

Frank Klaproth

Digitalisierung und Digitale Bibliotheken – 25 Jahre

Digitalisierung im Kontext der Deutschen

Universitätsbibliotheken

Zusammenfassung: Dieser Beitrag liefert einen Einblick in die Entwicklungsphasen der Digitalisierung deutscher Universitätsbibliotheken und stellt dabei auch die Einflüsse primär technischer Disziplinen auf die Gestaltung von Digitalen Bibliotheken dar.

Schlüsselwörter: Massendigitalisierung, Digitale Bibliothek, Technologiewandel

Digitization and Digital Libraries – 25 years of Digitization in the Context of German University Libraries

Abstract: This article provides an insight into the development phases of the digitization of German university libraries and presents the influence of primarily technical disciplines on the design of digital libraries.

Keywords: Mass digitization, digital library, technology change

1 Einleitung

Mitte der 1990er Jahre hatten die deutschen Universitätsbibliotheken bereits eine sehr gute Digitalisierungsreife in ihrem Kernbereich der Informationserfassung, Recherche und Bereitstellung inkl. der Softwarelösungen für Bibliotheksnutzer erreicht. Zu der Zeit lag der Schwerpunkt der Digitalisierung in der Unterstützung der manuellen Erfassung von bibliographischen Metadaten analoger Materialien und Medien inkl. Verschlagwortungen und Klassifizierung sowie deren Austausch zwischen Standorten der Bibliotheken. Datenträger wie Schallplatten, Videotapes, Disketten und CD-ROMs waren Bestandteil der Medien. Die Metadatenhaltung erfolgte typischerweise in Lokalen und / oder Zentralen Bibliothekssystemen, deren Erfassungsoberfläche den Katalogisierungsrichtlinien folgte. Der Datenaustausch wurde bereits seit längerem organisiert in Bibliotheksverbundstrukturen. Indexierungen von Daten bezogen sich auf bibliographische Metadaten und die Bereitstellung der Ergebnisse erfolgte über Katalogrechercheterminals für den Bibliotheksnutzer vor Ort sowie über den Online Public Access Catalogue (OPAC) zunächst

noch per Telnet-Protokoll dann ab ca. 1995 per HTTP-Protokoll ¹ und Webbrowser. Ein Wandel mit der Einfachheit des Ausbaus des World Wide Web im Internet zeichnete sich ab. Dieser Wandel erfasste auch die Universitätsbibliotheken.

2 Digitalisierungserkenntnis in Schritten

Mit Beginn und Ausbau von fachdisziplinspezifischen Webservern (ab Beginn 1994), noch auf Basis von NCSA HTTPd ² und den dort vorgehaltenen primären digitalen Objekten (zunächst HTML-codierte Seiten, häufig mit GIF-Images), begann eine Diskussion und Entwicklung von vereinfachten Metadatenschemata zur Beschreibung und Kategorisierung für elektronische Onlineressourcen. Diese mündete 1994 in die Gründung der Dublin Core Metadata Initiative (DCMI)³. Eine international wegweisende Entwicklung, die sich im Weiteren auch auf die Aktivitätsfelder der deutschen Universitätsbibliotheken auswirkte. Der bisher vergebliche Versuch ausgewählte Online-Ressourcen in den Katalogen der Bibliotheken mit Katalogisierungsrichtlinien in Einklang zu bringen wurde nun, auch mit Hilfe von Mapping-tools oder Crosswalks zwischen dublin.core Elementen und den klassischen Bibliotheksstandards, möglich. Es schlossen sich allerdings Fragen der schnell wachsenden Anzahl von digitalen Formaten und deren Beschreibungen sowie eindeutiger Identifikation von entweder leicht flüchtigen oder veränderbaren digitalen Online-Objekten an. Die reine Menge an neuen frei zugänglichen Onlineressourcen und dem gleichzeitigen Wunsch in den Bibliotheken das Wissen um diese Ressourcen strukturiert und möglichst klassifiziert an die Nutzer zu kommunizieren führte zu einer gänzlich neuen Diskussion. Mit der Erkenntnis hier ggf. selektiv vorgehen zu müssen und gekoppelt mit dem zeitgleich skizzierten Bild einer verteilten und komplett digitalen Bibliothek nach Tim Berners-Lee, begann eine Profilsuche der Universitätsbibliotheken. Hinzu kamen die sich schnell entwickelnden Suchmaschinentechnologien (damals noch AltaVista ⁴1995, Inktomi ⁵1996) basierend auf Metatag-Suchen, semistrukturierten Inhalten und dem Ziel vollautomatisiert große Teile des WWW zu indexieren und mit wenigen Schritten den Suchenden vermeintlich vollständige Trefferlisten anzuzeigen.

¹ <https://datatracker.ietf.org/doc/html/rfc1945>

² <https://web.archive.org/web/20090415224404/http://hoohoo.ncsa.uiuc.edu/>

³ <http://dublincore.org/>

⁴ <https://de.wikipedia.org/wiki/AltaVista>

⁵ <https://en.wikipedia.org/wiki/Inktomi>

Auf der anderen Seite gab und gibt es in den Beständen der Universitätsbibliotheken z.T. große Mengen an analoger Information, die im Rahmen des deutschen Urheberrechts digitalisiert und zur Nutzung frei verfügbar gemacht werden können. Allerdings fehlte die Methodik und Prozesse für die Digitalisierung (vor allem für die Massendigitalisierung) sowie für die Bereitstellung der Daten.

Als Folge aus den obigen Situationen ergaben sich seitdem neue Herausforderungen und Chancen, für die Bibliotheken auch entlang neuer Technologien den Teil der Digitalen Bibliotheken zu erkunden und auszubauen.

3 Ausbildung von Komponenten der Digitalen Bibliotheken

3.1 Portale für die Wissenschaft

Mit Hilfe von Fördermitteln der Deutschen Forschungsgemeinschaft wurden bereits seit 1995 die Entwicklung von Portaltechnologien (mit Webis⁶ zunächst basierend auf Apache httpd) gefördert, die thematische Online Fach Guides (SSGFIs⁷, 1997) zum Ziel hatten. Diese zunächst händisch und intellektuell gepflegten sowie von wissenschaftlichen Fachpersonal mit Hilfe einer Fachcommunity kommentierten Guides basierten auf Linklisten und boten den WissenschaftlerInnen einen Einstieg in die Online-Fachressourcen. Ein weiterer Schritt, die jeweilige Fachcommunity mit Informationen zu versorgen waren Portale der Virtuellen Fachbibliotheken (ViFas). Hierbei wurden z.B. Recherchen und Ergebnislisten aus den thematischen OPAC-Suchen und den Online Fachguides verschnitten. Zudem wurden Facettierungen wie z.B. Erscheinungsdatum und Materialart (gedruckt, online, ...) mit Hilfe der technischen Schnittstellenanbindung einer verteilten Suche über mehrere Quellen (z.B. lokales Bibliothekssystem, Dokumentenrepository) hinweg ermöglicht. Bspw. auf Basis der Retrievalprotokolle Z39.50⁸ und SRU⁹. In den ViFas kommen Informationen zu verfügbaren Lizenzen und Allgemeine Informationen zu Veranstaltungen, Blogs o.ä. hinzu. Dafür werden häufig WebCMS Systeme wie z.B. TYPO3¹⁰ oder Drupal¹¹ usw. genutzt.

⁶ Ahlers, T.; Enderle, W. (1997). In: Bibliotheksdienst Jg. 31, Heft 2

⁷ <http://webdoc.sub.gwdg.de/ebook/aw/ssgfiwork/rutz.htm>

⁸ <https://de.wikipedia.org/wiki/Z39.50>

⁹ https://de.wikipedia.org/wiki/Search/Retrieve_via_URL

¹⁰ <https://typo3.org/>

¹¹ <https://www.drupal.org/>

Ab 2016 ist auch auf Grund der Einführung sogenannter Fachinformationsdienste (FID¹²) ein weiterer Wandel in den thematischen Portalzuschnitten, aber auch in der Ergänzung der Qualität der Services erfolgt. WissenschaftlerInnen und Fachcommunities erhalten mit Authentifizierungs- und Autorisierungsverfahren z.B. über Virtuelle Homeorganisationen per Single Sign On (SSO) und Shibboleth¹³-Verfahren deutschlandweit direkten Zugriff auf lizenzpflichtige Forschungsquellen. Dabei müssen sowohl die Universitätsinfrastrukturen (hier z.B. Verzeichnisdienste wie LDAP¹⁴), als auch die Angebots- und Dienststrukturen der Bibliotheken aufeinander abgestimmt bzw. automatisiert werden.

3.2 RetroDigitalisate für die Wissenschaft

Im Jahr 1997 schloss sich ein umfangreiches Programm zur Gestaltung und Aufbau einer Verteilten Digitalen Forschungsbibliothek in Deutschland an. Dabei wurden u.a. zwei Kompetenzzentren für die Retrodigitalisierung (MDZ¹⁵ an der Bayrischen Staatsbibliothek in München und das GDZ¹⁶ an der Niedersächsischen Staats- und Universitätsbibliothek in Göttingen) gefördert.

Bei der Digitalisierung der unterschiedlichen Materialien konnte auf Erfahrungen der amerikanischen Bibliotheken an den Universitäten Cornell, Harvard, Virginia sowie an der Library of Congress in Washington zurückgegriffen werden.

Eine Vielzahl von Retrodigitalisierungsvorhaben wurde über die Bibliotheksbestände hinweg durchgeführt, die sich der entwickelten Digitalisierungsstandards (TIFF – Formate und Metadaten hier bereits gemeinsam technische Bilddaten als auch bibliographische Metadaten, basierend auf XML/RDF) der Zentren bedienen. Dabei mit dem Ziel sowohl zeitliche Lücken (z.B. 15. oder 18. Jahrhundert), als auch thematische Lücken (z.B. Mathematica oder Zoologica) der digitalen Bestände zu schließen. Diese sind in einem Zentralen Verzeichnis digitalisierter Drucke (ZVDD¹⁷) gelistet.

¹²

https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/fachinfodienste_wissenschaft/index.html

¹³ <https://www.shibboleth.net/>

¹⁴ https://de.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

¹⁵ <https://www.digitale-sammlungen.de/de>

¹⁶ <https://gdz.sub.uni-goettingen.de/>

¹⁷ <https://www.zvdd.de/>

Für die Verwaltung und Anzeige der wachsenden digitalen Dokumentensammlungen wurden erfahrene Kooperationspartner auch aus dem privatwirtschaftlichen Sektor gesucht.

Es entwickelten sich somit Lösungen wie z.B. Agora oder Miles (später MyCore¹⁸), die auch bereits Java basierte Servlet-Technologien für die Präsentationsschichten der Rechercheergebnisse einsetzen.

Der weiter zunehmende Digitalisierungsanteil und know how Anwuchs in den Bibliotheken führte bei gleichzeitiger Mittelstagnation bzw. Reduktion und breiterer Verfügbarkeit von OpenSource-Softwarekomponenten zu einer stufenweisen Lösung von möglichen Anbieterabhängigkeiten (vendor lock-in).

Es entstanden gepflegte OpenSource-Lösungen wie z.B. Goobi¹⁹(2004) und zuletzt kitodo²⁰(2016) für die Abbildung und Steuerung eines kompletten Retrodigitalisierungsworkflows mit modernsten Softwaretechnologiestacks.

3.3 Suchmaschinen für die Wissenschaft

Mit Aufkommen der Webcrawler und Suchmaschinen, die einen hohen Grad der Integration in die Webbrowserfunktionalitäten und das Potential der Indexierung vieler digitaler Objekttypen besitzen, nahm das Bild einer globalen Digitalen Bibliothek Konturen an. Das Ziel mit einer Suchanfrage in „allen“ verfügbaren Informationsquellen eine relevante Treffermenge als Antwort zu erhalten klingt verlockend. Neben den Fragen der Relevanz, des Rankings, der Suchtiefe, der Menge, der Zeichensätze, der Semantiken ist dies oft nur domainspezifisch zu beantworten und somit auch in naher Zukunft nicht realistisch. Allerdings bietet die Technologie für Bibliotheken in Kopplung mit Erfahrungen von bibliographischen Indexsystemen eine sehr performante und gute Möglichkeit die Metadaten der digitalen Objekte zu indexieren. Und je nach Qualität der Metadaten auch Facettierungen und browsing anzubieten, so dass Treffermengen qualifiziert eingegrenzt werden können. Ein sehr gutes Beispiel stellt die Entwicklung des 2004 gelaunchten Service Bielefeld Academic Search Engine (BASE²¹) dar. Der Service wird ständig mit Inhalten und

¹⁸ <https://www.mycore.de/>

¹⁹ <https://goobi.io/>

²⁰ <https://www.kitodo.org/>

²¹ <https://www.base-search.net/>

tlw. Funktionen erweitert und bietet auch Zugriff auf Inhalte aus Repositorien, die nicht von den kommerziellen Suchmaschinen indiziert werden.

Aber auch Angebote wie Google Scholar ²²(2004 gelauncht) bietet mit Hilfe der Suchmaschinentechologie und den zur Verfügung gestellten Metadaten und Volltexten sehr gute, sogar disziplinübergreifende Suchergebnisse.

3.4 Repositorien für die Wissenschaft

Mit wachsender Zahl an digitalen Publikationen und der Erzeugung von Forschungsdaten entstanden neben den Angeboten der Verlage Initiativen aus den Fachdisziplinen und Bibliotheken heraus diese Objekte in Online-Repositorien zu erfassen, zu klassifizieren und der Community und auch der Öffentlichkeit zugänglich zu machen. Die eingesetzten Softwareumgebungen basieren u.a. auf DSpace²³, EPrints²⁴, OPUS²⁵, PURE ²⁶ usw. Eine gute Übersicht der weltweit betriebenen Repositorien findet sich im OpenDOAR ²⁷(gelaunched 2005) Verzeichnis. So sind laut OpenDOAR für Deutschland z.Z. 284 qualifizierte Repositorien dort gelistet. Die Repositorien bieten Schnittstellen mit denen in Form des Open Access auf die Inhalte zugegriffen werden kann. Entweder mit Hilfe einer eigenen Portalseite oder auch über die Integration in andere Portale und Suchmaschinen.

3.5 Virtuelle Forschungsumgebungen für die Wissenschaft

Um primär digitale Arbeitsergebnisse von WissenschaftlerInnen auch in einer kollaborativen digitalen Umgebung abzubilden, wurden virtuelle Forschungsumgebungen geschaffen. Bereits 2005 wurden dazu drei JISC geförderte Projekte von Michael Fraser ²⁸vorgelegt, die die Idee eines Virtual Research Enviroment (VRE) prototypisch umsetzen sollte. Dies wurde möglich, da sich einerseits die Web-Technologien und die Webstandards dergestalt weiterentwickelten, dass sich komplexe Aufgaben und Programmabläufe in einer Webbrowser Umgebung abbilden ließen. Andererseits erfordert dies auf der Serverseite entsprechend ausreichende Ressourcen wie z.B. Hauptspeicher, Plattenplatz und auch Bandbreite für den Netzwerkverkehr. Da sich diese Komponenten dynamisch

²² <https://scholar.google.com/>

²³ <https://duraspace.org/dspace/>

²⁴ <https://www.eprints.org/>

²⁵ <https://www.opus-repository.org/>

²⁶ <https://libguides.coventry.ac.uk/rsp/pure>

²⁷ <https://v2.sherpa.ac.uk/opensoar/>

²⁸ <http://www.ariadne.ac.uk/issue/44/fraser/>

weiterentwickelten, konnte z.B. auch 2011 ein Angebot in einen Forschungsverbund TextGrid²⁹ für die Geistes- und Kulturwissenschaften initialisiert werden. Die VREs nutzen mit Hilfe der Kombination bereits vorhandener technischer Lösungen sowie mit Hilfe von Datenschnittstellen Datenobjekte, die geändert, kommentiert oder gar neu geschaffen werden. Es werden mit Hilfe von Bearbeitungs- und Analysewerkzeugen auch Digitale Editionen gänzlich neu geschaffen. Universitätsbibliotheken erarbeiten dafür die passenden Softwarelösungen im Team mit den WissenschaftlerInnen.

3.6 Weitere Komponenten für die Wissenschaft

Hier ist vor allem die Archivierung (Langzeitarchivierung – LZA) der digitalen Daten zu nennen. Ein internationales Vorgehen im Bereich der Bibliotheken, insbesondere der Nationalbibliotheken mit ihren Sammlungsaufträgen führte 2004 in Deutschland zu einem BMBF geförderten Projekt „kopal³⁰“. In einer Kooperation von Nationalbibliothek, Universitätsbibliothek, Universitätsrechenzentrum und Industriepartner wurde eine Lösung auf international basierenden LZA-Standards (Open Archival Information System - OAIS³¹) entwickelt. In der Nachfolge schlossen sich mehrere LZA-Projekte an. Zudem entwickelten sich weitere LZA-Lösungen, sowohl kommerzieller Art closed source, als auch OpenSource basiert.

Eine weitere wichtige Kernkomponente stellt die Identifikation der digitalen Objekte dar.

Hier haben sich verschiedene Persistent Identifier (PID) Systeme auf Basis der handle-Technologie und unter Einsatz von URN-Resolver etabliert. So z.B. ePIC³², DOI³³, DNB-URN Service³⁴. So werden auf unterschiedliche Granularitätsebene von Objekten oder ganzen Sammlungen PIDs vergeben. Die Stabilität und langfristige Verfügbarkeit der PID-Services ist von entscheidender Bedeutung, weshalb sich auch Konsortien und Rechenzentren dafür zusammengeschlossen haben.

Volltexterzeugung, auch aus den digitalisierten analogen Vorlagen (Handschriften, Briefe, aber auch z.T. gedruckte Vorlagen) beschäftigt die Bibliotheken und Dienstleister schon seit

²⁹ <https://textgrid.de/>

³⁰ <http://kopal.langzeitarchivierung.de/>

³¹ <http://www.oais.info/>

³² <https://www.pidconsortium.net/>

³³ <https://www.doi.org/>

³⁴ https://www.dnb.de/DE/Professionell/Services/URN-Service/urn-service_node.html

der späten 1990er. Nicht alle Materialien können zufriedenstellend automatisiert Volltext erfasst werden. Jedoch bietet ein Teil der Volltexte zusammen mit den anwachsenden originär digital verfügbaren Daten inzwischen ein Fundus an Forschungsmaterial. Hier müssen noch die Angebote für die Wissenschaft deutlicher über die Bibliotheken hinweg herausgearbeitet und im Dialog kommuniziert werden.

Es sei erwähnt, dass eine Professionalisierung im Bereich der Softwareentwicklung und Softwarebereitstellung auch für die Universitätsbibliotheken über die letzten Jahre stark zugenommen hat. Dies gelingt u.a. durch Kooperationen mit Unternehmen, in Community Konsortien, aber auch durch konkrete Anwendung von gängigen Standards (z.B. ITIL³⁵) und Methoden (z.B. SCRUM³⁶) wie z.B. Nutzung von Softwarerepositorien (Github, Gitlab, ...). Zudem nutzen Entwicklungen im Bereich der Digitalen Bibliotheken internationale Frameworks wie z.B. Fedora³⁷.

4 Fazit und Ausblick

Die Universitätsbibliotheken haben sich innerhalb der letzten 25 Jahre einem steten und zuletzt intensiveren und anhaltenden Wandel gestellt. Kommend aus einer Situation als Anbieter von Information (Beschaffung, Recherche, physikalischer Ausleihe, Rücknahme), nun bereits als aktiver Mitgestalter für Informationsstrukturen, dort nahe am Wissenschaftsprozess positioniert und als Entwickler von Softwarelösungen für Digitale Bibliotheken etabliert. Letzteres erfolgt tlw. vor und bei Erzeugung von wissenschaftlichen Ergebnissen oder auch deren Zwischenprodukten. So kann z.B. die Rolle des Data Steward, für die Einhaltung der Qualität der Daten und Metadaten anteilig oder ganz aus der Bibliothek heraus beigesteuert werden. Eine Softwareentwicklung und der Einsatz von agilen Entwicklungsmethoden, sowie die Konzertierung von Hard- und Softwareumgebungen ist Teil der projektorientierten Vorgangsmodelle geworden. Die breiter und tiefer greifende Digitalisierung der Services für die Wissenschaft sowie stetig neue Technologien im Bereich der Hard- und Software stellen Herausforderungen aber auch neue Lösungsmöglichkeiten dar. Die Zahl der Komponenten für den Betrieb und Funktionserweiterung Digitaler Bibliotheken wächst stetig. Die Auswahl aus dem Angebot stellt neue Herausforderungen

³⁵ <https://de.wikipedia.org/wiki/ITIL>

³⁶ <https://www.scrum.org/>

³⁷ <https://duraspace.org/fedora/>

dar. Hier gilt es den Serviceaspekt und die Nachhaltigkeit abgestimmt zusammen mit Partnern in den Fokus zu rücken. Dabei ist das Zusammenspiel von Digitalen Infrastrukturanbietern wie z.B. lokal, aber auch überregional agierenden Rechenzentren, Verbundzentralen, Medienzentren und Bibliotheken weiterhin der Schlüssel für die passende Gestaltung der Wissensinfrastrukturlandschaft.

Die Universitätsbibliotheken sind gut gerüstet denn, der angenommene Wandel wird als steter Prozess, die ersten Bibliotheken als aktiver Partner der Wissensinfrastruktur wahrgenommen und die Zukunft ganz im Sinne „Stets die NutzerIn, den Nutzen und den Service der Bibliothek im Blick“ gelebt.

Literaturverzeichnis

Godby, Carol Jean; Young, Jeffrey A.; Childress, Eric (2004): A Repository of Metadata Crosswalks. In: *D-Lib Magazine*, 10(12). Verfügbar unter <https://dlib.org/dlib/december04/godby/12godby.html>.

Miller, James S. (1996): W3C and Digital Libraries. In: *D-Lib Magazine*, 2(11). Verfügbar unter <https://www.dlib.org/dlib/november96/11miller.html>.

Summan, F.; Lossau, N. (2004): Search engine technology and digital libraries – Moving from theory to practice. In: *D-Lib Magazine*, 10(9). Verfügbar unter <https://doi.org/10.1045/september2004-lossau>.

Frank Klaproth
Georg-August-Universität Göttingen
Niedersächsische Staats- und Universitätsbibliothek
Platz der Göttinger Sieben 1
D-37073 Göttingen
klaproth@sub.uni-goettingen.de
ORCID ID: <https://orcid.org/0000-0002-5957-3597>