



Die Abbildung zeigt die technische Infrastruktur des Bibliothekarischen Archivierungs- und Bereitstellungssystems (BABS).

## Aufbau einer Infrastruktur zur digitalen Langzeitarchivierung

### Der Weg der Bayerischen Staatsbibliothek

Im Januar 2011 wurde das von der DFG geförderte Projekt BABS2: „Vertrauenswürdige und skalierbare Langzeitarchivierung“ der Bayerischen Staatsbibliothek in Kooperation mit dem Leibniz-Rechenzentrum mit einem Workshop zum Thema „Langzeitarchivierung von Digitalisaten“ erfolgreich abgeschlossen. Dies ist ein willkommener Anlass, einmal die bisherigen Entwicklungen und Erfahrungen mit diesem sehr komplexen Thema an der Bayerischen Staatsbibliothek zu rekapitulieren.

Von Astrid Schoger

## Grundsätze der digitalen Langzeitarchivierung

Das Erbe unserer Gesellschaft – wissenschaftliche Erkenntnisse, historische Ereignisse und kulturelle Leistungen – wird zunehmend nur noch in digitaler Form dokumentiert. Aber auch die Werke vergangener Jahrhunderte werden in zunehmender Zahl digital wiederveröffentlicht und verbreitet. Dank rasanter Fortschritte in der Digitalisierungstechnik ist inzwischen die Vorstellung gar nicht mehr utopisch, dass schon in wenigen Jahrzehnten alle jemals durch die Menschheit gedruckten Bücher auch in digitaler Form vorliegen könnten.

Die Vorteile einer rein digitalen Herstellung beziehungsweise der Retrodigitalisierung liegen auf der Hand: Neben einer einfacheren und schnelleren Publikation und einer nahezu uneingeschränkten weltweiten Verfügbarkeit ist es vor allem der Mehrwert in der Nutzung. Selbst große, verteilte Datenbestände sind bei richtiger Erschließung und dem Einsatz moderner Werkzeuge binnen Augenblicken durchsuchbar, nach wenigen Klicks ist das Gesuchte da. Der grundlegende Wandel der Technologien bringt es mit sich, dass digitale Publikationen nunmehr einen wesentlichen Anteil am Bestand von Bibliotheken, die sich der Bewahrung und der Bereitstellung dieses Erbes annehmen, ausmachen. Der Wert der Dokumente ergibt sich nicht nur aus ihrer Einzigartigkeit oder dem Mehrwert der digitalen Form für die Nutzung. Die Digitalisierung beziehungsweise Produktion digitaler Dokumente ist teuer. Dem Bestreben, die Verfügbarkeit dieser Dokumente langfristig zu sichern, liegen also viele immaterielle, aber auch sehr materielle Motive zugrunde. Doch wenn die eigentliche Arbeit der Digitalisierung beziehungsweise elektronischen Produktion getan ist, fangen die Probleme und Herausforderungen erst an. Das Ziel, die Dokumente selbst und deren Nutzbarkeit auf Dauer zu erhalten, ist durch vielfältige organisatorische und technische Risiken bedroht. Als Beispiele seien genannt:

- die begrenzte Lebensdauer einzelner Datenträger: Einerseits verfallen die Trägermedien physisch, andererseits veralten sie in technischer Hinsicht, mit der Konsequenz, dass oft keine geeigneten Abspielumgebungen mehr zur Verfügung stehen,
- weitere Abhängigkeiten durch die permanente Fortentwicklung der Dateiformate und die Kurzlebigkeit der zugehörigen Anwendungsprogramme, Betriebssysteme und Geräte,
- die Flüchtigkeit von Publikationen im Internet, die häufig bereits nach kurzer Zeit nicht mehr auffindbar oder nur noch unter einer veränderten Internetadresse oder in inhaltlich stark veränderter Form erreichbar sind,

- elektronische Zeitschriften und Bücher, die durch befristete Lizenzverträge mit Verlagen nur in einem bestimmten Zeitraum zur Verfügung stehen.

Notwendig ist ein aktives Risikomanagement, welches seinerseits eine Vielzahl von Maßnahmen auf organisatorischer und technischer Ebene beinhaltet. Die Gesamtheit aller Maßnahmen zum Erhalt eines digitalen Bestandes und seiner Nutzbarkeit über lange Zeiträume hinweg nennt man digitale Langzeitarchivierung. Dazu gehören nach dem OAIS-Referenzmodell (Reference Model for an Open Archival Information System, ISO-Standard 14721:2003) für die Langzeitarchivierung:

- die gezielte Auswahl des archivwürdigen Materials, also letztlich die Entwicklung und die fortlaufende Pflege eines Archivierungsprofils,
- eine adäquate Erschließung durch die eindeutige und dauerhafte Identifikation der archivierten Materialien, die Ausbringung von beschreibenden und strukturellen Metadaten, die technische Charakterisierung der Dateien sowie die Dokumentation aller Veränderungen,
- die Archivspeicherung, das heißt der dauerhafte Erhalt der digitalen Daten auch über die Lebensdauer einzelner Speichermedien hinaus durch ständiges Umkopieren auf neue Medien,
- die Bereitstellung der digitalen Daten für die Nutzung,
- das Management und die Administration des digitalen Langzeitarchivs

und insbesondere

- die Planung sowie die Durchführung der Langzeiterhaltungsmaßnahmen, also die Migration (Konvertierung veralteter Dateiformate in aktuelle, standardisierte, offene oder offen gelegte Formate) oder auch die Emulation (die Nachbildung alter Betriebssysteme und Anwendungsprogramme auf neuen Rechnern), zum Erhalt der Interpretierbarkeit der Daten auch für zukünftige Nutzungen.

### Erste Schritte an der Bayerischen Staatsbibliothek: Langzeitarchivierung der Digitalisate

Die Bayerische Staatsbibliothek (BSB) beziehungsweise deren dafür zuständiges Referat Münchener Digitalisierungszentrum/Digitale Bibliothek (MDZ) hat sich bereits früh, das heißt im Rahmen ihrer weit ausholenden Digitalisierungsbemühungen, intensiv mit diesem Thema auseinandergesetzt. Der

**Münchener Digitalisierungszentrum:**  
[www.digitale-sammlungen.de/](http://www.digitale-sammlungen.de/)

**Bibliothekarische  
Archivierungs-  
und Bereitstellungssystem BABS:  
www.babs-  
muenchen.de/**

erste konkrete Schritt hin zu einer systematischen digitalen Langzeitarchivierung wurde 2004 unternommen. Damals wurde in Zusammenarbeit mit dem Leibniz-Rechenzentrum (LRZ) die gesamte bis dahin auf CD gespeicherte Produktion aus den MDZ-Retrodigitalisierungsprojekten (über 3.000 CDs) in das Archivsystem des LRZ überspielt. Der Schritt, die CDs für diesen Zweck ganz abzuschaffen, war sowohl von der Sorge um die Haltbarkeit der Datenträger getragen als auch im Bewusstsein, dass der schnell wachsende Datenbestand mit einem Archiv einzelner CDs oder DVDs schon bald nicht mehr zu handhaben sein würde. Bereits in den ersten Digitalisierungsprojekten hat das MDZ großen Wert auf die Archivierungsfähigkeit der Daten aus der retrodigitalen Produktion gelegt. So wurden offene, standardisierte beziehungsweise de-facto-Standards darstellende Dateiformate wie das Tagged Image File Format (TIFF) und Standard Generalized Markup Language (SGML) sowie später Extended Markup Language (XML) eingesetzt. Mit der Entwicklung und dem nachfolgenden Einsatz der ZEND (= Zentrale Erfassungs- und Nachweisdatenbank für Digitalisate) wurde bereits seit 2005 Sorge dafür getragen, dass der computerbasierte Digitalisierungsworkflow eine automatische Komponente zur Weitergabe der Daten zwecks Langzeitarchivierung an das LRZ beinhaltet.

Die strategische Kooperation mit dem LRZ im Bereich der digitalen Langzeitarchivierung wurde in den folgenden Jahren schrittweise vertieft.

### **BABS1: Das Bibliothekarische Archivierungs- und Bereitstellungssystem als Pilot**

Es wurde schnell deutlich, dass es mit der Abschaffung der CDs allein nicht getan ist. Zur Sammlung weiterer systematischer Erfahrungen mit der Konzeption und Umsetzung der Langzeitarchivierung von digital produzierten Netzpublikationen und Digitalisaten hat das MDZ deshalb in Kooperation mit dem LRZ und mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG) in den Jahren 2005 – 2007 eine organisatorisch-technische Infrastruktur zur Langzeitarchivierung als Pilotsystem

**nestor: www.langzeitarchivierung.de**

aufgebaut, das Bibliothekarische Archivierungs- und Bereitstellungssystem (BABS). Mit BABS wurde zunächst ein pragmatischer Weg beschritten. Ohne die Gesamtanforderungen der Langzeitarchivierung, insbesondere die Planung und Durchführung von Langzeiterhaltungsmaßnahmen, aus dem Auge zu verlieren, galten die Bemühungen zunächst dem Aufbau einer produktiven Infrastruktur zur Erfassung, Verwaltung, Archivierung und Bereitstellung digitaler Publikationen.

Die dadurch geschaffene technische Infrastruktur für BABS vereint seit rund fünf Jahren erfolgreich folgende in der BSB und dem LRZ vorhandene technische Komponenten:

- die ZEND, die Zentrale Erfassungs- und Nachweisdatenbank für Digitalisate, die als zentrale Produktions- und Bereitstellungsumgebung des MDZ im Rahmen von BABS auch die automatische Archivierung der Digitalisate ermöglicht,
- das Digital Asset Management System Digi-Tool der Firma Ex Libris, das im Rahmen von BABS für die Verwaltung und Bereitstellung von Netzpublikationen von Drittanbietern (Verlagen, Behörden, wissenschaftlichen Einrichtungen) eingesetzt wird und ebenfalls die Archivierung der Daten leistet, sowie
- das im Hintergrund angeschlossene Archiv- und Backupsystem Tivoli Storage Manager (TSM) als Archivspeicher am LRZ, wo die Dateien in robotergesteuerten Magnetbandbibliotheken gespeichert werden.

### **BABS2: Der Ausbau der Infrastruktur zu einem vertrauenswürdigen und skalierbaren digitalen Langzeitarchiv**

Nach dem erfolgreichen Aufbau von BABS begann eine Konsolidierungsphase. In einem auf den Ergebnissen und den einschlägigen Erfahrungen des oben genannten Projekts und der aktiven Beteiligung am Kooperationsprojekt nestor, dem deutschen Kompetenznetzwerk für Langzeitarchivierung, aufsetzenden weiteren DFG-Projekt wurde in den Jahren 2009 – 2010 die inzwischen geschaffene Infrastruktur zu einer skalierbaren und vertrauenswürdigen, den nestor-Kriterien genügenden Lösung für die Langzeitarchivierung ausgebaut. Ziel war es, einerseits den rasant steigenden Zahlen von Dateien und Datenvolumina (rund 300 Terabyte in über 500 Millionen Dateien aus der Produktion des MDZ und seiner Kooperationspartner) gewachsen zu sein und anderer-

seits den gestiegenen Qualitäts- und Nachhaltigkeitsanforderungen, die sich inzwischen auch aus einer neuen beziehungsweise erweiterten Ablieferungspflicht für gedruckte und digitale Veröffentlichungen ergeben, zu genügen. Konkret ist der 2009 in Kraft getretene Erlass der Bayerischen Staatsregierung zu nennen, in welchem die Abgabe amtlicher Veröffentlichungen und damit auch die Archivierungsverpflichtung der BSB auf Netzpublikationen ausgeweitet wurde (s. auch Nina Balz und Karin Knaf: Der Ablieferungspflicht unterliegen jetzt auch Netzpublikationen. In: BFB 3 (2009) 2, S. 128–129).

Zur Umsetzung des Erlasses wurde ein komfortables und sicheres Ablieferungsverfahren für die digitalen amtlichen Publikationen der Behörden, Dienststellen und Einrichtungen des Freistaates Bayern erarbeitet und implementiert. Über eine Webschnittstelle, das sogenannte Deposit, können die einschlägigen digitalen Dokumente von den abliefernden Stellen selbst relativ einfach und schnell übermittelt werden. Die Dokumente werden anschließend von Mitarbeitern der BSB nochmals einzeln gesichtet, katalogisiert und langzeitarchiviert. Begonnen wurde 2009 mit der Ablieferung von einschlägigen Veröffentlichungen durch die obersten Dienstbehörden, das heißt die bayerischen Staatsministerien. Deren damit befasste Mitarbeiter wurden vorab in einer eigens anberaumten Informationsveranstaltung sowohl über die Änderungen in den rechtlichen Bestimmungen als auch über das neue digitale Ablieferungsverfahren informiert und ganz konkret in die Bedienung der Web-Schnittstelle eingeführt. In einem zweiten Schritt wurden ab 2010 auch die Behörden der Mittelstufe, die über einschlägige Veröffentlichungen verfügen, einbezogen. Gleichwohl bleiben die Regelungen und Absprachen zur Abgabe von gedruckten Publikationen (im Falle von parallel erscheinenden digitalen Veröffentlichungen) davon bis auf weiteres unberührt.

Die Verfahren zur Sammlung und Archivierung von Netzpublikationen werden laufend ausgebaut. Einerseits wird der Kreis der potentiellen Ablieferer immer größer, andererseits erweitert sich das

Spektrum der Dokumente erheblich. Waren zunächst nur buch- oder zeitschriftenähnliche digitale Publikationen im Sinne von parallelen oder genuin digitalen Veröffentlichungen zu archivieren, so stellte sich zunehmend die Herausforderung, auch typische Internetquellen, wie z. B. vollständige Websites, zu sammeln, zu erschließen und zu erhalten. Dafür wurde an der BSB ein integriertes Verfahren von der Genehmigungseinholung über den Harvest-Prozess und die Archivierung bis hin zur Katalogisierung im lokalen OPAC implementiert, das das von der neuseeländischen Nationalbibliothek in Kooperation mit der British Library entwickelte Web Curator Tool verwendet. In einer Pilotphase werden 2011 damit zunächst die neu erfassten Internetressourcen der sechs Virtuellen Fachbibliotheken der BSB (Altertums-, Musik-, Bibliotheks-, Buch- und Informationswissenschaften, Osteuropa, Romanischer Kulturkreis) sowie jene mit Bezug zu Bayern archiviert.

Die großen Datenmengen und die komplexen Prozesse in der Praxis der digitalen Langzeitarchivierung verlangen nach einem besonders umfassenden strategischen Planungsansatz der Langzeitarchivierungsmaßnahmen. Letzterer muss neben möglichst detaillierten Verfahrensanforderungen beziehungsweise -vorgaben immer auch die jeweiligen technischen, organisatorischen und finanziellen Rahmenbedingungen berücksichtigen. Im Rahmen des oben genannten DFG-Projekts wurden verschiedene Alternativen im Bereich der Migration von Dateiformaten getestet und bewertet, um eine begründete und nachvollziehbare Entscheidung im Erhaltungsprozess von digitalen Objekten zu ermöglichen. Ferner wurden im Projekt Erfahrungen mit der Anwendung innovativer Methoden zur Bewertung der Vertrauenswürdigkeit – einem ganz wesentlichen Qualitäts- und Nachhaltigkeitsmerkmal der Langzeitarchivierung – auf der Grundlage der nestor-Kriterien gesammelt. Es zeigte sich, dass bei einigen Anforderungen erheblicher Handlungsbedarf bestand, weswegen umgehend Verbesserungsmaßnahmen in die Wege geleitet wurden, die zum einen dem stetig gewachsenen Problembewusstsein gerecht werden und zum anderen den wachsenden technischen Möglichkeiten Rechnung tragen. Darüber hinaus wurden Methoden des Risikomanagements für digitale Langzeitarchive erprobt. Diese reichten von der Identifizierung der Risiken über

die Bewertung von deren Eintrittswahrscheinlichkeiten und Schweregrade bis hin zur Spezifikation von Gegenmaßnahmen.

Die Skalierbarkeit der einzelnen Teile des Archivsystems und der gesamten technischen Infrastruktur von BABS wurde vom technischen Partner, dem LRZ, eingehend getestet und die Speicherstruktur daraufhin verbessert, indem sie virtuell weiter unterteilt wurde. So kann durch eine Reduzierung der Anzahl der Dateien und des Datenvolumens pro Einheit das Datenmanagement optimiert werden, und die Migration der Daten auf neue Generationen von Hardware (Plattenspeicher, Magnetbandsysteme) ist wesentlich einfacher und ohne Störung des laufenden Betriebs durchzuführen.



**DIE AUTORIN**

**Dr. Astrid Schoger**  
ist leitende Referentin für Langzeitarchivierung im Referat Münchener Digitalisierungszentrum/Digitale Bibliothek der Bayerischen Staatsbibliothek.

**Die Zukunft mit Rosetta**

Einen wichtigen Meilenstein in der Entwicklung der Bayerischen Staatsbibliothek und des Bibliotheksverbunds Bayern – wie es der Generaldirektor der Bayerischen Staatsbibliothek Rolf Griebel bei der Verbundkonferenz 2009 nannte (vgl. Rolf Griebel: Der Bibliotheksverbund Bayern im Jahr 2009 – Meilensteine und Perspektiven. In: BFB 4 (2010) 2, S. 99–101) – stellt die Lizenzierung und gemeinsame Weiterentwicklung des Rosetta Digital Preservation Systems der Firma Ex Libris dar. Mit Hilfe einer gezielten Projektförderung des Bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst wurde die Einführung eines landesweiten digitalen Langzeitarchivierungssystems in Angriff genommen. Mit Rosetta wird definitiv der Schritt hin zu einem alle wesentlichen Funktionalitäten des OAIS-Referenzmodells umfassenden Langzeitarchivierungssystem unternehmen. Mit ihm werden weltweit anerkannte Standards der digitalen Langzeitarchivierung konkret in die bibliothekarische Praxis umgesetzt und in den Routinebetrieb übernommen. Das System ist außerdem explizit für einen multi-institutionellen Einsatz ausgelegt.

Ein abteilungsübergreifendes Projektteam der BSB, bestehend aus Mitarbeitern des Referats Münchener Digitalisierungszentrum/Digitale Bibliothek und des Referats Virtuelle Bibliothek Bayern der Verbundzentrale des Bibliotheksverbunds Bayern, ist gerade mit der Pilotierung von Rosetta befasst. Das Projekt wird gemeinsam mit dem bewährten technischen Partner LRZ in der

gewohnten Arbeitsteilung angegangen: Alle die Dateneingabe, das Datenmanagement und die Datenausgabe sowie den Erhalt der Interpretierbarkeit der Daten betreffenden Vorgänge und die Administration des Systems werden von der BSB und der BVB-Verbundzentrale übernommen, während das LRZ die Archivspeicherung der Daten auf großen Plattenspeichern und in robotergesteuerten Magnetbandbibliotheken sowie die Bereitstellung der technischen Infrastruktur übernimmt.

Zunächst wird das System an drei ausgewählten Workflows der BSB getestet, welche wichtige und besonders dringliche digitale Material- und/oder Arbeitsvorgangstypen der Bibliothek aufgreifen:

- Digitalisate: hier liegt das Augenmerk vor allem auf der Skalierbarkeit des Systems,
- Pflichtablieferungen: hier werden vor allem die unterschiedlichen Anforderungen ein- und mehrbändiger sowie fortlaufender Publikationen berücksichtigt,
- Webarchive: hier spielt die Integration spezieller Werkzeuge zur Sammlung und Bereitstellung der einschlägigen Ressourcen eine wesentliche Rolle.

Im Verlaufe des Jahres 2011 werden planmäßig drei weitere bayerische Bibliotheken, die Universitätsbibliotheken Augsburg, Regensburg und Würzburg, als Pilottester in das Projekt einbezogen.

Der anhaltend rasanten Entwicklung gerade im Bereich der Informationstechnik einerseits und der rasch wachsenden Flut von digitalen Publikationen unterschiedlichster Typen andererseits ist nur mit einer steten organisatorisch-technischen Optimierung und Erneuerung zu begegnen. Die Bayerische Staatsbibliothek mit dem Bibliotheksverbund Bayern bleibt bis auf weiteres bei ihrer bisher erfolgreichen Strategie einer schrittweisen Entwicklung beziehungsweise Übernahme und Integration von neuer Technologie in bereits praxisbewährte digitale Archivierungs- und Bereitstellungskonzepte.