perl-Konfigurationsdatei kann das anzuwendende Metadatenschema für Datenmanipulationen via *exiftool* und somit auch *exiftoolGUI* festgelegt werden. Die inhärenten Vorteile von *exiftool* und *exiftoolGUI* sind deren Offenheit, Konfigurierbarkeit und Plattformunabhängigkeit. Die Nutzung von *exiftoolGUI* für größere Bildbestände ist jedoch eher unkomfortabel, da dieses Tool ausschließlich zur Metadatenmanipulation und nicht als Bildverwaltungssoftware ausgelegt ist.

Die CVMA-Arbeitsstelle in Potsdam nutzt hingegen die proprietäre Software *FotoStation* von *FotoWare* (Version 8.0). Zwar liegt der Quellcode dieser Software nicht offen, jedoch besteht auch bei diesem Tool die Möglichkeit zur Konfiguration eigener Metadatenschemata. *FotoStation* bietet deutlich mehr Bedienkomfort im Vergleich zu *exiftoolGUI* und kann als Bildverwaltungssoftware genutzt werden, welche den lokalen Bildbestand samt Metadaten für Recherchen indexiert und die Bilder auf einem digitalen Lichttisch darstellt. Für die Konfiguration des Metadatenschemas bietet *FotoStation* graphische Editoren, mit welchem nicht nur die zu verwendenden Namensräume und Felder definiert werden können, sondern auch das Metadateneingabeinterface frei gestaltet werden kann. Dabei ist es möglich kontrollierte Vokabulare zu anzulegen und die Nutzereingaben über reguläre Ausdrücke validieren zu lassen. Weiterhin lässt sich die Gesamtkonfiguration von *Fotostation* leicht ex- und importieren, sodass sämtliche Informationen zum verwendeten Metadatenschema, Editorinterface, Vorschlagslisten, Bearbeitungsaktionen und Rechercheeinstiege leicht innerhalb des Teams ausgetauscht und aktualisiert werden können.

Ziele des Workshops

Die Teilnehmer an diesem Workshop sollen einen Einstieg in die nachhaltige Bildmetadatenverwaltung mit XMP erhalten. Als Übung wird mit den Teilnehmern gemeinsam zuerst ein Beispiel-Metadatenschema definiert. Anschließend werden die Teilnehmer in die Benutzung von *exiftoolGUI* und *FotoStation* eingeführt, um daraufhin das zuvor definierte Metadatenschema für beide Tools umzusetzen und zu testen. Weiterhin wird das Abfragen von XMP-Metadaten mit dem *exiftool* über die Kommandozeile geübt, um so das Potential von XMP für die Erstellung von Tools, Services oder Datenbanken für Bildbestände aufzuzeigen.

Das CVMA-Deutschland erhofft sich, Interessierten und ähnlichen Projekten den Einstieg in die Erstellung und Manipulation von XMP-Daten zu erleichtern und gleichzeitig weitere Lösungsansätze und Tools für den Umgang mit XMP-Daten kennenzulernen.

Für die Teilnahme an diesem Workshop werden lediglich Grundkenntnisse in XML vorausgesetzt. Die Kenntnisse einer Programmier- oder Skriptsprache sind zwar von Vorteil, aber nicht erforderlich. Den Teilnehmern entstehen für die Nutzung von Software während des Workshops keine Kosten, da *exiftool* und *exiftool GUI* kostenfrei sind und *Fotostation* für 14 Tage kostenfrei getestet werden kann. Der Workshop ist für maximal 20 Teilnehmer ausgelegt.

Forschungsinteressen

Oliver Pohl ist wissenschaftlicher Mitarbeiter bei TELOTA an der Berlin-Brandenburgischen Akademie der Wissenschaften und betreut dort das CVMA, das Langzeitvorhaben *Corpus Coranicum* sowie das Kooperationsprojekt *Paleocoran* mit dem *Collège de France*. Seine Forschungsinteressen sind Webtechnologien und Semantic Web Technologien für digitale Geisteswissenschaften als auch maschinelle Übersetzung.

Torsten Schrade ist Leiter der Digitalen Akademie der Mainzer Akademie der Wissenschaften und der Literatur und beschäftigt sich vorrangig mit dem Forschungsdatenmanagement und dem Einsatz von Webtechnologien für die geisteswissenschaftliche Grundlagenforschung. Daneben zählen Methoden und Programmierparadigmen der agilen Softwareentwicklung sowie die Technologien des Semantic Web zu seinen Forschungsinteressen.

Fußnoten

1. <http://www.corpusvitrearum.de>
2. <http://www.corpusvitrearum.de/cvma/1.1/> (Stand: 25.08.2016)
3. <http://www.adobe.com/de/products/bridge.html>

4. <http://www.fotoware.com/products/fotostation-client>
5. http://www.iso.org/iso/catalogue_detail?csnumber=57421
6. <http://u88.n24.queensu.ca/~bogdan/> bzw. <https://hvdwolf.github.io/pyExifToolGUI/>
7. <http://www.sno.phy.queensu.ca/~phil/exiftool/>

Bibliographie

Abdillah, Leon Andretti (2013): „PDF Articles Metadata Harvester“, in: *arXiv Preprint arXiv:1301.6591*. <http://arxiv.org/abs/1301.6591> [letzter Zugriff 25. August 2016].

Adobe Systems (2005): *Extensible Metadata Platform (XMP) Specification*. Adobe Systems <https://partners.adobe.com/public/developer/en/xmp/sdk/XMPspecification.pdf> [letzter Zugriff 25. August 2016].

Adobe Systems (2006): *PDF Reference, Sixth Edition: Adobe Portable Document Format Version 1.7* http://www.images.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/pdf_reference_1-7.pdf [letzter Zugriff 25. August 2016].

Binder, Jennifer (2006): „Exchanging Assets and Metadata across Platforms“, in: *Journal of Digital Asset Management* 2 (5): 215–18 10.1057/palgrave.dam.3650045.

Braun, Kim / Buddenbohm, Stefan / Dobratz, Susanne / Herb, Ulrich / Müller, Uwe / Pampel, Heinz / Schmidt, Birgit (2010): *DINI-Zertifikat Dokumenten-Und Publikationsservice 2010* <https://pub.uni-bielefeld.de/publication/2491543> [letzter Zugriff 25. August 2016].

Bright, Jason (2006): „First Steps: XMP“, in: *Journal of Digital Asset Management* 2 (3–4): 198–202 10.1057/palgrave.dam.3650025.

Eriksson, Henrik (2007): „The Semantic-Document Approach to Combining Documents and Ontologies“, in: *International Journal of Human-Computer Studies* 65 (7): 624–639.

IPTC (2014): *IPTC - NAA Information Interchange Model Version 4.2*. International Press Telecommunicatoins Council <http://www.iptc.org/std/IIM/4.2/specification/IIMV4.2.pdf> [letzter Zugriff 25. August 2016].

ISO (2011): *ISO 19005-1:2005 - Document Management -- Electronic Document File Format for Long-Term Preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)* http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38920 [letzter Zugriff 25. August 2016].

Library of Congress / National Science Foundation (2003): *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation* http://www.digitalpreservation.gov/documents/about_time2003.pdf [letzter Zugriff 25. August 2016].

Regli, Theresa (2009): „The State of Digital Asset Management: An Executive Summary of CMS Watch's

Digital Asset Management Report“, in: *Journal of Digital Asset Management* 5 (1): 21–26.

Tesic, Jelena (2005): „Metadata Practices for Consumer Photos“, in: *IEEE MultiMedia* 12 (3): 86–92.

Weibel, Stuart / Kunze, John / Lagoze, Carl / Wolf, Misha (1998): *Dublin Core Metadata for Resource Discovery* <http://www.rfc-editor.org/info/rfc2413> [letzter Zugriff 25. August 2016].