

Metadaten – Manuelle Datenpflege vs. Automatisieren

Ein Praxisbericht zu Metadatenmanagement an der ETH-Bibliothek

Erica Pfister, Barbara Wittwer und Marianne Wolff

.....

Neue Entwicklungen im Bibliothekswesen und in der Technologie führen zu neuen Aufgaben, die spezialisierte Fachkräfte erfordern. Die ETH-Bibliothek reagierte darauf, indem sie das Pilotprojekt „Metadatenmanagement“ startete, welches sich mit Datenanalysen, -einspielungen und -mutationen, Datenmappings, der Erstellung eines Datenflussdiagramms, sowie mit der Einführung von RDA und GND beschäftigte. Nach zwei Jahren zeigte sich, dass zahlreiche Aufgabengebiete existieren, welche von Metadaten spezialisten, als Schnittstelle zwischen den Fachabteilungen und der IT, übernommen werden können. Dieser Bericht fasst die getätigten Arbeiten, Erfahrungen und Erkenntnisse der zweijährigen Pilotphase zusammen.

New trends and technologies in the information sciences require new special skills from librarians. In light of these challenges the ETH library started a pilot project „metadata management“. This new team concentrated on tasks like data analysis and mutations, batchloading of metadata, mappings, the creation of a data flow diagram and the introduction of RDA and GND. It became apparent in the two years after the start of the project that a wide field of tasks exists for metadataspecialists. The team therefore forms a bridge between the Department of Acquisition and Cataloguing and the IT-Department. This report summarizes the executed tasks, experiences and conclusions of the pilot phase.

.....

Einleitung

Wie ein Großteil der wissenschaftlichen Bibliotheken steht auch die ETH-Bibliothek seit mehreren Jahren vor der Herausforderung, mit gleichbleibenden Personalressourcen neue Aufgaben wahrzunehmen, die Produkte und Services an die veränderten Rahmenbedingungen durch rasche technologische Entwicklungen anzupassen und in den von Träger- und Förderinstitutionen verlangten Bereichen weiterzuentwickeln. Um dies zu ermöglichen, werden Wege und Möglichkeiten gesucht, die anfallenden Arbeiten in der Bibliothek effizienter und im besten Fall automatisierter zu erledigen.

Die ETH-Bibliothek hat hierzu einen Schwerpunkt bei den Aufgaben der Erwerbungs- und Katalogisierungsabteilungen gesetzt. Die Aufgaben der Mitarbeitenden in diesen Abteilungen lassen sich zum einen durch Vereinfachung wie Fremddatenübernahmen, Metadateneinspielungen und Ähnlichem erleichtern, zum

anderen lassen sie sich aber auch automatisieren und dadurch effizienter gestalten. Die Prozesse der Automatisierung erfordern allerdings von den Mitarbeitenden – zusätzlich zu den spezifischen Kenntnissen der Katalogisierung – neue Qualifikationen und vertieftes Know-how in den Bereichen der Datenformate, Datenanalyse und Regelwerke.

Daneben werden heute oftmals von IT-Fachkräften bibliothekarische Aufgaben übernommen. Die Bibliothekssysteme der nächsten Generation tragen diesem Umstand bereits Rechnung und haben entsprechende Workflows in die Datenverarbeitung und -erfassung integriert, welche es den Bibliothekarinnen und Bibliothekaren erlauben, Eingriff in den technischen Strukturen vornehmen zu können.

Unter Berücksichtigung all dieser Faktoren wurde an der ETH-Bibliothek Anfang 2014 das Pilotprojekt „Metadatenmanagement“ gestartet. Drei Mitarbeiterinnen, zwei aus der Medienbearbeitung, eine aus den Bibliotheks-IT-Services, bildeten ein Pilotteam. Mit diesem Pilot wurde dem Team Raum, Zeit und Infrastruktur zur Verfügung gestellt, um mögliche Aufgaben zu definieren, Abläufe zu automatisieren, neue Lösungen zu entwickeln und künftige Tätigkeiten auszuloten. Hierbei wurde ein Vorgehen „on the job“ gewählt: die drei Mitarbeiterinnen transferierten in regelmäßigen Meetings Know-how und vertieften die verschiedensten Aspekte von Metadatenmanagement. Zusätzlich wurden das bis 2015 gültige und das seit Januar 2016 anzuwendende Regelwerk¹ studiert, und als Ergänzung zum Wissensaufbau Fortbildungen zu relevanten Themen besucht.

Rückblick

Im ersten halben Jahr 2014 lag der Fokus vor allem auf der Analyse und Bereinigung von Metadaten sowie dem Einspielen von Titelaufnahmen oder einzelnen Feldern, wobei sich diese Arbeiten auf das Bibliothekssystem Aleph und Daten der ETH-Bibliothek

¹ KIDS und RDA

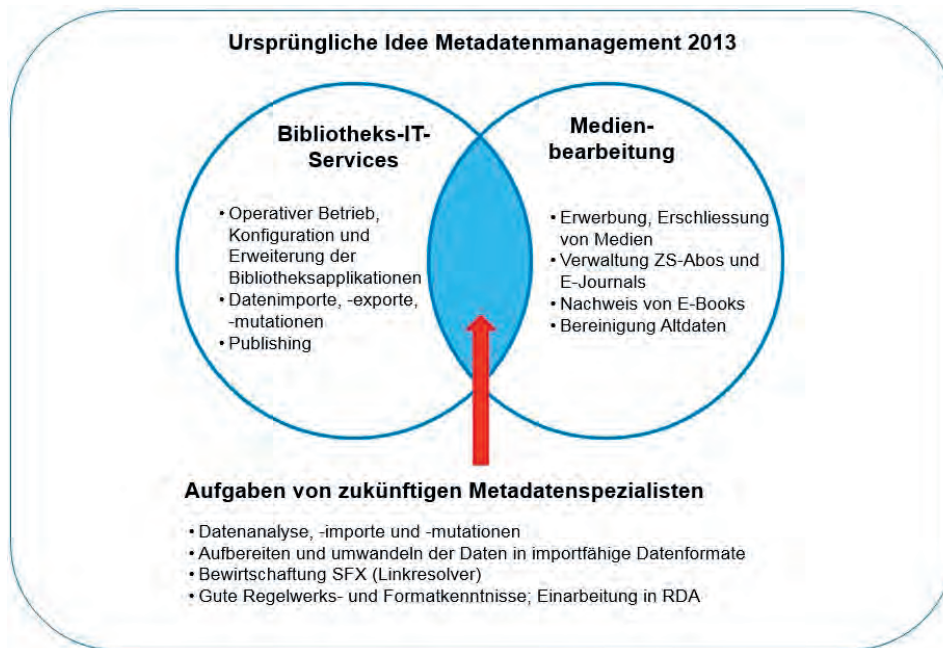


Abbildung 1: Ursprüngliche Idee Metadatenmanagement 2013

beschränkten. Darüber hinaus wurden verschiedene Möglichkeiten geprüft, um Arbeiten der traditionellen Katalogisierung zu automatisieren (z. B. Abgleichen von Daten mit dem Bibliothekskatalog, Erstellen von Angleichungsroutinen für Katalogdaten, Einspielen von Bestellaufnahmen etc.).

Das Bibliothekssystem bietet viele Optionen für Datenrecherchen und kann für Selektionen, Analysen und Datenabgleiche innerhalb der Datenbank genutzt werden. Datenmutationen lassen sich allerdings nur mit Zugriff auf die Server bewerkstelligen. Um diese Arbeiten durchführen und die Möglichkeiten, die Linux bietet (u. a. komplexere und schnellere Datenanalysen) nutzen zu können, wurde das Pilotteam entsprechend in Linux geschult.

In der zweiten Hälfte 2014 wurde der Fokus auch auf Metadaten im Discovery System und auf Datenflüsse zwischen den verschiedenen Systemen ausgeweitet. Die Metadaten der ETH-Bibliothek werden inzwischen in mehreren Quellsystemen verwaltet, welche von verschiedenen Teams betreut werden. Eine Übersicht oder eine Darstellung davon gab es bislang noch nicht, woraus sich die Aufgabe ergab, ein Metadatenschema zu erstellen. Anhand eines Datenflussdiagramms wurden die Wege der Metadaten der ETH-Bibliothek aufgezeigt. Als Weiterführung ist auch ein sogenanntes Data Dictionary geplant, das ergänzend dazu die in den Systemen enthaltenen Metadaten detailliert beschreibt.

Anhand des Diagramms lassen sich die Abhängigkeiten der Systeme erkennen und Änderungen an Metadaten können auf ihre Auswirkungen in anderen Systemen eingeschätzt werden. So kann z. B. das Wegfallen oder Hinzukommen eines MARC-Feldes

im Bibliothekssystem Anpassungen in den Normalisierungsregeln für das Discovery System und im Statistiktool nach sich ziehen. Zudem ist das Diagramm durch die Sichtbarkeit der Abhängigkeiten eine Hilfe bei der Evaluation und Einführung von neuen Systemen. Und nicht zuletzt konnte sich das Pilotteam beim Erstellen des Diagramms vertiefte Kenntnisse in verschiedenen Metadatenformaten (Dublin Core, METS etc.) aneignen.

Einführung RDA und GND²

2015 tat sich ein zusätzliches Arbeitsfeld auf, als die Detailplanung zur Einführung von RDA und GND an der ETH-Bibliothek und im NEBIS-Verbund startete³. Die Datenanalyse und -bereinigung sowie Systemanpassungen erwiesen sich im Laufe des Jahres als geeigneter Aufgabenbereich für das Pilotteam.

Erste Analysen von Katalogdaten ergaben, dass auf technischer Seite sehr viel Arbeit anfallen und die Konsequenzen der Umstellung weitreichend sein würden.

Durch den Wegfall von einigen lokalen Feldern und den Umstieg von IDS-MARC⁴ auf FULL MARC mit RDA mussten neue und ablösende Felder definiert, Suchabfragen optimiert und zahlreiche Indices angepasst werden. Hier war das Datenflussdiagramm wiederum eine Hilfe, da abhängige Systeme erkannt werden konnten. Diese Systemkonfigurationen muss-

2 Zusammen mit dem neuen Regelwerk RDA wurde auch auf die Gemeinsame Normdatei (GND) umgestellt.

3 Die ETH-Bibliothek ist Mitglied des NEBIS-Verbundes (NEBIS = Netzwerk von Bibliotheken und Informationsstellen in der Schweiz).

4 Für den IDS angepasstes MARC-Format. Der IDS ist der Informationsverbund Deutschschweiz, bei dem der NEBIS-Verbund Mitglied ist.

ten getestet und laufend angepasst werden. Auch in diesen Aufgaben konnte sich das Pilotteam Metadatenmanagement mit seiner erworbenen Fachkompetenz einbringen.

Des Weiteren erstellte das Pilotteam neue Arbeitshilfen für die Katalogisierenden, zum Beispiel in Form neuer interner Workflows und dazu passenden Satzschablonen. Schlussendlich organisierte es neben den offiziellen Schulungen des NEBIS-Verbundes und in Zusammenarbeit mit weiteren Mitarbeitenden für alle Katalogisierenden interne Kurzschulungen zu speziellen Themen (z.B. fortlaufende Ressourcen, Tipps und Tricks im Umgang mit Aleph, etc.).

In der ETH-Bibliothek war die Mitarbeit des Pilotteams bei der Umstellung auf das neue Regelwerk RDA und die Gemeinsame Normdatei GND sowie auf das erweiterte MARC-Format von großem Nutzen, da für diese Arbeit das interdisziplinäre Wissen der Metadatenspezialistinnen optimal eingesetzt werden konnte. Das neu aufgebaute Team hat maßgeblich dazu beigetragen, dass die Umstellung auf RDA und GND professionell und reibungslos ablaufen konnte.

Erkenntnisse nach zwei Jahren Pilotphase und Entwicklung

Eine der ersten Erkenntnisse des Pilotteams war, dass Arbeitsfelder, die nur mit Serverzugriff effizient erledigt werden können, bei der aktuellen Systemarchitektur nicht von einem breiten Feld von Mitarbeitenden, sondern weiterhin von geschultem IT-Perso-

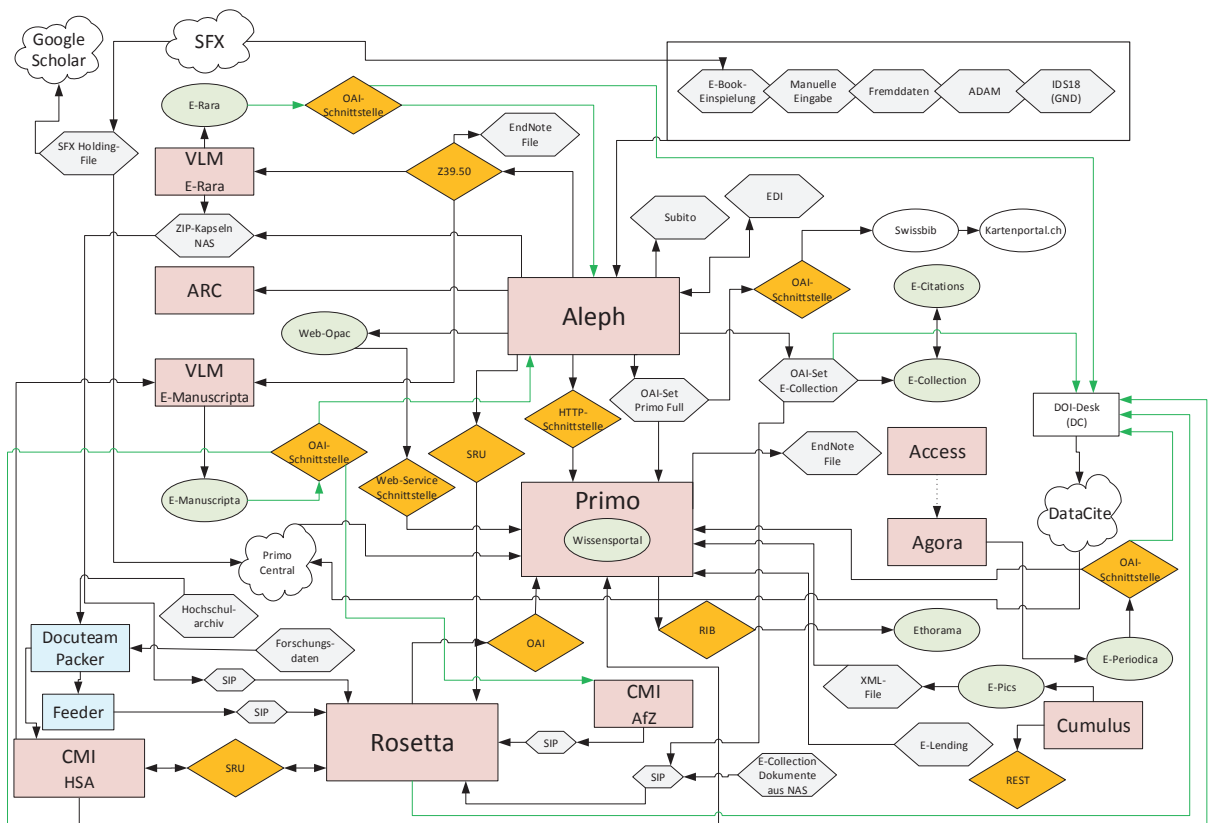
nal durchgeführt werden müssen. Die zwei Pilotteam-Mitarbeiterinnen aus der Medienbearbeitung hatten daher nur Zugriff auf den Testserver des Bibliothekssystems. Arbeiten auf dem Produktivserver wurden immer in Zusammenarbeit mit den Bibliotheks-IT-Services ausgeführt.

Im Laufe der Pilotphase wurden verschiedene Arbeitsfelder getestet und evaluiert. Schnell zeigte sich, dass die Datenanalyse eines der Hauptthemen im Metadatenmanagement ist. Für die Integration oder den Transfer von Metadaten in ein System oder auch um Metadaten mit weiteren Daten anreichern oder mutieren zu können, ist eine vorgängige Datenanalyse unumgänglich. Dabei müssen zum Teil hunderttausende von Records analysiert werden. Hierzu können geeignete Tools und Methoden eingesetzt werden, um die Daten effizient zu analysieren.

Ein weiteres Arbeitsfeld, das eng mit der Datenanalyse verbunden ist, ist die Erstellung von Datenmappings, die zur Unterstützung von Arbeiten in Projekten erstellt und bearbeitet werden. Durch diese und weitere unterstützende Mitarbeit in Projekten hat sich gezeigt, dass in der ETH-Bibliothek bereichsübergreifend ein Bedarf an vertieftem Wissen im Umgang mit Metadaten besteht.

Das Team wird sich nun vermehrt auch mit Linked (Open) Data, RDF und neuen Datenmodellen, wie zum Beispiel BIBFRAME, befassen. Zudem stehen immer mehr Tools für die Metadatenverwaltung und -analyse zur Verfügung, welche auch von Nicht-Informatikern

Abbildung 2: Datenflussdiagramm der ETH-Bibliothek



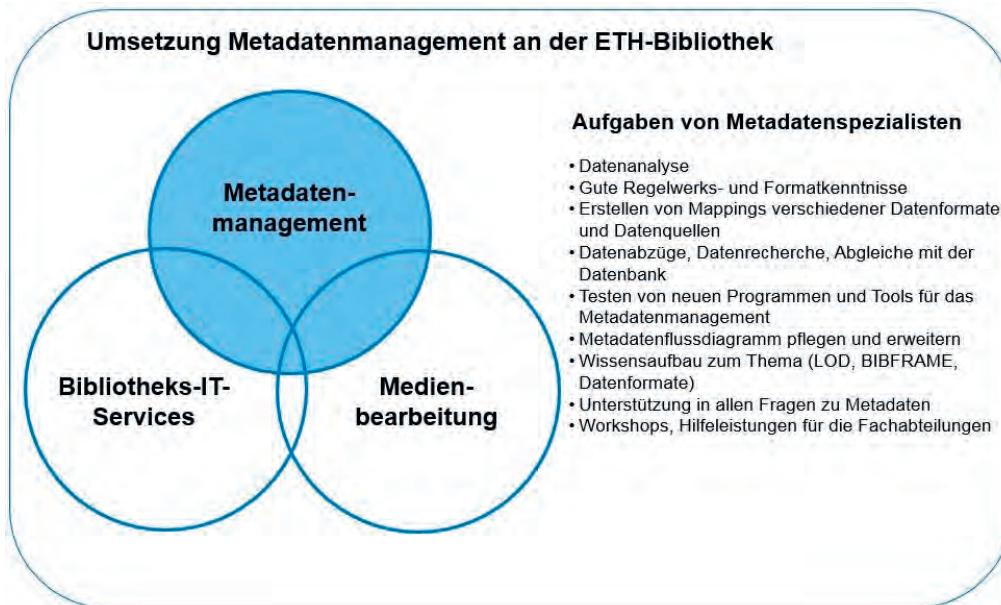


Abbildung 3: Umsetzung Metadatenmanagement an der ETH-Bibliothek

angewendet werden und somit neue Möglichkeiten für die Metadatenbearbeitung durch Bibliothekare und Bibliothekarinnen schaffen können. Diese Tools zu testen und deren Eignung für den Einsatz an der ETH-Bibliothek zu prüfen, wird ebenfalls eine Aufgabe der Metadatenspezialistinnen sein.

Fazit

Die bisher geleisteten Arbeiten zeigen, dass zahlreiche Aufgabengebiete existieren, in denen sich Metadatenspezialisten und -spezialistinnen zur Unterstützung einbringen können.

Mit den neu erlangten Qualifikationen ist es möglich, Hilfsmittel (z. B. Editor-Programme, Bibliothekssystem-Services, etc.) optimiert einzusetzen und damit Arbeitsabläufe effizienter zu gestalten. Zudem kann die Herangehensweise an verschiedene Aufgaben und deren Auswirkungen optimal geplant bzw. beurteilt werden.

Spezialisierte IT- und Katalogisierungsstellen sind weiterhin nicht wegzudenken. Neben verschiedenen Automatisierungsvorgängen wird die intellektuelle Erschließung auch in Zukunft ein wichtiger Teil der Bestandsbearbeitung sein, da korrekte und vollständige Datensätze eine solide Grundlage für die weitere automatisierte Verarbeitung und Anreicherung bilden. Durch die neuen Technologien hat sich jedoch ein zusätzliches Aufgabengebiet aufgetan, wo Metadatenspezialistinnen und -spezialisten mit ihrem interdisziplinären Wissen Aufgaben übernehmen können. Sie bilden somit eine wichtige Schnittstelle zwischen den Fachabteilungen und der IT.

Die teamübergreifende Zusammenarbeit hat sich in der ETH-Bibliothek sehr bewährt, ist etabliert und nicht mehr wegzudenken. Der gegenseitige Know-how-Transfer funktioniert gut und ist für beide Seiten

gewinnbringend. Über die Dauer der Pilotphase wurde das Team Metadatenmanagement ein gern genutzter Ansprechpartner bei allen Fragen zu den in der ETH-Bibliothek vorhandenen Metadaten, aber auch zur Verbesserung der Qualität.

Im Hinblick auf die Entwicklungen auf dem Markt, die sich hin zu Next-Generation-Systemen und weg von den klassischen Bibliothekssystemen bewegen, werden sich in Zukunft auch weitere Bibliothekarinnen und Bibliothekare vermehrt mit Metadatenmanagement befassen müssen. Aufgaben werden nicht mehr in einzelne Arbeitsschritte unterteilt, Verantwortung und Selbständigkeit werden mehr und mehr bei den Mitarbeitenden liegen. Auf diesem Weg zu neuen Systemen und Arbeitsabläufen werden Metadatenspezialistinnen und -spezialisten ihre Kenntnisse und Fähigkeiten optimal einbringen können. |



Erica Pfister

Monographische Medien
ETH-Bibliothek
erica.pfister@library.ethz.ch



Barbara Wittwer

Bibliotheks-IT-Services
ETH-Bibliothek
barbara.wittwer@library.ethz.ch



Marianne Wolff

Zeitschriften
ETH-Bibliothek
marianne.wolff@library.ethz.ch