

Zugriff für die Ewigkeit

Digitale Langzeitarchivierung an der SLUB

von **HENRIKE BERTHOLD** und **ACHIM BONTE**

Anzahl und Umfang digitaler Dokumente nehmen bekanntlich weltweit rasch zu, als primär digital produzierte Informationen (born digital) wie eBooks und eJournals oder als Digitalisate von analogen Quellen wie Büchern, Fotos, Musik- und Videoaufnahmen. Bibliotheken als Gedächtnisorganisationen reflektieren diesen Wandel, indem sie Digitale Bibliotheken aufbauen und pflegen, die die komfortable Nutzung digitaler Information ermöglichen.

Der Lebenszyklus digitaler Dokumente ist mit der Bereitstellung in einer Digitalen Bibliothek jedoch nicht beendet. Im Gegensatz zu Büchern aus säurefreiem Papier, die eine Haltbarkeit von hundert und mehr Jahren aufweisen, kann ein digitales Dokument bereits nach wenigen Jahren aufgrund eines Medienfehlers oder eines veralteten Datenformats nicht mehr zugänglich sein. Die UNESCO rief deshalb bereits 2003 zur Bewahrung des Digitalen Kulturerbes auf. Die garantierte, langfristige Bewahrung digitaler Dokumente ist das Ziel der digitalen Langzeitarchivierung (LZA). Wer sich an seine persönlichen Erfahrungen mit Diarähmchen, Tonbändern oder VHS-Kassetten erinnert, kann ansatzweise ermessen, welche gewaltigen Herausforderungen das bedeutet.

Moderne Speicherverfahren sichern digitale Dokumente in ihrem aktuellen Format redundant an zwei oder drei Standorten und gewährleisten damit den Erhalt der Korrektheit (Bitstream Preservation). Diese Verfahren können freilich noch nicht die Interpretierbarkeit der Inhalte (Content Preservation) über lange Zeiträume sicherstellen, da wesentliche Informationen über die Dokumente (Metadaten) kein automatisch auswertbarer Bestandteil des Speichersystems sind. Risiken für die digitalen Dokumente können folglich systemseitig nicht erkannt und damit auch nicht behandelt werden.

Die digitale Langzeitarchivierung umfasst deshalb neben technischen Maßnahmen zum Erhalt der Korrektheit auch bibliothekarische und technische Maßnahmen zur nachhaltigen Interpretierbarkeit digitaler Dokumente.

Eine Reihe von internationalen und nationalen For-

schungsprojekten haben Strategien, Richtlinien und Prototypen für die digitale Langzeitarchivierung erforscht und dokumentiert. Seit etwa zwei Jahren werden digitale Langzeitarchive aufgebaut; einige sind inzwischen produktiv, zum Beispiel in der Bayerischen Staatsbibliothek oder der Deutschen Nationalbibliothek und mit ersten Workflows ab Juni nun auch in der SLUB.

Digitale Langzeitarchivierung an der SLUB

Die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) ist die Archivbibliothek des Freistaates Sachsen und der Technischen Universität Dresden, womit ihr auch die Bewahrung digitaler Dokumente obliegt. Das Dresdner Digitalisierungszentrum der SLUB ist mit rund drei Millionen Scans pro Jahr und seiner besonderen Medienvielfalt (Handschriften, Archivalien, Drucke, grafische Medien, Tonträger) eines der führenden deutschen Zentren in öffentlicher Hand. Die digitalen Sammlungen und der Publikationsserver der SLUB umfassen derzeit ein Datenvolumen von rund 200 TB. Die Bestände werden intensiv nachgefragt und repräsentieren einen hohen materiellen wie ideellen Wert.

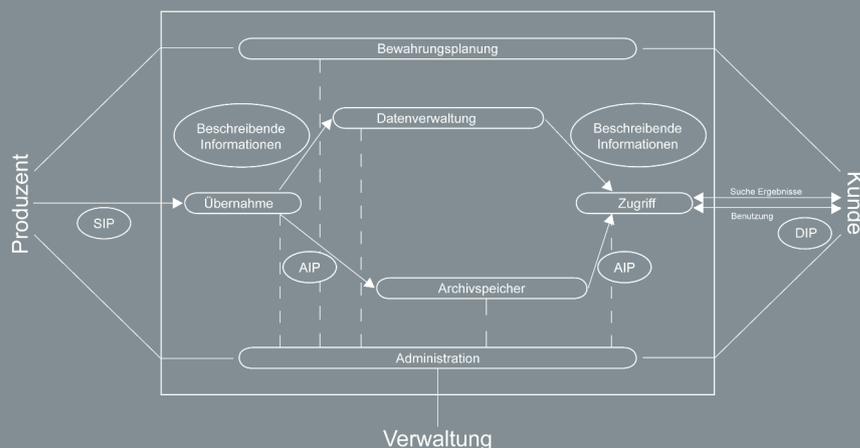
Schon seit mehreren Jahren war die SLUB an der Entwicklung von Konzepten für die Langzeitarchivierung als Partner in Forschungsprojekten wie DP4lib und LuKII aktiv beteiligt. Sie engagierte sich auch in Arbeitsgruppen von nestor, dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung.

Nach gründlicher Prüfung der verfügbaren Systeme begann die Bibliothek im Sommer 2012, ein digitales Langzeitarchiv auf der Basis von Ex Libris Rosetta aufzubauen und in Betrieb zu nehmen. Mit den gewonnenen Erfahrungen und im Wege einer Lizenzerweiterung sollen in einem Folgeprojekt Dienstleistungen für die sächsischen Hochschulbibliotheken und andere Institutionen im Freistaat entwickelt und angeboten werden.

Wie funktioniert digitale Langzeitarchivierung?

Die Funktionalitäten und organisatorischen Rah-

OAIS-DARSTELLUNG DER KERNPROZESSE UND DATENPAKETE



menbedingungen sind von der Internationalen Standardisierungsorganisation ISO beschrieben. Die Abbildung oben zeigt die Kernprozesse der digitalen Langzeitarchivierung nach dem sogenannten Open Archival Information System (OAIS) und die behandelten Datenpakete.

Bei der Übernahme eines digitalen Dokumentes ins Langzeitarchiv (Ingest) wird ein Transferpaket (Submission Information Package, SIP) vom Produzenten entgegengenommen, aufbereitet und geprüft und daraus ein Archivpaket (Archival Information Package, AIP) erzeugt. Das Transferpaket enthält die Dateien, die das eigentliche digitale Dokument bilden, und Metadaten mit inhaltsbeschreibenden und technischen Informationen.

Für die Metadaten gibt es eine Reihe von Beschreibungsstandards. Einige davon sind spezifisch für die Langzeitarchivierung (zum Beispiel PREMIS), andere sind allgemeinerer Natur wie etwa Dublin Core.

Die Prüfung eines Transferpakets umfasst unter anderem das Feststellen der Vollständigkeit aller Metadaten und Dateien, das Überprüfen oder Erzeugen von Prüfsummen, die Kontrolle der Metadaten, das Ermitteln der tatsächlich verwendeten Datenformate sowie die Validierung der Datenformate (das heißt die Prüfung, ob eine Datei die Anforderungen des Datenformates erfüllt). Sofern die Prüfung erfolgreich ist (gegebenenfalls auch erst nach der Korrektur des Pakets), wird ein Archivpaket erzeugt und im Archivspeicher abgelegt.

Der Archivspeicher sichert die Korrektheit der Dateien. Zu seinen Funktionen zählen das Erneuern von Speichermedien, das Prüfen auf Fehler und das Bereitstellen von Möglichkeiten zur Wiederherstellung nach einem Fehlerfall.

Die Erhaltungsplanung überwacht die Umgebung und bestimmt, basierend auf den Metadaten, Risiken für die Archivpakete, zum Beispiel das Veralten eines Datenformats. Die Interpretierbarkeit risikobehafteter Dokumente wird in diesem Fall über Maßnahmen wie die Konvertierung in ein anderes Datenformat oder die Bereitstellung einer Emulationsumgebung erhalten.

Die Zugriffskomponente stellt Funktionen bereit, die es den Kunden erlaubt, die Existenz, Beschreibung und Verfügbarkeit von archivierten Dokumenten zu ermitteln sowie die Dokumente in Form von Ausgabepaketten (Dissemination Information Package, DIP) zu erhalten.

An ein digitales Langzeitarchiv, das digitale Dokumente 50 und mehr Jahre nutzbar halten soll, werden auch besondere Anforderungen an die Vertrauenswürdigkeit gestellt. Die DIN 31644 „Information und Dokumentation – Kriterien für vertrauenswürdige digitale Langzeitarchive“, beschreibt 34 Kriterien. Diese Kriterien betreffen inhaltliche, organisatorische rechtliche und betriebliche Regelungen, wie zum Beispiel:

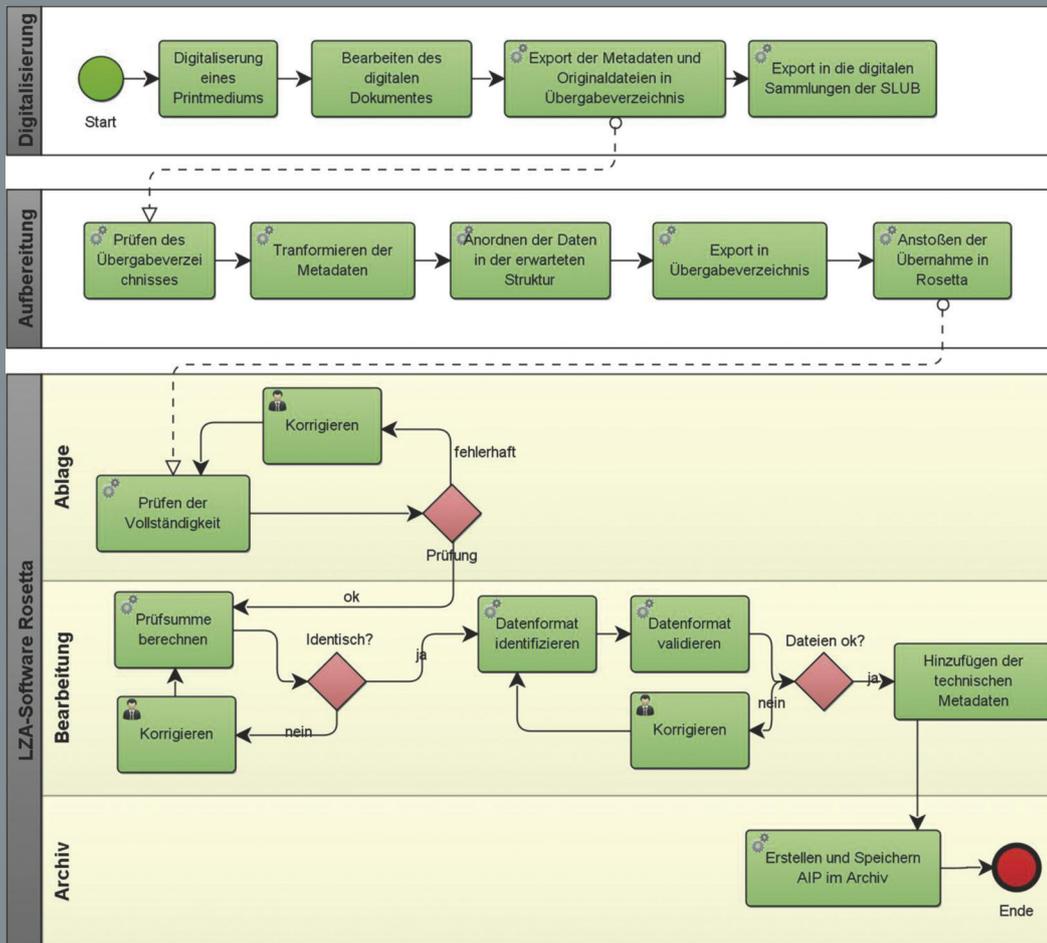
- Auf welcher Grundlage werden Dokumente und Repräsentationen ausgewählt?
- Wer übernimmt die Verantwortung für den Erhalt?
- Für welche Zielgruppen werden die Dokumente bewahrt und welche Anforderungen haben diese?
- Existieren Verträge mit den Produzenten zur Übernahme, Archivierung und Nutzung?
- Ist die Finanzierung und personelle Betreuung langfristig sichergestellt?

Wo steht die SLUB?

Der Betrieb eines LZA-Systems unterliegt ständigen Herausforderungen. Workflows und Bibliothekssysteme ändern sich, neue Medientypen kommen hinzu, neue Technologien werden eingesetzt. Die Entwicklung und der Test neuer Lösungen auf einem Produktivsystem würde für die digitalen Dokumente eine unvermeidbare Gefahr bedeuten. Aus diesem Grund wurden zwei Systeme realisiert – ein Produktivsystem und ein Entwicklungs- und Testsystem, die ähnlich aufgebaut, aber unterschiedlich dimensioniert sind.

Die IT-Infrastruktur für das Produktivsystem wurde gemeinsam mit dem Rechenzentrum der TU Dresden (ZIH) geplant, auf die am ZIH vorhandene Infrastruktur abgestimmt und in diese integriert. Perspektivisch, soll der Betrieb der IT-Infrastruktur durch das ZIH übernommen werden. Das Test- und

LANGZEITARCHIVIERUNGSWORKFLOW



Entwicklungssystem benutzt die vorhandene IT-Infrastruktur der SLUB.

Der Aufbau der LZA findet in zwei Schritten statt. Bis Ende Juni 2013 wurden die SLUB-Produzenten „Retrodigitalisierung von Drucken“ und „digitalproduzierte Dokumente (Qucosa)“ an die digitale Langzeitarchivierung angebunden. Danach begann der Testbetrieb. Im nächsten Schritt, ab Sommer 2013, werden die weiteren Datenproduzenten der SLUB – die „Deutsche Fotothek“ und die „Mediathek“ mit grafischen Medien, Audio- und Videomaterial integriert. Der Workflow für die Retrodigitalisate ist in der Abbildung oben dargestellt.

Neben der Übernahme der Daten ins Langzeitarchiv werden auch Workflows für die Benutzung und Ergänzung der digitalen Dokumente realisiert.

Die Übernahme von Dokumenten aus Qucosa ins Langzeitarchiv entspricht grundsätzlich dem Ablauf bei den Retrodigitalisaten, jedoch sind die Formate für die Daten und Metadaten verschieden.

Wie geht es weiter?

Der Aufbau eines Langzeitarchivs ist sehr aufwändig und teuer. Neben dem bereits erwähnten Versorgungsauftrag im Freistaat Sachsen sieht die SLUB

deshalb vor, mittel- und langfristig Dienstleistungen für alle Bereiche der Langzeitarchivierung aufzubauen und so gerade kleinere Kultur- und Wissenschaftseinrichtungen zu unterstützen, die ihre digitalen Objekte nicht hinreichend werden schützen können. Idealerweise könnte sich die SLUB im Rahmen eines Leistungsverbundes Langzeitverfügbarkeit einbringen und so ein Angebot von passgenauen Dienstpaketen für verschiedene Systeme, Sparten und Fächer erbringen – von der Beratung, über die Validierung digitaler Dokumente oder der Einrichtung einer manuellen Übergabe bis hin zur Implementierung einer automatisierten Archivierungslösung. Ein adäquates Kosten- und Ertragsmodell für diese Dienstleistungen kann die nötige Transparenz und Nachhaltigkeit sichern. Vor diesem Hintergrund führt die SLUB aktuell Gespräche mit vergleichbar leistungsfähigen Partnern und Fördereinrichtungen. Zugriff für die Ewigkeit: Das darf für die reiche und bewahrenswerte digitale Überlieferung in Deutschland keineswegs Utopie bleiben.



HENRIKE
BERTHOLD



ACHIM
BONTE