

Kooperative Arbeitsprozesse in Digitalen Bibliotheken am Beispiel des deutsch-russischen Projektes RusDML

Heike Neuroth / Ralf Stockmann¹

urn:nbn:de:0009-9-6560

¹ Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen. Bei der Erstellung dieses Berichts hat – im Rahmen ihres Referendariats an der SUB Göttingen – Antje Blomeyer mitgewirkt.

Einführung: neue Anforderungen an "digitale Bibliotheken"

<1>

Die wissenschaftlich ausgerichteten Bibliotheken befinden sich – wie der gesamte Wissenschaftssektor – in einer Phase der Umstrukturierung und Neuausrichtung. Der ungehinderte Zugang zu relevanten digitalen Informationen² von jedem Ort aus und zu jeder Zeit spielt in der digitalen Informationsgesellschaft eine immer bedeutendere Rolle. Völlig zu Recht erwarten die Nutzer der Bibliothek – Studierende wie Wissenschaftler – einen schnellen, verlässlichen und unkomplizierten Zugriff auf digitale Ressourcen. Angebote für Wissenschaftler wie von "JSTOR"³ oder "DigiZeitschriften"⁴ zeigen, dass der Anteil an digitalen wissenschaftlichen Artikeln immer weiter zunimmt und als Arbeitsinstrument mehr und mehr eingesetzt wird.

<2>

Im Zuge der Retrodigitalisierung werden hierzu in großem Umfang "analog" vorliegende Werke – Monographien wie Zeitschriften – digitalisiert und über das Internet verfügbar gemacht. Fernziel ist eine vollständige "World Digital Library (WDL)", in der ortsunabhängig alle jemals gedruckten Informationen abrufbar sind. Das DFG geförderte Projekt "Zentrales Verzeichnis Digitalisierter Drucke (ZVDD)"⁵ baut hierzu ein zentrales Nachweisportal zu bereits digitalisierten Bibliotheksmaterialien auf, um den heterogen verteilten, digitalen Content über eine einheitliche Oberfläche durchsuchbar zu machen.

<3>

Die Herangehensweise, die Literatur *eines* Fachgebietes weltweit zu digitalisieren und der Fach-Community bereitzustellen, ist in diesem Zusammenhang ebenfalls nicht neu. Die EU wird in den kommenden Jahren vielfältige Programme starten, um europäisches Kulturgut in digitaler Form über geeignete Plattformen sichtbar zu machen, wie sie es zum Beispiel mit dem Programm eContentPlus⁶ zur Zeit realisiert. (Ist eContentPlus ein Beispiel für fachbezogene Digitalisierung? Wenn nicht, sollte man sagen: "Die EU wird zudem . . .")

<4>

Es gibt gute Gründe, die Retrodigitalisierung der Altbestände direkt vor Ort, also in den Bibliotheken selbst, vorzunehmen. Zu nennen sind hier: Einsparen der Transportkosten, kontrolliert pfleglicher Umgang mit gefährdeten Beständen oder schnelle Wiederverfügbarkeit am Standort. Ausschlaggebend gegenüber einer auswärtigen Lösung ist jedoch vor allem der Vorteil, bibliothekarisch erfahrenes Personal in den Digitalisierungs- und Erschließungsprozess einbinden zu können.

<5>

Gerade der inhaltlichen Erschließung der Digitalisate durch Meta- und Strukturdaten kommt eine entscheidende Bedeutung zu: Suchen im Volltext oder direkte Navigation in den (digitalen) Inhaltsverzeichnissen im Webbrowser wird erst durch Erschließung möglich. Auch die Wissenschaftsinfrastruktur der nächsten Generation (etwa: Semantic Web) ist auf Metadaten angewiesen und kann mit reinen Scan-Images nichts anfangen.

² How much information? 2003 <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>>

³ JSTOR <<http://www.jstor.org/>>

⁴ DigiZeitschriften <<http://www.digizeitschriften.de/>>

⁵ ZVDD <<http://www.zvdd.de/>>

⁶ eContentplus <http://europa.eu.int/information_society/activities/econtentplus/index_en.htm>

<6>

Die Digitalisierung und Erschließung der eigenen Bestände stellt Bibliotheken und Archive jedoch auch vor Probleme, vor allem organisatorischer Art. Schon ab einem mittleren Scanvolumen von mehreren Bänden pro Tag erscheint eine Aufteilung der Arbeitsschritte (etwa: Arbeitsvorbereitung, Digitalisierung, Qualitätskontrolle, Imageoptimierung, Meta- und Strukturdatenerfassung) auf verschiedene, spezialisierte Mitarbeiter sinnvoll. In Drittmittelprojekten kommt häufig erschwerend eine räumliche Komponente hinzu: unterschiedliche Arbeitsschritte werden arbeitsteilig an unterschiedlichen Standorten vorgenommen.

<7>

Um diese arbeitsteilige Produktion orts- und zeitunabhängig kontrollieren und organisieren zu können, wurde am Göttinger Digitalisierungszentrum (GDZ) in Kooperation mit der Abteilung Forschung und Entwicklung der Staats- und Universitätsbibliothek Göttingen (SUB) eine Workflowsoftware für digitale Bibliotheken entwickelt. Im von der DFG geförderten Projekt "RusDML" (Russian Digital Mathematics Library)⁷ mit dem Ziel, in einer zweijährigen Laufzeit ein digitales Kernarchiv russischer Zeitschriftenartikel im Fach Mathematik in der Originalsprache aufzubauen, wurde so prototypisch die verteilte Produktion von insgesamt 283.000 Seiten koordiniert.

Kooperative Arbeitsprozesse: Workflow

<8>

Der komplexe Workflow von Beginn der Digitalisierung bis hin zur Online-Stellung einer Zeitschrift lässt sich am Beispiel des RusDML Projektes in folgende Einzelschritte gliedern:

I. Vorbereitende Tätigkeiten (verschiedene Partner)

1. Lizenz (TU Berlin, GPNTB Moskau)
2. Bestandsaufnahme der gedruckten Version (jeweiliger Standort, Koordination SUB Göttingen, GPNTB Moskau)
3. ZDB-Aufnahme (TIB/UB Hannover)
4. Bibliothekarische Aufnahme für den lokalen OPAC (SUB Göttingen, GPNTB Moskau)
5. Normierung der Zeitschriftentitel (TIB/UB Hannover)

II. Anlegen eines Prozesses und Digitalisierung (alle Partner, zur Zeit nur die SUB)

6. Anlegen eines Prozesses auf Bandebene (SUB Göttingen)
7. TIFF-Header erstellen (Partner und Beteiligte, die digitalisieren)
8. Scannen der Images (Partner und Beteiligte, die digitalisieren)

⁷ Die Homepage des Projektes RusDML finden Sie unter <<http://www.rusdml.de>> bzw. <<http://www.rusdml.org>> oder auch <<http://www.rusdml.net>>.

- Plattform-Unabhängigkeit (Web-Applikation), da Partner weltweit auf das Tool zugreifen müssen
- Zentrale Metadaten-Verwaltung, d.h. Katalogisieren und Vervollständigung der Metadaten von verschiedenen Standorten aus (zum Beispiel Erstellung der russischen Metadaten in Moskau, Transliterationen in Hannover)
- Zentrale Verwaltung der Digitalisate (Images)
- Import- und Exportschnittstellen für Metadaten und für Fremd-Digitalisate (zum Beispiel aus Russland)
- Controlling-Mechanismen: welcher Partner ist wie weit mit seinen Arbeiten, welche Zeitschrift ist in welchem Bearbeitungsstadium etc.
- Fehlermeldungen, Abschließen eines Arbeitsschrittes und Weiterleiten zur nächsten Arbeitsstufe (eventuell hier auch Wechsel/Übergabe von einem Partner zum anderen Partner)



Abbildung 2

<10>

Komponenten für das verteilte Workflowmanagement sind integriert worden, um die Verwaltung einer verteilten Produktion und Kommunikation zwischen den unterschiedlichen Partnern zu gewährleisten. Für jeden einzelnen Workflowschritt kann eingestellt werden, ob er parallel zum vorherigen durchgeführt werden kann oder erst nach dessen erfolgreichem Abschluss. Ferner werden Korrekturschleifen abgebildet, um in späteren Schritten festgestellte fehlerhafte Daten früherer Schritte zu vermerken. Die Applikation ist transparent konzipiert, somit ist es für jeden Projektpartner jederzeit möglich, den Status der Teilschritte des Workflows inklusive Personenzuordnung für einzelne Werke detailliert einzusehen.



Abbildung 3

<11>

Mit der Vergabe persistenter URLs kann auf die freigegebenen Digitalisate referenziert werden. Zuvor werden die Dateien mit den mehrsprachigen Metadaten auf Artikelebene verknüpft, die aus den verschiedenen Datenbanken zusammengestellt werden. Das Projekt widmet sich an dieser Stelle insbesondere der Übersetzung und Transliteration: Um einen benutzerfreundlichen Sucheinstieg zu gewährleisten, werden alle Titel in der Sprache der Vorlage und ergänzend wenigstens in englischer und/oder russischer Übersetzung angeboten. Alle Personennamen und Titel, die im Original in kyrillischer Schrift vorkommen, werden ins lateinische Alphabet transliteriert (gemäß ISO 9 und DIN 1460).⁸

<12>

Mit dem Import in das Document Management System sind die Materialien online verfügbar. Darüber hinaus werden die Artikel in verschiedenen Datenbanken nachgewiesen: Neben der Referenz in den Referateorganen JFM und Zentralblatt MATH sind dies vor allem die verschiedenen Bibliothekssysteme.

Resumee und Nachnutzung

<13>

RusDML stellt einen Teil der weltweiten Bestrebungen dar, eine globale mathematische Weltbibliothek, die sogenannte Digital Mathematics Library (DML), einzurichten.⁹ Die im Rahmen von RusDML konzipierten Werkzeuge erlauben einen verteilten, kooperativen Workflow zur Digitalisierung und Metadatenerschließung eines gemeinsamen Datenbestandes. Diese Werkzeuge sind von allgemeinem Interesse für ähnlich gelagerte Kooperationsprojekte jedweder Fachrichtung. Insofern hat RusDML eine Pilotfunktion für ähnliche Projekte oder Folgemaßnahmen. Vor allem im Hinblick auf die Kooperation verteilter Partner mit gleichen Aufgaben am selben Objekt, die Bilingualität der Metadaten und die Besonderheiten der in Kyrillisch abgefassten russischen Dokumente bieten die Werkzeuge besondere Leistungen.

<14>

Die Workflowsoftware wurde von Beginn an auf möglichst große Nachnutzbarkeit hin konzipiert. Folgerichtig wird sie nunmehr für die gesamte Produktion am GDZ eingesetzt. Interessierte Nachfragen anderer Bibliotheken haben nachdrücklich das generelle Interesse an einer solchen Produktionssoftware aufgezeigt. Folgerichtig wird die Software weiterentwickelt und mit anderen Modulen – etwa der Access-Steuerung oder der Nutzungsauswertung – kombiniert.

<15>

Hierdurch entsteht eine umfassende Lösung für die Umsetzung der (Massen-) Digitalisierung unter dem Namen "Goobi" (Göttingen online-objects binaries). Das folgende Schaubild verdeutlicht den modularen Aufbau:

<16>

Die Software wird lizenzfrei entwickelt und steht als open source zur Verfügung. Mittelfristiges Ziel ist es, eine verteilte Entwicklercommunity zu gründen, die die nachhaltige Pflege und Weiterentwicklung der Programme garantiert.

⁸ Ein von der TIB/UB Hannover entwickeltes Transliterationstool wurde zu diesem Zweck in das RusDML-Tool integriert.

⁹ Vgl. hierzu insbesondere Bernd Wegner: DML proposals and questions <<http://www.library.cornell.edu/dmlib/wegner.html>>.

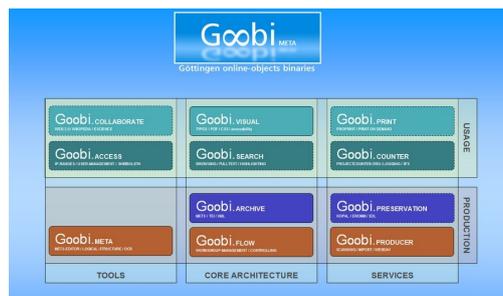


Abbildung 4

Weitere Informationen und eine Testumgebung sind unter <http://goobi.sub.uni-goettingen.de> erreichbar.

Abbildungen

Abb. 1

Arbeitsschritte im Workflow des Projekts mit Statusanzeige

Abb. 2

Erfassung und Bearbeitung der Metadaten

Abb. 3

Prozessverwaltung

Abb. 4

Funktions- und Modulübersicht der Goobi-Softwaresuite. Gestrichelte Module sind derzeit noch in der Konzeption/Entwicklung

Autoren:

Dr. Heike Neuroth

Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen

Abteilung Forschung und Entwicklung

Papendiek 14

37073 Göttingen

neuroth@sub.uni-goettingen.de

Ralf Stockmann

Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen

Göttinger Digitalisierungszentrum (GDZ)

Papendiek 14

37073 Göttingen

stockmann@sub.uni-goettingen.de