

# Alte Drucke in neuem Gewand

## Ein Beispiel für den erfolgreichen Einsatz der freien Digitalisierungsplattform Goobi an der UB Mannheim

Christian Hänger, Irene Schumm und Stefan Weil

### Kommerzielle oder Open-Source-Software

Im vorliegenden Beitrag wird exemplarisch der Umstieg von einer kommerziellen Software zu einer Open-Source-Software beschrieben. Am Beginn eines jeden IT-Projekts steht die grundsätzliche Frage, ob eine kommerzielle Software oder eine Open-Source-Software eingesetzt werden soll. Beim Einsatz einer kommerziellen Software hat ein Kunde den Vorteil, das Produkt nach dem Prinzip „Plug and Play“ einsetzen zu können, ohne selbst aufwendige Anpassungen und Implementierungen vornehmen zu müssen. Rechte und Pflichten der beiden Seiten sind vertraglich geregelt und die Supportanfragen werden von der Firma übernommen. Auf der Negativseite stehen die hohen Kosten, die kommerzielle Produkte mit sich bringen. Dazu kommen häufig noch Einschränkungen in der Umsetzung von Anforderungen der eigenen Einrichtung, da eine Firma ein einheitliches Produkt für einen großen Kundenkreis anbietet und auf individuelle Wünsche einzelner Kunden nicht eingehen möchte. Beim Einsatz einer Open-Source-Software kehren sich Vor- und Nachteile um. Der Anwender vermeidet hohe einmalige und dauerhafte Kosten für den Kauf und die Pflege der Softwarelizenz, muss aber eigenes Personal für die Installation, Weiterentwicklung und den Support vor Ort aufwenden. Hinsichtlich der Anpassung der Software an die eigenen Anforderungen besteht eine größere Freiheit als bei einem kommerziellen Produkt, da eine Weiterentwicklung häufig lediglich mit der Community abgestimmt werden muss und die Umsetzung dann mit eigenen Mitteln realisiert werden kann.

Bei dem konkreten Fall der UB Mannheim ging es um eine Software zur Unterstützung des Digitalisierungsworkflows und zur Präsentation der Digitalisate. Die UB Mannheim hatte sich für diese Anwendung zunächst für eine kommerzielle Software entschieden und zu Beginn des Jahres 2012 die Software *Visual Library* der Firma *Semantics* eingeführt. Mit *Visual*

*Anfang 2012 begann die Universitätsbibliothek Mannheim damit, ihre Alten Drucke und weitere Digitalisate mit Hilfe einer speziellen Software zu erschließen und im Internet zu veröffentlichen. Dabei setzte sie zunächst auf ein kommerzielles Produkt, die Software Visual Library von semantics Kommunikationsmanagement GmbH. 2014 fiel die Entscheidung, auf eine freie Softwarelösung zu wechseln. Ende September 2014 ging die freie Digitalisierungsplattform Goobi in den Produktivbetrieb.*

*Since 2012 Mannheim University Library uses specialized software to make their digitised books accessible and to present them in the Internet. Initially the commercial software Visual Library (semantics Kommunikationsmanagement GmbH) was used. In 2014 this software was replaced by the free digitisation workflow solution Goobi.*

Library stand eine stabile und überzeugend gestaltete Software zur Verfügung, die alle Aspekte des Digitalisierungsworkflows in vollem Umfang unterstützte. Dies wurde durch die kompetenten Mitarbeiter der Firma *Semantics* komplementiert, die die UB Mannheim bei der Einrichtung der erforderlichen Geschäftsprozesse berieten und bei der Einhaltung der relevanten Standards unterstützten. Ohne die Software *Visual Library* und die Unterstützung der Firma *Semantics* wäre es der UB Mannheim nicht möglich gewesen, in so kurzer Zeit das Wissen für die Durchführung großer Digitalisierungsprojekte aufzubauen und das DFG-Projekt *Digitalisierung der Reiseberichte der Sammlung Desbillons* erfolgreich durchzuführen. Allerdings wurde hier bereits ein Nachteil dieser Lösung offensichtlich, da mit dem Auslaufen der Finanzierung durch die DFG die hohen Lizenz- und Pflegekosten von der UB Mannheim selbst erbracht werden mussten. Es ging letztlich darum, auch kleinere Digitalisierungsprojekte ohne eine Fremdfinanzierung durchzuführen, ohne dass der Sachmittletat der UB Mannheim mit hohen Fixkosten belastet wird.

Daher bestand das Erfordernis, eine kostenfreie Software auszuwählen, die über eine große Anwender-

community verfügt und die als Open Source angeboten wird, so dass eigene Anforderungen in Absprache mit den anderen Nutzern realisiert werden können. Nach einer intensiven Marktsichtung, der Evaluation alternativer Produkte und einer Testinstallation fiel die Entscheidung für Goobi, das die aufgeführten Kriterien vollständig erfüllt.

### Was genau ist Goobi?

Goobi ist ein freies Softwarepaket für die Digitalisierung und besteht aus den beiden Modulen Goobi.Production und Goobi.Presentation.

Goobi.Production unterstützt große Teile des Digitalisierungsworkflows vom Anlegen eines Titels bis zur Freischaltung des erschlossenen Digitalisats im Internet. Die Anwendung ist in Java realisiert und wird auf der Basis von Apache Tomcat und einer Datenbank (MySQL, MariaDB oder andere kompatible Datenbank) betrieben. Die Nutzung erfolgt über ein Webinterface, das mit allen gängigen Webbrowsern funktioniert, und ist damit betriebssystemunabhängig.

Goobi.Presentation ist eine Erweiterung für das Content-Management-Framework TYPO3 und unterstützt die Präsentation der Digitalisate im Internet. Die Funktionalität umfasst unter anderem die Seitenvorschau und eine sehr mächtige Seitenanzeige, Anzeige der Metadaten eines Werkes, Auflisten der Sammlungen und der darin enthaltenen Werke mit umfangreichen Browsing-Funktionen, eine Solr-basierte Suchfunktion und Download-Möglichkeiten für PDF-Dateien. Darüber hinaus implementiert die Erweiterung auch eine Schnittstelle (Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH) für den automatisierten Zugriff auf alle Daten durch zentrale Dienste der DNB und anderer Einrichtungen.

Seit 2012 werden die beiden Softwaremodule unter der Trägerschaft des Vereins Goobi. Digitalisieren im Verein e.V.<sup>1</sup> weiterentwickelt. Für beide Module der Software Goobi gibt es Entwicklerplattformen, die den Wissensaustausch innerhalb der Community unterstützen und eine Release-Kontrolle ermöglichen.

Die Software wird an rund 50 Institutionen mit Schwerpunkt in Deutschland, aber auch in Österreich, den Niederlanden, Großbritannien und Spanien, eingesetzt (Stand Anfang 2015).

### Installation und Einführung von Goobi

Häufig wird die Installation und Einführung von Open-Source-Anwendungen von dem Anwender in Eigenleistung erbracht. Bei Goobi besteht die erfreuliche Situation, dass zwei Firmen ihre Dienste für die Soft-

ware anbieten. Daher konnte die UB Mannheim auf das Wissen der Firma Zeutschel GmbH zurückgreifen und einen Teil der Arbeiten bei der Einführung der Software durch diese Firma durchführen lassen. Insbesondere die Durchführung von Schulungen für Endanwender und Administratoren war eine große Unterstützung. Die Mitarbeiterinnen der Abteilung Medienbearbeitung wurden dabei in die Basisfunktionen des Systems eingeführt und sind dadurch in der Lage, den Digitalisierungsworkflow vom Scanprozess über die formale Erschließung bis zur Freigabe eines digitalisierten Werks in Goobi abzubilden. In gleicher Weise unterstützte Zeutschel die Administratoren bei der Konfiguration des Systems, da die technische Dokumentation von Goobi.Production sehr lückenhaft ist und das erforderliche Wissen ohne fremde Hilfe nur im Trial- und Error-Verfahren erworben werden kann.

Die übrigen Arbeiten erbrachte die UB Mannheim selbst. Die Installation der Software, die typischerweise auf Linux-Servern vorgenommen wird, verlief erfolgreich und beanspruchte einen Zeitaufwand von ca. einem halben Tag. Diese Tatsache bestätigte die Auswahl der UB Mannheim, da die geringe Qualität einer Software sich häufig schon zeigt, wenn eine Erstinstallation nicht ad hoc möglich ist. An der UB Mannheim laufen Goobi.Production und Goobi.Presentation auf virtuellen Servern unter der Debian GNU Linux Version Jessie.

Da mehrere Digitalisierungsprojekte ständig neue Daten lieferten und die Medienbearbeitung möglichst rasch mit Goobi.Production arbeiten sollte, hatte dieses zunächst die höchste Priorität. Bei den ersten Schritten und späteren Detailfragen konnten wir auf die Unterstützung durch die Firma Zeutschel zurückgreifen, wurden dann aber auch zunehmend selbständig und verstanden immer besser die nicht ganz einfache, wenig dokumentierte Konfiguration. Trotzdem dauerte dieser Prozess einige Wochen.

Mit zunehmenden Ergebnissen aus Goobi.Production stieg natürlich der Wunsch, diese im Internet zu präsentieren, also Goobi.Presentation in Betrieb zu nehmen. Auch die Webpräsenz der Universitätsbibliothek ist mit TYPO3 realisiert, und so war die Anpassung an unsere Stilvorgaben zwar zeitaufwändig, aber gut zu realisieren. Mitten in der Inbetriebnahme stellten wir aber zufällig fest, dass die Version 4.5 von TYPO3, die wir für die Webpräsenz und jetzt auch für Goobi.Presentation verwendeten, abgekündigt war: nur noch bis Anfang 2015 sollte es dafür Sicherheitsupdates geben. Wir standen also vor der Frage, ob wir noch vor der offiziellen Freischaltung von Goobi.Presentation auf die Nachfolgeversion 6.2 wechseln sollten oder

<sup>1</sup> <http://www.goobi.org/>

Westerstrasse 114-116 | D-28199 Bremen  
fon: (0421) 50 43 48 | fax : (0421) 50 43 16

## Erwerbungspartner, mit denen Sie rechnen können

**Flexibel**

**Erfahren**

**Innovativ**

**Konditionsstark**

**Serviceorientiert**

**Engagiert**

**Klar**



ob wir das noch ein paar wenige Monate aufschieben. Hilfreiche Vorbilder gab es keine, denn alle anderen Anwender von Goobi setzten noch ältere Versionen von TYPO3 ein. Ein Versuch, für den wir uns maximal zwei Tage zugestanden, sollte zeigen, ob der Versionswechsel funktionierte. Tatsächlich konnte die neue Version innerhalb eines Tages zum Laufen gebracht werden – leider mit einem schwerwiegenden Mangel: es funktionierte alles bis auf die visuelle Darstellung der Digitalisate. Im HTML-Quellcode konnte man sehen, dass das Bild darin enthalten war, nur angezeigt wurde es eben nicht. Durch Vergleichen mit funktionierenden Webseiten fanden wir schließlich heraus, dass das neue TYPO3 den Dokumententyp für alle Webseiten auf HTML5 umgestellt hat – eine durchaus sinnvolle Änderung, die aber verantwortlich für unser Anzeigeproblem war. Mit diesem Wissen war es dann nur noch etwas Fleißarbeit, die Webseite mit der Seitenansicht HTML5-kompatibel umzugestalten und eine funktionsfähige Website mit Goobi.Presentation zu erhalten. Hier zeigt sich deutlich, dass die Einführung einer Open-Source-Software den Anwender vor viele unerwartete Herausforderungen stellen kann, und daher Abweichungen vom ursprünglichen Projektplan häufig auftreten. Zur Lösung dieser Schwierigkeiten ist auch teilweise erhebliches Fachwissen notwendig, das entweder bei den eigenen Mitarbeitern oder bei einem externen Dienstleister vorhanden sein muss.

Bis zur heutigen Implementierung dauerte es daher ein wenig. Wir hatten uns vorgenommen, gleich mit einer zweisprachigen Website (deutsch und englisch) zu starten. Das erwies sich als unerwartet schwierig und kostete Zeit. Goobi.Presentation verfügt über Funktionen, die eine Mehrsprachigkeit unterstützen, die aber leider nicht fehlerfrei sind. So war es uns mit den über Goobi gebotenen Mitteln nicht möglich, die Bezeichnungen der Sammlungen für die englische Oberfläche darzustellen.

Noch nicht optimal gelöst ist in Goobi der Übergang zwischen dem Produktionssystem Goobi.Production und der Präsentation in Internet, also Goobi.Presentation. Dabei müssen die Seitenbilder, die üblicherweise als TIFF-Dateien vorliegen, in kompaktere JPEG-Dateien unterschiedlicher Auflösung konvertiert werden. Zusätzlich können PDF-Dateien für die einzelnen Werke erstellt werden, idealerweise mit eingebauten Metadaten, die dem Leser beim Navigieren im Text helfen, indem sie beispielsweise die Kapitel anzeigen. Beides konnten wir mit freier Software (graphicsmagic, pdftk) und kleinen selbstgeschriebenen Skripten realisieren. Im Goobi-Umfeld gibt es dafür einen Content Server, aber dieser wird anscheinend

nur in einer kommerziellen Version aktiv weiterentwickelt und wird auf den offiziellen Goobi-Webseiten gar nicht mehr erwähnt. Nach der dürftigen verfügbaren Dokumentation zum Content Server erschien uns die hausgemachte Lösung ausreichend und schneller realisierbar. Auch die geplante weitere Erschließung der Digitalisate mit OCR lässt sich in unsere Lösung noch integrieren.

### Konfiguration von Goobi

Bevor ein erstes Digitalisierungsprojekt in Goobi.Production starten kann, sind zunächst noch einige Konfigurationen nötig. Da eine Digitalisierung in der Regel im Rahmen von zeitlich und / oder inhaltlich abgeschlossenen Projekten stattfindet, gibt es analog auch in der Logik von Goobi.Production *Projekte*. Zunächst muss also ein Projekt angelegt werden. Dies kann dann bestimmten *Benutzern* zugeordnet werden, d. h. lediglich autorisierte Benutzer können in diesem Projekt Bearbeitungen vornehmen.<sup>2</sup> Des Weiteren lassen sich in Goobi.Production *Produktionsvorlagen* erstellen. Diese legen den Scan-Workflow fest, also die Abfolge der einzelnen *Aufgaben*, die im Digitalisierungsprozess bearbeitet werden müssen. Jedem Projekt muss eine Produktionsvorlage zugeordnet werden, damit für die zu digitalisierenden Werke des Projekts klar ist, wie der Workflow ist. Schließlich sind noch *Regelsätze* zu definieren, in denen festgelegt wird, welche Metadaten und Strukturdaten für die Digitalisate zur Verfügung stehen sollen, die interne Logik der Strukturdaten (beispielsweise kann ein Kapitel Unterkapitel haben, ein Stempel jedoch nicht) sowie Export-Einstellungen. Soll nun ein Titel digitalisiert werden, wird zu dem betroffenen Goobi-Projekt ein *Vorgang* erstellt. Mit der Produktionsvorlage, welche dem Projekt zugeordnet ist, ist definiert, welche Aufgaben im Digitalisierungsworkflow wann und von wem abzarbeiten sind. Schließlich muss für einen Vorgang noch der geltende Regelsatz ausgewählt werden.

### Migration der Altdaten der UB Mannheim

Die neue Webpräsenz mit Goobi.Production sollte nicht nur neu digitalisierte Werke zugänglich machen, sondern auch alle Werke, die bisher mit *Visual Library* erfasst und präsentiert waren. Dazu wurden zunächst alle Metadaten aus dem Altsystem extrahiert. Ein erster Versuch, direkt auf die Dateien auf

<sup>2</sup> Jeder Benutzer kann einer oder mehreren Benutzergruppen zugeordnet werden. So gibt es bspw. die Benutzergruppen „Scanner“, „Medienbearbeiter“, „Administrator“ usw., welche in den jeweiligen Projekten unterschiedliche Rechte bei den abzuarbeitenden Aufgaben haben können.

Startseite | Drucken | English

UNIVERSITÄT MANNHEIM

Universitätsbibliothek Mannheim

UB MANNHEIM

**Sammlungen**  
 Nachlass Otto Selz (113)  
 Sammlung Desbillons (590)  
 Sammlung Mammelsdorf (17)  
 Sammlung Weickum (8)  
 Varia (49)

**Service**  
 Bestandsgeschichte  
 Kurzüberblick  
 Detailsuche

**Links**  
 Universität Mannheim  
 Rechenzentrum  
 Multimedia-Zentrum

**Die historischen Bestände der Universitätsbibliothek Mannheim**  
**Digitale Sammlungen der Universitätsbibliothek Mannheim**

Die Digitalen Sammlungen der Universitätsbibliothek Mannheim präsentieren Alte Drucke und handschriftliche Nachlässe vom 15. bis zum 19. Jahrhundert. Ein Großteil der Bücher sind in lateinischer und französischer Sprache verfasst. Ziel ist, Digitalisate von Objekten anzubieten, die besonders nachgefragt oder im Rahmen von Erschließungs- und Forschungsprojekten erstellt werden.

Suchen nach:

Impressum Datenschutzerklärung © 2015 UB Mannheim

Abbildung 1: Goobi-Portal für die historischen Sammlungen der UB Mannheim

dem Webserver zuzugreifen, war nicht erfolgreich. Glücklicherweise unterstützt aber auch Visual Library Datenzugriffe per OAI-PMH. So war es möglich, mit Hilfe des OAI-Browsers, der bei Debian Teil des Softwarepakets libhttp-oai-perl ist, alle Metadaten als METS/MODS-Dateien zu erhalten. Prinzipiell kann Goobi.Presentation solche Dateien, wie sie auch von Goobi.Production erzeugt werden, direkt importieren. Allerdings unterscheiden sich METS/MODS-Daten je nach Erzeuger in Details. So führte gleich unser erster Versuch, ein Werk von Visual Library zu importieren, nach längerem Warten zur Fehlermeldung „zu wenig Speicher“. Ursache war eine Referenz in den METS/MODS-Daten, die auf sich selbst zeigte, was in der Goobi-Software zu einer Endlosrekursion führte. Die neueste Goobi-Software kann mit diesem Sonderfall umgehen, aber wir mussten uns behelfen, indem wir die Daten nachbearbeiteten und die kritische Referenz entfernten. Nach einigem Experimentieren konnten wir auch weitere Besonderheiten in den Daten anpassen beziehungsweise ergänzen und so erfolgreich alle Metadaten der Alten Drucke in Goobi.Presentation importieren.

### Die Portale für Digitalisate

Zum Zeitpunkt der Veröffentlichung des Beitrags präsentiert die UB Mannheim ihre Digitalisate in zwei Portalen. Dabei handelt es sich um zwei eigenständige Konfigurationen der Software Goobi.Presentation,

die die Metadaten im Format METS/MODS jeweils von der Goobi.Production Instanz der UB Mannheim einsammeln. Letztlich ist die Entscheidung für diese getrennte Präsentation aufgrund der inhaltlichen Unterschiede der Medien gefallen.

Eine inhaltliche Einheit stellen die historischen Drucke der UB Mannheim dar, die alle Werke von der Frühen Neuzeit bis zur Mitte des 20. Jahrhunderts umfassen. Hier finden sich überwiegend französische, lateinische und spanische Werke, die für die wissenschaftliche Arbeit der Geschichtswissenschaften und Philologien relevant sind.<sup>3</sup>

Die obige Abbildung zeigt die Eingangsseite der digitalisierten historischen Werke. Bei jeder Sammlung ist in Klammern hinter dem Eintrag die Anzahl der Titel eingetragen. Die tatsächliche Anzahl der Bände ist höher, da bei mehrbändigen Werken immer nur der Titel der Gesamtaufnahme gezählt wird. Aufgrund des großen Interesses von internationalen Wissenschaftlerinnen und Wissenschaftlern ist das Angebot durchgehend mit deutsch- und englischsprachigen Texten gestaltet.

Eine weitere Einheit sind die Bände des Aktienführers, mit denen hauptsächlich Wirtschaftswissenschaftlerinnen und Wirtschaftswissenschaftler arbeiten.<sup>4</sup> Der Aktienführer enthält Unternehmensprofile zu Unter-

<sup>3</sup> <http://goobi.bib.uni-mannheim.de/>

<sup>4</sup> <http://goobi.bib.uni-mannheim.de/aktienfuehrer/>

UNIVERSITÄT MANNHEIM

Universitätsbibliothek Mannheim

UB MANNHEIM

Aktienführer:

### Hoppenstedt Aktienführer (Archiv)

**Über den Aktienführer**

Für die auf Deutschland bezogene Unternehmens- und Finanzmarktforschung ist eine wichtige Forschungsdatenquelle der "Aktienführer", welcher von 1951-1999 jährlich in gedruckter Form erschien - zunächst im Verlag Saling, später im Verlag Hoppenstedt. Der Aktienführer enthält in standardisierter Form Berichte über deutsche und ausländische Aktiengesellschaften, die an einer deutschen Börse gehandelt werden bzw. wurden. In den Berichten finden sich u.a. Informationen und Daten zum Firmensitz und Vorstand, zu Tätigkeitsbereichen und Beteiligungen, zur Aktionärs- bzw. Kapitalstruktur und zu wesentlichen Positionen der Bilanz sowie Gewinn- und Verlustrechnung. Die Daten des Aktienführers sind aus mehreren Gründen in der Finanzmarktforschung sehr wichtig. Zunächst werden im Aktienführer heterogene Daten aus den verschiedenen Geschäftsberichten der Unternehmen in homogener Darstellungsweise zusammengefasst. Im Gegensatz zu internationalen Datenanbietern werden beim Aktienführer hierfür die nach nationalen Rechnungslegungsstandards publizierten Daten verwendet, was sie sehr wertvoll für die auf Deutschland bezogene empirische Kapitalmarktforschung macht. Die hohe Qualität der Aktienführer-Daten ist ein weiterer zentraler Aspekt. Zudem erlaubt die lange Publikationshistorie des Aktienführers Untersuchungen über lange Zeiträume hinweg.

**Über das Projekt**

Mit Unterstützung durch die **Deutsche Forschungsgemeinschaft** wurden im Jahr 2013 zunächst 21 für die wirtschaftswissenschaftliche Forschung besonders relevanten Exemplare digitalisiert. Die Digitalisierung erfolgte durch die Universitätsbibliothek Mannheim.

**Aktienführer**  
(1 Titel / 21 Bände)

Suchen nach:

Suchen

Abbildung 2: Goobi-Portal für den Aktienführer

nehmen, die an deutschen Börsen notiert sind oder waren. Neben den eigentlichen Digitalisaten sind hierbei für die Forschenden vor allem die enthaltenen Daten interessant, welche in empirischen Studien ausgewertet werden. Daher ist eng mit dem Präsentationsportal für die Digitalisate eine Datenbank verknüpft, in welcher die strukturierten Daten aus dem Aktienführer hinterlegt und exportierbar sind. Historische Bestände aus der Frühen Neuzeit sind für die wissenschaftliche Arbeit in diesem Bereich eher weniger relevant, weshalb sich eine eigenständige Sicht als zweckmäßig erwies.

Die Einstiegseite für den Aktienführer zeigt alle Bände des Aktienführers von 1884 bis 1979. Hier ist zu beachten, dass allein die urheberrechtsfreien Werke, d. h. vor mehr als 70 Jahren publizierten, frei zugänglich sind. Für den Zeitpunkt der Veröffentlichung des Betrags liegt die Grenze im Jahr 1945. Alle nach diesem Stichjahr gedruckten Bände sind aufgrund vertragsrechtlicher Bestimmungen lediglich für die nicht-kommerzielle, wissenschaftliche Verwendung zur kostenfreien Nutzung zugänglich. Für diese aktuelleren Aktienführer-Bände musste daher eine Authentifizierung realisiert werden. Hierzu müssen sich interessierte Einrichtungen zunächst bei der Deutschen Zentralbibliothek für Wirtschaftswissenschaften (ZBW) registrieren. Nach Prüfung der Zugriffsberechtigung werden die IP-Adressen der Einrichtung dann für den

Zugriff freigeschaltet. Realisiert ist dies mit einer entsprechenden Konfiguration des Apache2 Webservers. Wir bekommen von der ZBW die IP-Adressbereiche der zugelassenen Einrichtungen als XML-Datei und erzeugen daraus eine Apache2-Konfigurationsdatei mit inzwischen mehr als 3000 Zeilen. Das ist per Hand natürlich nicht mehr möglich und wird deshalb von einem kleinen Programm erledigt.

### Digitalisierungs-Workflow in Goobi.Production

Der allgemeine, typische Digitalisierungs-Workflow in Goobi.Production sieht folgendermaßen aus, wobei sich Details natürlich in Abhängigkeit von lokalen Konfigurationen unterscheiden können:

#### 1. Verbundaufnahme für das geplante Digitalisat

Der erste Schritt im Prozess ist normalerweise die Erstellung einer Titelaufnahme für das Digitalisat im Verbundkatalog durch Mitarbeiter der Abteilung Medienbearbeitung.

#### 2. Digitalisierungsvorgang in Goobi anlegen und Laufzettel erzeugen

Für jeden zu scannenden Band legt das Digitalisierungsteam in der Abteilung Medienbearbeitung einen Vorgang in Goobi an und erfasst dessen Titel-Metadaten. Eine große Erleichterung bei diesem Schritt ist

die Schnittstelle zur SWB-Verbunddatenbank, welche es ermöglicht, über den Identifier der Digitalisat-Titelaufnahme die Titeldaten nach Goobi zu laden. Jeder Vorgang erhält einen eindeutigen Vorgangstitel in Goobi, mit dessen Hilfe er später wieder aufgerufen werden kann. Mit dem Erstellen des Vorgangs in Goobi wird auf dem Goobi-Server ein gleichnamiges Verzeichnis angelegt, welches mit diesem Vorgang verknüpft und für die Digitalisate vorgesehen ist. Abschließend wird für den Digitalisierungsvorgang ein Laufzettel erzeugt, welcher alle für den weiteren Scan-Vorgang relevanten Informationen enthält.

### 3. Digitalisierung

Der Scan-Operator übernimmt den Vorgang in Goobi. Production, scannt den betroffenen Titel und führt eine Sicht-Qualitätskontrolle durch. Die Digitalisate werden dabei im entsprechenden Samba-Verzeichnis abgelegt, welches in Schritt 2 automatisiert erstellt worden war. Die Benennung der Digitalisate muss nach einem vorab konfigurierten Muster erfolgen, welches es Goobi ermöglicht, die Bilddateien korrekt zu sortieren. An der UB Mannheim lautet das Schema für die Benennung: Verbund-ID\_LaufendeBildnummer (z. B. 405844417\_0001, 405844417\_0002, ...).

### 4. Qualitätskontrolle

Nach Abschluss der Digitalisierung erfolgt eine weitere Qualitätskontrolle durch das Digitalisierungsteam der Abteilung Medienbearbeitung in einem Standard-Bildbetrachtungsprogramm. Spätestens hier sollten fehlerhafte, falsch benannte oder komplett fehlende Scans identifiziert werden, da eine nachträgliche Korrektur nur noch mit erheblichem Aufwand möglich ist.

### 5. Meta- und Strukturdatenvergabe

Ebenfalls in der Abteilung Medienbearbeitung angesiedelt ist die Vergabe der Meta- und Strukturdaten für die Digitalisate. Zunächst wird in Goobi.Production bei diesem Schritt den Digitalisaten die Seitenzählung aus dem Print-Werk zugeordnet. Anschließend werden Gliederungselemente sowie weitere besondere Merkmale (z. B. Abbildungen) ausgezeichnet. Welche Auszeichnungselemente hierbei in Goobi.Production zur Verfügung stehen sollen, kann bei der Konfiguration vorab festgelegt werden.

Da die Seitenzählung und die Strukturdaten denen des Original-Titels entsprechen sollen, ist während deren Auszeichnung ein ständiger Abgleich mit dem Digitalisat nötig. In Goobi.Production gibt es zwar



**Kein Warten auf Godot**

Eröffnen Sie Ihren Kunden höchste Flexibilität beim Lesen – mit Open Library.

Die Open-Library-Lösungen von EasyCheck machen maßgeschneiderte Öffnungszeiten auch in Ihrer Bibliothek möglich.

**easyCheck**  
library technologies

Ein Unternehmen der **ekz** Gruppe [www.easycheck.org](http://www.easycheck.org)

eine integrierte Seitenvorschau, diese ist jedoch zu klein, als dass ihr Seitenzahlen, Überschriften oder anderer Text entnommen werden könnten. Für die Unterstützung der Seitenzahl- und Strukturdatenvergabe ist diese Vorschau damit ungeeignet. An der UB Mannheim kommt stattdessen ein externes Bildbetrachtungsprogramm zur Unterstützung dieses Arbeitsschritts zum Einsatz, was natürlich zu Brüchen bei den Abläufen in Goobi.Production führt. Hier konnte aus Anwendersicht noch ein größeres Verbesserungspotential identifiziert werden.

## 6. Export

Als letzter Schritt erfolgt der Export des bearbeiteten Digitalisats. Nach dessen Abschluss steht das Digitalisat über Goobi.Presentation zur Verfügung.

### Einbindung in nationale und internationale Nachweissysteme

Bei der Datenmigration von *Visual Library* nach Goobi musste die persistente Adressierbarkeit auch mit einer neuen Verzeichnisstruktur und neuem DNS-Eintrag gewährleistet bleiben. Die UB Mannheim setzt den URN-Dienst der Deutschen Nationalbibliothek ein. Da für eine Übergangszeit der Visual-Library-Server noch unter der bisherigen URL<sup>5</sup> erreichbar ist, sind die auf diesen Server verweisenden URNs noch valide.

Die in Goobi neu vergebenen URNs müssen noch automatisiert erfasst und in der URN-Datenbank der Deutschen Nationalbibliothek eingetragen werden. Bei den ersten Diskussionen mit der Deutschen Nationalbibliothek kam es zu der Schwierigkeit, dass das von Goobi gelieferte Datenformat nicht von der Deutschen Nationalbibliothek automatisiert verarbeitet werden konnte. Während Goobi qualifiziertes xepicur liefert, konnte die Deutsche Nationalbibliothek lediglich unqualifiziertes xepicur verarbeiten. Qualifiziertes xepicur unterscheidet sich von unqualifiziertem darin, dass jedem XML-Tag das Präfix *epicur* vorangestellt ist und dass die Tags eine zusätzliche Hierarchieebene aufweisen. Bedauerlicherweise konnte die Deutsche Nationalbibliothek keinen verbindlichen Termin für die Anpassung ihres Importfilters anbieten, so dass auch qualifiziertes XML verarbeitet werden kann. Folglich musste die UB Mannheim das Format ihres Datenexports umstellen und von dem in Goobi vorgegebenen Standard abweichen. Mit Hilfe einer Lösung der Firma Zeutschel konnte der Sourcecode für die eigene Anwendung entsprechend geändert werden, so dass ein Datensammeln durch die

Deutsche Nationalbibliothek jetzt möglich ist, und der URN-Dienst funktioniert.

Des Weiteren werden die Metadaten von Goobi.Presentation durch das Zentrale Verzeichnis der Deutschen Drucke (zvdd) gesammelt und auf der angebotenen Übersichtsseite nachgewiesen. Über diesen Dienst werden die Metadaten und die digitalen Inhalte auch an die Deutsche Digitale Bibliothek und die Europeana gemeldet.

### OCR für Digitalisate

Die meisten unserer Alten Drucke sind mit Antiqua-Schriften gesetzt und somit mehr oder weniger gut maschinenlesbar. Wir möchten aus den Digitalisaten den Text erfassen, in PDF-Dateien zusammen mit den Digitalisaten bereitstellen und so für Suchmaschinen und automatisierte Auswertungen zugänglich machen. Eine gewisse Fehlerquote bei der Texterfassung (sogenanntes schmutziges OCR) ist dabei akzeptabel. Erste Versuche mit freier OCR-Software wie Tesseract und OCRopus beziehungsweise dem Nachfolger OCRopy waren vielversprechend, konnten aber noch nicht die Erkennungsrate der kommerziellen Produkte der Firma ABBYY erreichen. Deshalb wird die *ABBYY FineReader Engine CLI für Linux* mit einer Lizenz für das Fraktur-Paket zum Einsatz kommen. Diese Lizenzform beinhaltet auch Wörterbücher für OldSpanish, OldFrench und OldEnglish, die für die mehrheitlich in diesen Sprachen gedruckten historischen Werke der UB Mannheim wichtig sind.

### Erste Softwareaktualisierung

Gestartet sind wir mit Goobi.Production 1.10.3 Community Edition. Ende März 2015 erschien die neue Version Goobi.Production 1.11.0 CE mit vielen Verbesserungen und auch neuen Funktionen wie beispielsweise dem Zeitungsmodul und der Normdatenunterstützung.<sup>6</sup> Diese neue Version sollte in der etwas ruhigeren Woche nach Ostern installiert werden. Das hat auch gerade so funktioniert, denn mit einem guten Tag Zeitbedarf war der Aufwand doch größer als erwartet. Wie schon bei der Erstinbetriebnahme hatten uns auch hier Feinheiten der Konfiguration lange aufgehalten. Beispielsweise hatte sich die Erweiterung für PICA-CBS-Systeme geändert, aber die neue Erweiterung wurde nicht gefunden, und so funktionierte der Zugriff auf die Metadaten des SWB nicht mehr.

### Resümee

Bei der Einführung des Produkts zeigten sich viele für

<sup>5</sup> <http://digi.bib.uni-mannheim.de/>

<sup>6</sup> <http://www.goobi.org/new/single/release-goobiproduction-1110-ce-fuer-die-zeitungsdigitalisierung-und-mehr/>



Open-Source-Software typische Phänomene. Auf der positiven Seite ist die Möglichkeit der umfangreichen Mitwirkung bei der Ausgestaltung von neuen Funktionen hervorzuheben, die allein mit der Community abzusprechen sind. Auf der negativen Seite steht die veraltete oder teilweise gar nicht vorhandene Dokumentation. Dieses Defizit wurde teilweise durch die Unterstützung des externen Dienstleisters Zeuschel kompensiert. Während des gesamten Projektzeitraums stand die Firma mit ihrem Wissen als Unterstützung bereit und ermöglichte eine zügige Durchführung des Projekts. Dennoch konnten aufgrund der fehlenden Dokumentation viele Funktionen erst im Trial-und-Error-Verfahren implementiert werden. Daher musste die UB Mannheim mehr Personalkapazität als erwartet einsetzen. Dennoch hielt sich der Gesamtaufwand im Rahmen und hätte bei einer vollständigen Dokumentation sicherlich erheblich reduziert werden können. So bleibt zu hoffen, dass sich in der Gemeinschaft der Goobi-Anwender noch mehr Teilnehmer finden, die bereit sind, hier mitzuhelfen und ihren Erfahrungsschatz für andere aufzuschreiben. Es gibt bereits ein Wiki für alle Themen rund um Goobi, aber im letzten Jahr gab es darin kaum Korrekturen oder gar neue Beiträge. Vielleicht ist ja unser Artikel ein Anstoß, das zu verbessern.

Aufs Ganze gesehen war die Einführung von Goobi und die Übernahme der Daten aus Visual Library erfolgreich: Zum heutigen Zeitpunkt steht eine funktionstüchtige Software zur Unterstützung des Digitalisierungsworkflows und zur Präsentation der Digitalisate zur Verfügung, die in die nationalen und internationalen Nachweissysteme eingebunden ist. **I**



#### Dr. Christian Hänger

Abteilungsleiter Digitale Bibliotheksdienste  
Universitätsbibliothek Mannheim  
Schloss Schneckenhof West  
68131 Mannheim  
[christian.haenger@bib.uni-mannheim.de](mailto:christian.haenger@bib.uni-mannheim.de)



#### Dr. Irene Schumm

Stellv. Abteilungsleiter Medienbearbeitung  
Universitätsbibliothek Mannheim  
Schloss Schneckenhof West  
68131 Mannheim  
[irene.schumm@bib.uni-mannheim.de](mailto:irene.schumm@bib.uni-mannheim.de)



#### Stefan Weil

Stellv. Abteilungsleiter  
Digitale Bibliotheksdienste  
Universitätsbibliothek Mannheim  
Schloss Schneckenhof West  
68131 Mannheim  
[stefan.weil@bib.uni-mannheim.de](mailto:stefan.weil@bib.uni-mannheim.de)



OBID i-scan® HF



Neuer Handheld Reader

## Inventur mit Power.

- Leistungsstarker „Boost-Mode“ bis zu 4 W
- Flüssiges Arbeiten durch großen Datenpuffer
- Lange Betriebszeiten bis zu 16 Stunden
- Integrierte Antenne und WLAN-Modul
- Automatische Mediensuche
- Automatische Überprüfung / Änderung des AFI-Bytes



ID ISC.PRH200

OBID® – RFID by FEIG ELECTRONIC

**FEIG**  
ELECTRONIC

FEIG ELECTRONIC GmbH  
Lange Straße 4 · D-35781 Weilburg  
Tel.: +49 6471 3109-0  
Fax: +49 6471 3109-99 · [www.feig.de](http://www.feig.de)