

Rudolf Lindpointner, Gregor Neuböck

Landeskundliche Bücher ins Netz bringen

Werkstattbericht der Oberösterreichischen Landesbibliothek

Neben der Katalogisierung des aktuellen und historischen Bestandes im Rahmen der (Retro-)Erfassung der Metadaten ist die Digitalisierung von landeskundlichen Monographien und Periodika gerade für eine Landesbibliothek eine 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 E-Books anzubieten. Die Oö. Landesbibliothek hat deshalb einen Weg gesucht, ausgewählte Bestände an urheberrechtsfreien Büchern selber zu digitalisieren und ins Internet zu stellen. Wichtig war neben einem einwandfreien Ergebnis auch das „learning by doing“ und die Kontrolle des Arbeitsprozesses von der Herstellung der „Images“ über die Erlangung einer besonderen „Erschließungstiefe“ mittels „Strukturdaten“ bis zur Veröffentlichung im Netz.

Die Digitalisierung von Handschriften, Rara und landeskundlicher Literatur war schon seit längerer Zeit ein Ziel der Oberösterreichischen Landesbibliothek. Der eingeschlagene Weg dorthin und die einzelnen Stationen auf diesem Weg sollen im Folgenden kurz dargestellt werden. 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.

Die Zielsetzung besteht natürlich hauptsächlich darin, die betreffenden Dokumente (und hier insbesondere auch die Handschriften) einer breiteren Öffentlichkeit zugänglich zu machen und damit verbunden

- das Ansprechen möglicher neuer Benutzerschichten,
- die Beförderung der landeskundlichen Forschung sowie auch der Forschung an den Handschriften durch die Vereinfachung des Zugangs,
- die Schonung der Originale und
- die Schärfung des Profils der Bibliothek.

Die Vorbereitung und Durchführung des Projekts von der technischen Seite erfolgte mehr oder minder im Alleingang durch den Projektleiter.

In einem ersten Schritt kam es zur Vorbereitung des Projekts (von Oktober 2009 bis Februar 2010), und diese bestand in der Auseinandersetzung mit den Standards im Digitalisierungsbereich, d.h. Fragen wie

- geeignete Datenformate
- 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

in Form des MODS (Metadata-Objects-Description- Schema) eingebettet werden.

- Zitierfähige persistente URLs(PURLs, URNs = Uniform Resource Names, DOI = Digital Objects 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.
- OAI-(Open Archives Initiative)-Schnittstelle ermöglicht das automatische Harvesten über OAI-PMH (OAI-Protocol Metadata Harvesting) und somit die freie Weitergabe von Metadaten (z.B. Europeana)

Der zweite Schritt (von März bis April 2010) bestand in der Ausformulierung der konkreten Anforderungen, die die Oö. Landesbibliothek an ein künftiges System stellt. Das waren konkret:

- Abbildung des gesamten Workflows der Digitalisierung (Produktionssystem und Präsentationssystem mit allen 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
- Generierung persistenter URLs
- OCR-Schnittstelle
- Verwendung unterschiedlicher Metadatenformate (MARC 21, MARC XML, MAB 2, MAB XML, Dublin Core, METS, PICA, EAD) und Schnittstellen (z.B. Z39.50) sollte möglich sein
- Verwendung unterschiedlicher Grafikformate (TIFF, JPEG, PNG, GIF, JPEG2000)
- Räumlich und zeitlich voneinander unabhängige Zusammenarbeit muss möglich sein (z.B.: die Einbindung externer Scandienstleister bzw. anderer externer Beiträger)
- Flexible Benutzerverwaltung (Rechteverwaltung: Scanoperator, Metadatenbearbeiter, Qualitätskontrolle, Administrator,...) inklusive Zuordnung zu Teilschritten im Workflow
- Betriebssystemunabhängigkeit: Sowohl das Produktionssystem als auch das Präsentationssystem sollten im Browser funktionieren

Der letzte Punkt war uns dabei sowohl unter Kostengesichtspunkten als auch unter dem Gesichtspunkt der Flexibilität bei der Arbeit besonders wichtig.

Der dritte Schritt (von Mai bis Oktober 2010) bestand in der Sichtung und der Prüfung einer Auswahl relevanter am Markt befindlicher Systeme im Hinblick auf die oben angeführten Anforderungen. 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 schließlich im November 2010 für GOOBI, wobei folgende Gründe den Ausschlag gaben:

- *Betriebssystemunabhängigkeit*: Alle anderen untersuchten Systeme benötigten jeweils einen Windows-Client (dazu muss ein Programm auf einem Windows-Computer installiert werden, und oftmals fallen dann pro installiertem Client erhebliche Lizenzkosten an). In GOOBI können völlig frei Benutzer 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 Open-Source-Komponenten. Große Institutionen wie die Universität Göttingen, die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden oder die Staatsbibliothek Berlin ermöglichen eine nachhaltige Weiterentwicklung der Software.

Im März 2011 konnten wir schließlich nach einer eintägigen Schulung in unserem Haus mit dem Testbetrieb beginnen, wobei in dieser Einarbeitungsphase unsere Digitalisate auf einem Hosting-Server gespeichert wurden. Diese Phase (bis Jahresende 2011) war außerdem dadurch gekennzeichnet, dass es einen ständigen intensiven Kontakt und Austausch mit den Technikern der Fa. Intranda gab und wir dabei Anforderungen und Wünsche, die sich oft erst im Zuge der Arbeit mit dem System herauskristallisierten, laufend formulierten und weitergaben.

So wurden bisher ca. 150 Monographien sowie mehrere Periodika und Handschriften digitalisiert und bearbeitet, insgesamt ca. 500 Bände. Angesichts der großen Datenmengen erfolgte schließlich auch die Entscheidung für ein eigenes Serversystem, wobei vor allem die Uploadzeiten, aber auch die Kosten (Hostinggebühren) für den Betrieb eines eigenen Servers sprechen.

Die Ausschreibung erfolgte 08/2011. Das System wurde 11/2011 geliefert. Es handelt sich dabei um einen Digitalisierungsserver mit 24GB RAM und 8TB Speicherplatz (netto) sowie einem Sicherungsserver und einer eigenen Firewall.

Seit Dezember 2011 erfolgt die Installation des Produktions- und Präsentationssystems von GOOBI auf unseren Servern. Gleichzeitig erfolgt der Datentransfer der schon vorhandenen Digitalisate vom Hostingserver auf unseren eigenen.

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

- motorische Buchwippe 10cm/10kg
- manuelle gefederte Glasplatte
- Auflösung optisch 600 dpi.

Derzeit sind wir also dabei, den Vollbetrieb der „Digitalen Landesbibliothek Oberösterreich“ (ab ca. Februar 2012) aufzunehmen.

Da GOOBI eine starke Diversifikation ermöglicht, erfolgt in den nächsten Wochen die Einschulung einzelner MitarbeiterInnen auf die zugewiesenen Arbeitsprozesse im Digitalisierungsworkflow.

Der derzeitige Workflow besteht aus folgenden Schritten:

- Bibliographische Aufnahme (Datenübernahme aus der Aleph-Datenbank möglich)
- Scannen
- Qualitätskontrolle
- Automatischer Kopier- und Tiffheaderschritt
- OCR-Volltexterkennung
- Struktur- und Metadatenbearbeitung (Kapitel, Abbildungen, Abschnitte, Tabellen etc.)
- URN-Generierung (automatisch)
- Export in DMS
- URN-Eintragung (dabei wird die URN des Digitalisates im Verbundkatalog eingetragen)
- Archivierung

Automatische Schritte starten, sobald der vorhergehende Schritt korrekt abgeschlossen wurde (z.B. wurde der Punkt „Qualitätskontrolle“ abgeschlossen, wird automatisch der Punkt „Automatischer Kopier- und Tiffheaderschritt“ angestoßen). Manuelle Schritte können jeweils von „berechtigten“ Benutzern angenommen werden. Auch besteht die Möglichkeit, einmal angenommene Vorgänge wieder abzugeben (z.B. weil es Unklarheiten/Schwierigkeiten bei der Bearbeitung gibt). Generell lässt sich sowohl die Anzahl der Workflowschritte als auch der Ablauf völlig frei gestalten und an die individuellen Bedürfnisse anpassen, und es können auch verschiedene Workflows für verschiedene Arten von Dokumenten eingerichtet werden.

Noch ein Wort 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.

Deswegen hat sich die Oö. Landesbibliothek für eine Speicherung der Metadaten im METS-Format und bei den Bilddateien für das TIFF-Format entschieden.

Bei der Archivierung in GOOBI ist zu beachten, dass das System vollautomatisch die Mastertiffs kopiert und durch JPEG-Komprimierung herunterrechnet. Die Mastertiffs können also aus GOOBI entfernt werden, ohne Auswirkungen auf Produktions- und Präsentationssystem (und auf geeigneten anderen Datenträgern, Festplatten, NAS ... gesichert werden). Dadurch verringert sich der Speicherplatz in GOOBI um bis zu 4/5 der Mastertiffs.

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 präsentieren zu können.

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. Weiters wird eine Verbindung zu den jeweiligen Landesausstellungen hergestellt, insbesondere werden urheberrechtsfreie Werke zur Stadt- und Ortsgeschichte ins Netz gestellt.