

# ■ OBERÖSTERREICHISCHE LANDESBIBLIOTHEK: NEUES LEBEN FÜR ALTE BÜCHER – ODER WIE MAN EIN DIGITALISIERUNGSPROJEKT AUF SCHIENE BRINGT!

von Gregor Neuböck

## **Inhalt**

1. Erster Projektabschnitt
2. Ausformulierung der konkreten Anforderungen an ein künftiges System
3. Auswahl und Prüfung der am Markt befindlichen, für uns relevanten Systeme
4. Beschreibung des Aufbaus und der einzelnen Prozesse

**Zusammenfassung:** Die Oberösterreichische Landesbibliothek hat seit 2009 ihre Digitalisierungsambitionen sukzessive verstärkt und ab 2010 mit der konkreten Umsetzung eines Digitalisierungsprojektes begonnen. Mittlerweile digitalisieren viele Bibliotheken in Österreich urheberrechtsfreie Werke, bzw. „lassen digitalisieren“.

Die Oö. Landesbibliothek hat sich für einen zweiten Weg entschieden. In einem längeren Prozess wurde Eigenkompetenz im Bereich der Digitalisierung aufgebaut. Im Besitz der Oö. Landesbibliothek befinden sich ca. 50.000 alte Drucke (vor 1850), 1.000 Handschriften und eine Fülle an landeskundlich relevantem Material. Viele dieser Bücher werden aus konservatorischen Gründen nicht mehr ausgehoben, trotzdem wollen wir den kulturellen Schatz Oberösterreichs einer möglichst breiten Bevölkerungsschicht zugänglich machen.

Die Digitalisierung von landeskundlichen Monographien und Periodika ist für eine Landesbibliothek eine große Herausforderung für ihr »Produktportfolio«. Immer mehr Leser und Forscher (m/w) bevorzugen die gezielte Suche nach digitalisiertem »Content« gegenüber der Arbeit mit den physischen Exemplaren. Landesbibliotheken haben die Chance, ihre »identitätsstiftenden«, analogen Sammlungen auch im Internet als eBooks anzubieten. Die Oö. Landesbibliothek hat dazu den Digitalisierungs-Workflow der „Digitalen Landesbibliothek Oberösterreich“ auf die Beine gestellt. <http://digi.landesbibliothek.at>

**Schlagwörter:** GOOBI, Repository, Digitalisierung, Digitalisierungsworkflow, Oberösterreichische Landesbibliothek, „Digitale Landesbibliothek Oberösterreich“, elektronische Pflichtexemplare

## **UPPER AUSTRIA FEDERAL STATE LIBRARY: „NEW LIFE FOR OLD BOOKS – HOW TO START A DIGITIZATION PROJECT!“**

**Abstract:** *The Upper Austria Federal State Library in Linz started planning their digitalisation project in 2009. Since then they have gradually implemented their digitalisation process and in 2010 they began to physically start their digitization work. Many libraries in Austria now digitalise copyright-free works. However The Upper Austria National Library has opted for a slightly different method of this digitalisation process to many other libraries. They use a very specific digitalisation system called Goobi. The Library has about 50.000 old prints dating from before 1850, over 1.000 manuscripts of various ages and a wealth of expert sensitive material much of which has never been documented or viewed by the public before. They also have access to many Private collections of books and documents as well. Many of these books are now out of circulation due to the need to conserve and protect them. However, at the Library they want to make these cultural treasures of Upper Austria accessible to the broad population not just in Austria but throughout the world, they are making this possible through their current digitalisation work.*

*The digitalisation of cultural books and periodicals is a great challenge for The Upper Austria National Library. More and more readers and researchers prefer to carry out specific searches for digitalised documents and books rather than work with the physical copies. They also can do this from anywhere in the world as opposed to have to go to where the book or document is. Many regional libraries have the chance to offer analogue collections on the Internet as eBooks, this is something in which The Upper Austria National Library have become market leaders in. They are working hard to increase the digitalisation work flow and have through this put the Library's digitalised content on the internet therefore making it assessable to all. This content can be found at: <http://digi.landesbibliothek.at>*

**Keywords:** *GOOBI, Repository, digitizing, digitizing workflow, „Upper Austria National Library, Digitale Landesbibliothek Oberösterreich“, born digital*

Neben der Katalogisierung des aktuellen und historischen Bestandes im Rahmen der (Retro-)Erfassung von Metadaten ist die Digitalisierung von landeskundlicher Literatur, Rara und Handschriften gerade für eine Landesbibliothek eine Herausforderung für ihr „Produktportfolio“. Immer mehr LeserInnen und ForscherInnen bevorzugen die gezielte Suche nach digitalisiertem Content gegenüber der Arbeit mit den physischen Exemplaren. Landesbibliotheken haben damit die Chance, ihre identitätsstiftenden, analogen Sammlungen auch im Internet als hauseigene Qualitäts-E-Books anzubieten.

Die Oö. Landesbibliothek hat deshalb einen Weg gesucht, ausgewählte Bestände an urheberrechtsfreien Büchern selbst zu digitalisieren und ins

Internet zu stellen. Im Gegensatz zu vielen anderen Bibliotheken, die Digitalisierungsprojekte auslagern, war für die Oö. Landesbibliothek die Kontrolle des Arbeitsprozesses von der Herstellung bis zur Veröffentlichung ein zentrales Anliegen. Ein qualitativ hochwertiges Ergebnis auf Basis der Digitalisierung und eine besondere Erschließungstiefe mittels Strukturdaten gehörten zu den vorab festgelegten Grundanforderungen in diesem Projekt.

Zunächst wurde das Wissen über Standards, Komponenten und die wichtigsten Projektteile bei den beteiligten KollegInnen sukzessive aufgebaut. Im Anschluss daran wurden verschiedene Softwarehersteller und Scanneanbieter im Bereich Digitalisierung im Hinblick auf unser Anforderungsprofil und den zu berücksichtigenden Standards überprüft.

Digitalisierung meint nicht nur das bloße Scannen von Dokumenten, die dann auf Datenträgern abgelegt werden, sondern einen integrierten Workflow, der von der Übernahme der bibliographischen Daten bis zur Präsentation und Archivierung reicht.

So gibt es bei Handschriften andere Aufgaben als bei landeskundlichen Büchern des vorwiegend 17. bis 20. Jahrhunderts.

Die Zielsetzung bestand darin, die betreffenden Dokumente und hier insbesondere die Handschriften, einer breiteren Öffentlichkeit zugänglich zu machen und das Profil der Bibliothek zu schärfen. Damit verbunden war das Ansprechen möglicher neuer Benutzerschichten, die Förderung landeskundlicher Forschung sowie der Forschung an Handschriften bei gleichzeitiger Schonung der Originale.

Viele Bibliotheken meiden die interne Digitalisierung und tendieren dazu diese auszulagern, da es häufig an entsprechendem Fachpersonal mangelt und die Bibliothekare vor Ort sich in Bezug auf Anforderungen an ein Digitalisierungsprojekt überfordert fühlen. Unsere Bibliothek hat gezeigt, dass mit entsprechender Vorbereitungszeit und Weiterbildung ein solches Projekt auch ohne externe Spezialisten durchführbar ist.

Der eingeschlagene Weg dorthin und die einzelnen Stationen auf diesem Weg sollen im Folgenden kurz dargestellt werden.

Die Projektleitung und damit die Vorbereitung und Durchführung des Projekts von der technischen Seite erfolgte mehr oder minder durch mich.

Bibliothekarische Fragen und Anforderungen wurden durch Rudolf Lindpointner bereitgestellt und geprüft.

## 1. Erster Projektabschnitt (Oktober 2009 bis Februar 2010)

In diesem Zeitraum wurden grundlegende Fragestellungen bearbeitet, wie:

- Auseinandersetzung mit Standards
- Fragen zu Datenformaten
- Auflösung der Digitalisate
- Validität der von den Digitalisierungssystemen generierten METS-Dateien nach der Library of Congress (Metadata Encoding and Transmission Standards), auf deren Basis eine langfristige Archivierung dieser XML-Dateien, die in Form des MODS (Metadata-Objects-Description-Schema) eingebettet werden etc.
- Ebenso wurde der große Bereich OCR (Optical Character Recognition) näher beleuchtet. Es sollte die Möglichkeit bestehen Fremdserver oder lokale Server einzusetzen. Da der OCR-Prozess extrem rechnerintensiv ist, haben wir uns zu einer Auslagerung entschlossen.
- Zitierfähige persistente URLs (PURLs, z. B. URNs = Uniform Resource Names, DOI = Digital Object Identifier, Handle) – dabei handelt es sich um einen Identifikator, der nicht auf eine URL verweist, sondern auf einen Linkresolver, der dann die Weiterleitung an die entsprechende URL vornimmt, somit den klassischen 404-Fehler toter Links vermeidet, da der Link-Resolver ständig überprüft, ob eine URN auch auf eine gültige URL verweist.

## 2. Ausformulierung der konkreten Anforderungen an ein künftiges System (März bis April 2010)

Das waren konkret:

- Abbildung des gesamten Workflows der Digitalisierung – Produktionssystem und Präsentationssystem mit allen für uns notwendigen Teilschritten.
- Flexibilität hinsichtlich möglicher Entwicklungen: Erweiterung um neue Strukturelemente, elektronische Pflichtexemplare verwalten, Einbindung externer Dienstleister, flexible Gestaltung des Viewers durch eigene digitale Kollektionen für externe Institutionen (z.B. „Handschriften Oö. Klöster und Stifte“), freie Gestaltung bzw. Abänderung des Workflows
- OAI-Schnittstelle mit gängigen Formaten
- Generierung persistenter URLs (URNs)
- OCR-Schnittstelle

- Verwendung unterschiedlicher Metadatenformate, MARC 21, MARC XML, MAB 2, MAB XML, Dublin Core, METS/MODS, Xepicur, PICA, EAD und Schnittstellen z. B. OAI-PMH, Z39.50, sollte möglich sein
- Verwendung unterschiedlicher Grafikformate (TIFF, JPEG, PNG, GIF, JPEG2000)
- räumliche und zeitlich voneinander unabhängige Zusammenarbeit muss möglich sein, z.B. die Einbindung externer Scandienstleister bzw. anderer externer Beiträge
- flexible Benutzerverwaltung in den Bereichen Scanoperator, Metadatenbearbeiter, Qualitätskontrolle, Administrator,... inklusive Zuordnung zu Teilschritten im Workflow
- Betriebssystemunabhängigkeit: Sowohl das Produktionssystem als auch das Präsentationssystem sollten im Browser funktionieren

### 3. Auswahl und Prüfung (Mai bis Oktober 2010) der am Markt befindlichen, für uns relevanten Systeme

In Betracht gezogen wurden dabei:

- CCS (<http://www.ccs-digital.info/de/produkte>)
- GOOBI (<http://wiki.goobi.org/index.php/Hauptseite>)
- OCLC (<http://www.oclc.org/contentdm/about/features/default.htm#completesolution>)
- Visual Library (<http://www.walternagel.de/bildungs-und-kultureinrichtungen/software/visual-library>)

Die **Entscheidung fiel im November 2010 auf GOOBI**, wobei folgende Gründe den Ausschlag gaben:

- Betriebssystemunabhängigkeit: läuft z.B. auf UNIX Systemen wie SUN Solaris, BSD Systemen, Mac OSX, verschiedenen Linux-Distributionen, Microsoft Windows. Alle anderen untersuchten Systeme benötigten jeweils einen Windows-Client – dazu muss ein Programm auf einem Windows-Computer installiert werden. Pro installiertem Client fallen wiederum erhebliche Lizenzkosten an. In GOOBI können völlig frei Benutzer mit abgestimmten Berechtigungen und Aufgaben angelegt werden, es wird nur ein Browser benötigt und ein Internetzugang.
- hohe Flexibilität: GOOBI ist in seinen Konfigurationsmöglichkeiten sehr vielseitig und lässt sich an unterschiedlichste Bedürfnisse anpassen.

- Open Source: GOOBI basiert auf den Open-Source-Komponenten Java Server Faces (JSF) Technologien, der Suchindex auf Apache Lucene oder Apache Solr und das Dateisystem auf dem SMB-Protokoll, dadurch wird die Dateisystem-Unabhängigkeit garantiert und jedes Betriebssystem kann auf die Dateien von Goobi zugreifen. Es unterstützt alle gängigen Java-Application-Server (z.B. Apache Tomcat, GlassFish). SQL-Standard konformen Datenbanken (z.B. MySQL) ebenso wie gängige Verzeichnisdienste auf dem x.500 Protokoll basierend (z.B. LDAP, Active Directory, eDirectory von Novell, OpenDirectory von MAC).
- nachhaltige Entwicklungsgarantie: Große Institutionen wie die Universität Göttingen, die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden oder die Staatsbibliothek Berlin ermöglichen eine nachhaltige Weiterentwicklung der Software. Mittlerweile arbeiten mehr als 40 Einrichtungen aus 5 verschiedenen Ländern mit GOOBI.

Im September 2012 kam es zur Gründung des eingetragenen Vereines „GOOBI. Digitalisieren im Verein“ mit folgenden Aufgabenschwerpunkten:

- Organisation des Release Managements für Goobi
- Schutz der Markenrechte an Goobi
- Durchführung oder Beauftragung von gemeinnützigen Forschungs- und Entwicklungsvorhaben
- Durchführung von eigenen Seminaren, Kursen und Ausstellungen sowie anderen Formen der Öffentlichkeitsarbeit
- Wahrnehmung von eigenen Anliegen gegenüber Bund, Ländern und der Europäischen Union sowie gegenüber Wissenschaftsorganisationen und Öffentlichkeit
- Förderung des fachlichen Nachwuchses, zum Beispiel durch internationale bzw. institutionelle Austauschprogramme

#### **4. Beschreibung des Aufbaus und der einzelnen Prozesse innerhalb der digitalen Landesbibliothek Oberösterreich**

Im März 2011 konnten wir schließlich nach einer gerade mal eintägigen Schulung in unserem Haus mit dem Testbetrieb von GOOBI beginnen, wobei in dieser Einarbeitungsphase unsere Digitalisate auf einem Hosting-Server gespeichert wurden. Diese Phase (bis Jahresende 2011) war auch durch einen ständigen intensiven Kontakt und Austausch mit der Betreuer-

firma gekennzeichnet. Wir haben innerhalb dieses Zeitraumes Anforderungen und Wünsche, die sich oft erst im Zuge der Arbeit mit dem System herauskristallisierten, laufend formuliert und weitergegeben.

So wurden bisher ca. 1.000 Monographien und Periodika sowie 50 Handschriften digitalisiert und bearbeitet. Insgesamt sind derzeit schon mehr als 1.600 Bände oder 75.000 Images ins System eingearbeitet worden.

#### *Unser Personaleinsatz:*

Ab März 2011 arbeiteten 2 Personen sowohl am Scanner, als auch in der Struktur- und Metadatenbearbeitung. Seit Jänner 2012 arbeiten zusätzlich 2 Kollegen einige Stunden pro Woche am Scanner. Im Sommer 2012 arbeitete für 4 Wochen ein Ferialpraktikant am Scanner und seit November 2012 arbeitet eine weitere Kollegin stundenweise im Bereich der Struktur- und Metadatenbearbeitung.

#### *Hardware Serversystem:*

Angesichts der großen Datenmengen (wir scannen unsere Mastertiffs mit 600 DPI) fiel schon im Juni 2011 die Entscheidung für ein eigenes Serversystem, wobei vor allem die Uploadzeiten, aber auch die Kosten (Hostinggebühren steigen im TB-Bereich rapide an, spezielle Sicherung → Langzeitarchivierung kostet nochmals entsprechend bzw. läßt sich bei vielen Hosting-Systemen gar nicht oder nur mit einem sehr hohen Kostenaufwand realisieren) für die Anschaffung eines eigenen Servers sprachen.

Die Ausschreibung erfolgte 08/2011. Das angekaufte System wurde 11/2011 mit folgenden Spezifikationen ausgeliefert:

- Virtualisierungs-Server: 2HE 19" Intel Barebone (max. 12 HDD), 2x Intel Xeon 4-Core CPU 2,4GHz, 48 GB RAM, 2x 147 GB SAS II HDD (Kapazität brutto 294 GB), 6x 2000 GB SAS II, HDD (Kapazität brutto 12000 GB, RAID ca. 8000 GB), 1x SAS II RAID Controller 4 intern/4 extern.
- Hardware Backup-Server: 2HE 19" Intel Barebone (max. 8 HDD), 1x Intel G620 2-Core CPU 2,6 GHz, 4GB RAM, 2x 250 GB SATA II HDD (Kapazität brutto 500 GB), 4x 2000 GB SATA II HDD (Kapazität brutto 8000 GB).
- Hardware Embedded Firewall: ALIX 2d3, 2 GB SanDisk CF Speicherkarte, Netzteil.

Ab Dezember 2011 erfolgte die Installation des Produktions- und Präsentationssystems von GOOBI auf den hauseigenen Servern. Gleichzeitig begann der Datentransfer vom Hostingserver auf unseren eigenen.

### *Hardware Buchscanner:*

Mitte November 2011 wurde unser neuer Produktionsscanner i2S Copibook HD 600 geliefert. Er zeichnet sich durch folgende Spezifikationen aus:

- Motorische Buchwippe 10cm/10kg mit einer manuell gefederten Glasplatte.
- Das Gerät verfügt über eine maximal optische Auflösung von 600 dpi über ein Fläche von A2+ und eine Farbtiefe von 24/36 Bits RGB. Es können die Bildformate TIFF, JPEG, BMP, JPEG 2000, PDF und Multipage (TIFF, PDF) ausgegeben werden.
- Die mitgelieferte Workflowsoftware mit automatischer Seitenfreistellung und Vorschau, sowie praktischen Shortcuts zum nachträglichen Einfügen, Löschen bzw. Überschreiben von Scans entsprach unseren Vorstellungen. Zusätzlich Features (z.B. gerade ausrichten) werden bewusst nicht genutzt.

Da wir mit sehr hochauflösenden Digitalisaten arbeiten, zeigt sich nun der große Vorteil eines lokalen Serversystems. Weil sich der Scanner direkt im LAN des Digitalisierungs-Servers befindet, werden die Scans direkt in GOOBI eingespielt und stehen daher sofort zur weiteren Bearbeitung zur Verfügung. Bei einem Hostingssystem müssen derartige Datenmengen via FTP in den Nachtstunden übertragen werden und würden erst zeitverzögert zur Verfügung stehen.

Für sehr große Institutionen ist aber auch dieser Weg mit GOOBI leicht zu realisieren, wie man am Beispiel der „Welcome Library“ (<http://welcome.library.org/>) sieht.

In dieser werden die meisten Arbeitsschritte automatisiert. Derzeit werden dort ca. 1.000 Bände pro Tag online gestellt und man möchte in absehbarer Zeit 2.000 Bände pro Tag schaffen.

Seit Februar 2012 befindet sich die „Digitale Landesbibliothek Oberösterreich“ im Vollbetrieb.

GOOBI bietet eine hohe Diversifikation, infolgedessen wurden einzelne MitarbeiterInnen auf bestimmte Arbeitsprozesse im Digitalisierungsworkflow eingeschult. Diese Spezialisierung verringert die Fehleranfälligkeit und verbessert die Qualität. Manchmal werden auch PraktikantInnen oder externe Dienstleister (z.B.: Digitalisierung von 20 Handschriften im Digitalisierungszentrum Graz SOSA) eingesetzt.

Dazu haben wir in GOOBI verschiedene User angelegt. Diese Benutzer werden exakt den Erfordernissen angepasst.

In GOOBI können grundsätzlich extern via FTP, VPN oder intern (LAN) via Netzlaufwerke (SAMBA) Images hochgeladen werden.

Derzeit haben wir 6 verschiedene „Produktionsvorlagen“ mit unterschiedlichen Workflowschritten in GOOBI angelegt.

Auf die einzelnen Workflowschritte können nur „berechtigte“ User zugreifen.

Die Arbeitsteilung sorgt auch für einen „sauberen“ Arbeitsplatz, weil nur über Benutzerrechte zugeteilte Aufgaben sichtbar werden. Gleichzeitig hat aber auch die/der Berechtigte die Möglichkeit, Aufgaben NICHT anzunehmen, oder kann bereits angenommene Aufgaben zum Abschließen wieder abgeben, z.B., wegen Unklarheiten, Problemen, Urlaube, ....

Über ein internes WIKI wird eine dauerhafte und nachhaltige Kommunikation, z.B. Arbeiten an einem Band, die später zu erledigen sind, Fehlermeldungen etc., innerhalb von GOOBI garantiert.

Nachfolgend werde ich einige verschiedene Produktionsvorlagen und die darin enthaltenen Workflowschritte an der Oö. Landesbibliothek vorstellen und erläutern.

#### **4.1. Standard-Workflow**

##### *1. Bibliographische Aufnahme*

Es kommt zur Datenübernahme aus der Aleph-Datenbank. Die vorhandenen Daten können nachträglich noch angepasst und angereichert werden.

Für Zeitschriften werden Vorlagen angelegt, da im OPAC nur ein Identifier (AC-Nummer) vorhanden ist. Zum Anlegen der einzelnen Bände werden diese Vorlagen genutzt und nur die Bandangaben werden dabei ergänzt.

##### *2. Scannen*

Die Bücher werden direkt von unserem Scanner in den GOOBI-Server eingespielt. Um große Mengen in kurzer Zeit bewerkstelligen zu können, kann der Cache am Scanner mit bis zu 1 TB aktiviert werden.

##### *3. Qualitätskontrolle*

Die Digitalisate werden einer Endkontrolle unterzogen. Für diese Aufgaben sowie eventuelle Bildbearbeitungen verwenden wir einen iMAC mit 12 GB RAM.

Mit der Annahme eines Prozesses in GOOBI wird der Bilderordner für den jeweiligen Benutzer gemountet (Netzlaufwerk wird verbunden). So

können die Bilder im Finder (so nennt man auf einem MAC den Explorer) einer Kontrolle unterzogen werden. Durch den großen Bildschirm (27 Zoll) und die gute Auflösung (fast 3,7 Mio. Pixel) ist eine außerordentlich rasche und gute Qualitätskontrolle möglich.

Wie schon weiter oben erwähnt, gibt es zwischen den einzelnen Workflowschritten ein Rückmeldesystem zur Kommunikation zwischen den handelnden Personen. Es kann zum Festhalten von Fehlern, für Anmerkungen oder Aufgaben verwendet werden. Der erfolgreiche/nicht erfolgreiche Abschluss automatischer Schritte, wie z.B. die Ergebnisse von OCR-Prozessen, werden ebenfalls in diesem WIKI dokumentiert.

#### *4. Automatischer Kopier- und Tiffheaderschritt*

Die Mastertiffs werden kopiert und einer JPEG-Komprimierung unterzogen. Da für die Bildanzeige im Viewer keine hohe Auflösungen notwendig ist, werden die Mastertiffs in einem späteren Schritt verschoben.

#### *5. OCR-Volltexterkennung*

Es wird zwischen Antiqua, Fraktur und anderen Schriften unterschieden. Dadurch wird eine Volltextsuche mit Highlighting ermöglicht.

#### *6. Struktur- und Metadatenbearbeitung*

Diese gliedert sich in die Schritte Paginierung, Strukturdaten- und Metadateneingabe.

Im Schritt Paginierung werden den einzelnen Images die Seiten des Werkes zugeordnet. Es kann zwischen arabischer und römischer Paginierung, unnummeriert und Freitext unterschieden werden. Zusätzlich kann noch zwischen Seiten-, Spalten-, Blatt-, Blatt-recto-verso- und Seiten-recto-verso-Zählung gewählt werden.

Im Punkt Strukturdaten wird meist das Inhaltsverzeichnis (sofern vorhanden) abgebildet. Es können aber zusätzlich auch Tabellen, Bilder, Anmerkungen, Zeitschriftenartikel, usw. erfasst werden. Grundsätzlich lässt sich GOOBI in diesem Bereich extrem gut anpassen und erweitern. Die verwendeten Strukturelemente stützen sich dabei auf die Standards der Library of Congress, sodass eine valide METS-Datei ausgegeben wird.

Im Bereich Metadaten werden abschließend die zuvor angelegten Strukturelemente mit Inhalten gefüllt.

#### *7. URN-Generierung*

In GOOBI werden unsere URNs laut Policy der OBVSG erzeugt → urn:nbn:at:AT-OOeLB: optionaler Unternehmensraum-9373

- urn: Uniform Resource Name
- nbn: National Bibliography Number
- at: Länderkennzeichen
- ISIL der Vergabestelle z.B.: AT-OOeLB für Oö. Landesbibliothek
- optionale Unternehmensräume der Vergabestelle z.B. für Institute an Universitätsbibliotheken.
- Identifikationsnummer der Vergabestelle inklusive Prüfziffer am Ende

### *8. Export in DMS*

Dabei werden alle Meta- und Strukturdaten, eventuell vorhandene OCR-Ergebnisse und die heruntergerechneten Bilder in den Viewer kopiert. Das heißt der Viewer ist völlig unabhängig von GOOBI. In unserem Fall sind GOOBI und der Viewer auf zwei unterschiedlichen virtuellen Maschinen installiert, so beeinflussen sich die beiden Systeme nicht gegenseitig. Grundsätzlich kann GOOBI mit unterschiedlichen Viewern zusammenarbeiten. Neben TYPO3 gibt es MyCoRe, DigiTool, den DFG-Viewer und den Intran-da-Viewer.

### *9. URN-Eintragung*

Derzeit werden die URNs noch manuell im Aleph-Katalog eingetragen. Sobald der URN-Resolver der OBVSG in Betrieb ist, erfolgt dieser Schritt automatisiert über die OAI-PMH Schnittstelle von GOOBI.

### *10. Archivierung*

Dabei werden alle Meta- und Strukturdaten und die OCR-Ergebnisse vom GOOBI-Server auf eine NAS (Network Attached Storage) kopiert. Die Mastertiffs werden vom GOOBI-Server zur Archivierung auf die NAS verschoben, da diese in GOOBI und im Viewer nicht weiter benötigt werden. So können bei uns ca. 3/5–4/5 des Speicherplatzes am Server eingespart werden.

Zusätzlich werden die Daten auf externen Festplatten nochmals gesichert, mit einem Datumsstempel versehen und im Tresorraum aufbewahrt. Mittelfristig ist aus Sicherheitsgründen angedacht diese Platten bei einem Schwesterinstitut im Land Oberösterreich unterzubringen.

## **4.2. Workflow Elektronische Pflichtexemplare**

Dieser Workflow wurde an die im Folgenden beschriebenen Erfordernisse angepasst:

Anstatt des Schrittes „Scannen“ erfolgt nun der *PDF Import*. Die vorhandene PDF-Datei wird in das gemountete Verzeichnis kopiert oder per FTP hochgeladen.

Die *Automatische PDF Konvertierung* erfolgt nach validem Abschluss des vorangegangenen Schrittes. Für jede enthaltene Seite wird ein TIFF-Bild angelegt und der zugehörige Text wird extrahiert. Diese so erzeugten Daten werden in die entsprechenden Bild- und OCR-Ordner in Goobi verschoben. Zusätzlich wird die Struktur des PDF-Dokuments analysiert und in eine METS-Datei überführt. Da innerhalb des PDF-Dokuments keine Information zum Typ der Struktur enthalten ist, wird jedes erkannte Strukturelement in das Element „Sonstiges“ überführt. Sofern vorhanden werden auch Titelinformationen für diese Strukturelemente übernommen.

Im Schritt *Automatischer Kopier- und Tiffheader* werden für die zuvor erzeugten Bilddaten Tiff/JPEG Derivate für den Viewer erzeugt sowie die Tiffheader gesetzt.

Der Bereich *Meta- und Strukturdatenbearbeitung* ermöglicht eine optionale Kontrolle und Nachbearbeitung der importierten und generierten Daten im schon dargestellten Metadateneditor von Goobi, dieser Schritt kann aber auch übersprungen werden.

Alle folgenden Schritte laufen derzeit automatisch ab, können aber jederzeit sowohl auf Exemplarebene, als auch global verändert werden.

Zuerst werden die *URNs erzeugt* und danach erfolgt der *automatische Export* aller relevanten Daten in den Viewer.

Die Archivierung läuft ident zur Produktionsvorlage Standard-Workflow.

Allgemein starten **Automatische Schritte**, sobald der vorhergehende Schritt valide abgeschlossen wurde. Manuelle Schritte können jeweils von berechtigten Benutzern angenommen bzw. abgeschlossen werden. Diese können auch wieder abgegeben werden (z.B. weil es Unklarheiten/Schwierigkeiten bei der Bearbeitung gibt, siehe auch Rückmeldesystem). Generell lässt sich sowohl die Anzahl der Workflowschritte als auch deren Ablauf völlig frei gestalten und an die individuellen Bedürfnisse der Bibliothek anpassen.

### **4.3. Anmerkungen zur Archivierung**

Das Thema der Langzeitarchivierung ist mit vielen Unwägbarkeiten versehen. Neben der Datensicherheit, garantiert durch mehrere voneinander unabhängige Sicherungen, geht es vor allem um die „Lesbarkeit“ von Daten in der Zukunft.

Aus diesem Grund hat sich die Oö. Landesbibliothek für eine Speicherung der Metadaten im METS-Format und bei den Bilddateien für das TIFF-Format entschieden. Beide Formate scheinen nach dem heutigen Wissensstand eine zumindest mittelfristige Sicherheit zu gewährleisten.

Die „Digitale Landesbibliothek Oberösterreich“ möchte in Zukunft auch verschiedenen anderen Institutionen (Bibliotheken, Gemeinden, Stifte ...) in Oberösterreich, die weder über die technischen – noch über die finanziellen Möglichkeiten zur Realisierung eines derartigen Systems verfügen, eine Plattform bieten, ihre urheberrechtsfreien Werke und kulturellen Schätze individuell zu präsentieren.

Bei der Auswahl der Werke für die Digitalisierung wird eher pragmatisch vorgegangen. Werke, die von Heimat- und Regionalforschern angefordert werden, werden ebenso digitalisiert wie Werke renommierter oberösterreichischer Landeskundler. Neben relevanter Literatur für aktuelle Landesausstellungen werden bevorzugt urheberrechtsfreie Werke zur Stadt- und Ortsgeschichte ins Netz gestellt.

Abschließend möchte ich noch einige Anmerkungen zu den Auswirkungen der digitalen Landesbibliothek auf die Organisation der Oö. Landesbibliothek anführen.

Alle Mitarbeiterinnen und Mitarbeiter erzählen mir geschlossen über die positiven Erfahrungen bei der Arbeit in GOOBI, ganz egal ob es sich um Scanoperatoren oder Metadatenbearbeiter handelt.

Wahrscheinlich hängt dies einerseits mit der Möglichkeit zusammen, Vorgänge dann bearbeiten zu können, wenn die notwendige Zeit vorhanden ist, andererseits können MitarbeiterInnen innerhalb kurzer Zeit die Früchte ihrer Arbeit im Internet unter [digi.landesbibliothek.at](http://digi.landesbibliothek.at) bewundern.

Die Leitung der digitalen Landesbibliothek fühlt sich ebenfalls durch die Mitarbeiterzufriedenheit und statistischen Daten in ihrem Weg bestätigt. Mittlerweile erfolgen pro Woche ca. 7.000–12.000 Seitenaufrufe aus beinahe 60 verschiedenen Ländern von Kanada bis Uruguay, Portugal bis Russland, Norwegen bis Südafrika und Saudi Arabien bis Tuvalu.

Dipl.-Päd. Gregor Neuböck MAS MSc  
Oberösterreichische Landesbibliothek  
Schillerplatz 2, 4021 Linz  
E-Mail: [gregor.neuboeck@ooe.gv.at](mailto:gregor.neuboeck@ooe.gv.at)