

Langzeitarchivierung digitaler Bestände in der WLB... mehr als „nur“ ein Backup

Langzeitarchivierung, präziser formuliert „digitale“ Langzeitarchivierung (LZA), ist ein Themenkreis, der nun schon seit mehreren Jahren auf der IT-Agenda der dringend für die WLB zu erledigenden Aufgaben stand.

Die schnell wachsende digitale Bibliothek machte die Lösung der LZA-Frage dringend erforderlich. Die Realisierung der LZA ist jedoch keineswegs trivial und kann nicht einfach „mal so nebenbei“ umgesetzt werden, denn digitale Langzeitarchivierung beinhaltet viele Facetten, die sowohl technischer als auch organisatorischer Art sind. Nach außen sieht die LZA nicht spektakulär aus – sie sollte eigentlich nicht sichtbar sein, da hoffentlich nicht notwendig – und es stellt sich vielleicht die Frage, was LZA so schwierig und aufwendig macht und worin sie sich von einem „normalen“ Backup unterscheidet. Zunächst soll daher in einigen Vorbemerkungen der Themenkreis der digitalen Langzeitarchivierung kurz erläutert werden. Im Anschluss daran wird gezeigt, wie die WLB in Partnerschaft mit der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) und deren Dienstleistungsangebot SLUBArchiv die LZA organisiert und realisiert.

Wir möchten Sie zu einem kleinen Gedankenexperiment einladen und einige einfache Fragen stellen:

- Können Sie noch Ihre ersten digitalen Dokumente öffnen?
- Ist Ihnen bekannt, mit welchen Softwareprodukten und -versionen diese erstellt wurden?
- Haben Sie noch Lesegeräte für Ihre ersten Datenträger?
- Ist diese Software heute noch lauffähig und können Sie diese Software noch installieren?
- Liegen Ihnen noch die Datenträger vor?
- Welche Dokumente bzw. welche Inhalte wären heute noch relevant, sind aber technisch bedingt nicht mehr zugreifbar?
- Machen Sie sich bei neuen Dokumenten Gedanken darüber, ob der Inhalt für die Zukunft bedeutend ist und welche Schlussfolgerungen ziehen sie daraus?

Das Hardwareproblem

Vermutlich konnten Sie nicht alle Fragen positiv beantworten.

Digitale Dokumente sind gefährdete Dokumente und per se nicht für die Ewigkeit gemacht. Jedes technische Speichermedium hat eine Ausfallwahrscheinlichkeit größer Null. Festplatten können „crashen“, Sicherungsbänder verkleben und DVDs können aufgrund von chemischen Prozessen unlesbar werden. Nicht immer sind solche Datenfehler aber offensichtlich. Daten können auch im Laufe der Zeit schleichend verfälscht werden. Einzelne Bits können „umkippen“ („Bitfäule“) und die Daten sind nicht mehr lesbar. Unter Umständen werden solche Fehler zunächst lange nicht erkannt. Leseprobleme treten bei Datenträgern üblicherweise schon nach wenigen Jahren auf. Im Fehlerfall sind die darauf gespeicherten Daten verloren. Mit Maßnahmen der „Bitstream Preservation“ möchte man diese „technischen“ Speicherprobleme beherrschbar machen. Regelmäßiges, mehrfaches Kopieren der Daten an mehreren räumlich getrennten Standorten, die Verwendung von Prüfsummen, der Einsatz von robusten Dateisystemen und die Sicherung auf unterschiedlichen Speichertechniken (Bänder, Plattensysteme, usw.) spielen hier eine wichtige Rolle. Auch die statistische Untersuchung von Stichproben und Testdaten dient dazu, eine Aussage über den technischen Zustand der Daten und Geräte zu erhalten, um dann entsprechend reagieren zu können.

Das Format- und Softwareproblem

Die Gefährdungslage manifestiert sich jedoch nicht nur in den technischen Speichersystemen, sondern vor allem auch durch die Datenformate und die Anwendungen, die mit diesen Formaten umgehen können. Der inhaltliche Zugang zu den Daten ist gefährdet. Die korrekten „Bits and Bytes“ sind nur die halbe Miete. Wenn kein Programm die Dateien mehr öffnen bzw. lesen kann, ist der Zugang zu den Dokumenten nicht mehr möglich und die Inhalte sind trotz korrekter Datenlage verloren oder nur mit sehr großem Aufwand einsehbar.

Maßnahmen, um die Dateiinhalte zugänglich zu halten, werden als „Content Preservation“ bezeichnet. Notwendige Voraussetzung für die Content-Preservation ist, dass die Bitstream Preservation gewährleistet ist.

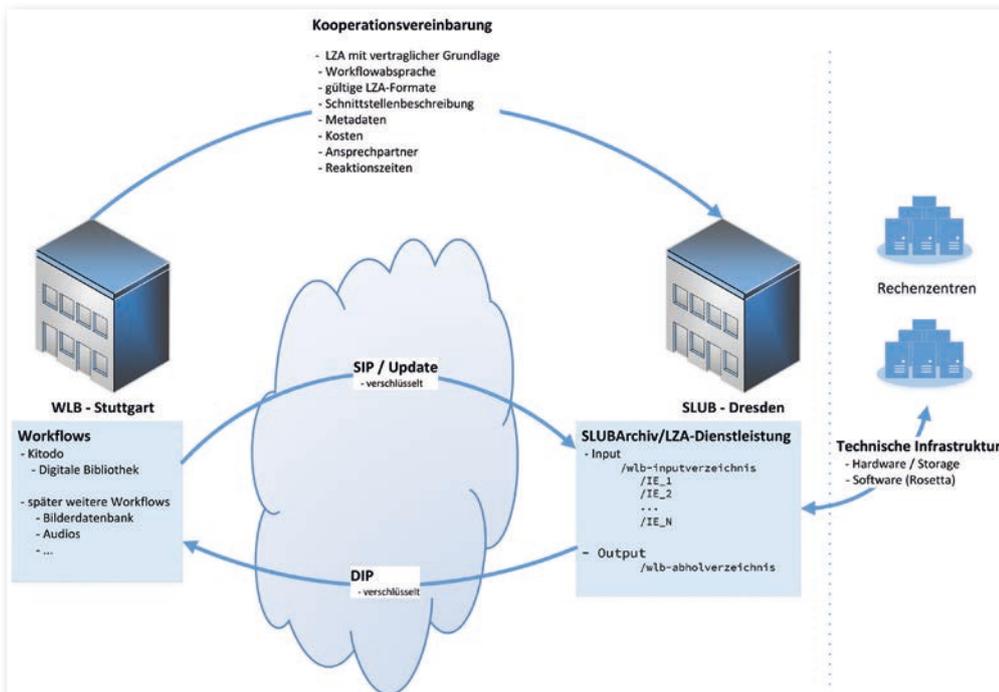
Während „Bitstream Preservation“ als technischer Prozess gut automatisiert werden kann, gestaltet sich die „Content Preservation“ als ungleich aufwendiger. Die Formatproblematik kann nicht einfach automatisiert werden. Formate müssen analysiert, die Darstellungssoftware auf Lauffähigkeit geprüft und der Formatkosmos laufend beobachtet werden. Üblicherweise wird eine Liste langzeitarchivfähiger Formate erstellt. Formate, die beispielsweise proprietär und nicht offen spezifiziert sind, die kaum Verbreitung aufweisen und für die es kaum Software zur Darstellung gibt, taugen nicht

Metadaten müssen daher sorgsam erstellt und vor dem Einbringen in die LZA geprüft werden. Die Aufnahme von digitalen Objekten in das Langzeitarchiv wird als Ingest bezeichnet.

Auch digitale Langzeitarchivierung denkt nicht in ewigen Zeiträumen, sondern zunächst einmal für die nächsten 50 Jahre. Das ist für den Bereich der Informationstechnik ein sehr langer Zeitraum. Damit die Langzeitarchivierung für diesen Zeitraum gewährleistet ist, muss dieser technische Prozess auch eine nachhaltige Organisationsstruktur, gute Geschäftsgänge, einen zuverlässigen Partner und eine vertragliche Grundlage haben.

Langzeitarchivierung mit dem SLUBArchiv

Auf dem Kitodo-Workshop¹ in Mannheim wurde im Mai 2016 die Langzeitarchivierungslösung SLUB-



Überblick: Langzeitarchivierung von digitalen WLB-Beständen im SLUBArchiv

zur LZA. Sind Formate am Ende ihrer Lebensdauer angekommen, so müssen Migrationsstrategien gesucht und entwickelt werden, um Dokumente im alten Format in ein neues Format automatisiert zu überführen. Dabei muss der Inhalt möglichst gut erhalten werden.

Neben der Gewährleistung der technischen Grundlage ist die Metadatenlage von entscheidender Bedeutung. Ohne administrative, technische und inhaltliche Metadaten sind die archivierten digitalen Objekte (Intellektuelle Einheiten = IE) wertlos.

Viele Gespräche und eine rege E-Mail-Kommunikation haben dann zu einer LZA-Kooperationsvereinbarung mit der SLUB geführt. Neben den ähnlichen Workflows ist vor allem auch die Tatsache ausschlaggebend, dass mit dem SLUBArchiv bereits jetzt ein LZA-System im umfassenden Sinne produktiv zur Verfügung steht und ein vertrauenswürdiges Archiv betrieben werden kann.

Die oben genannte Vereinbarung wurde zum 01.07.2017 geschlossen und auf beiden Seiten die LZA-Verantwortlichen benannt.

1) Kitodo: Key to Digital Objects / Kitodo ist der neue Name des bisherigen Goobi e.V. und bezeichnet eine Workflowsoftware für Digitalisierungsgeschäftsgänge.

Vorbereitend wurden Testdaten ausgetauscht und die Digitalisate auf Langzeitarchivierungstauglichkeit analysiert und validiert. Es zeigte sich, dass an einigen Stellen Korrekturen in Attributen und bei den Farbprofilen vorgenommen werden mussten. Die Anforderungen an die LZA werden zukünftig im Digitalisierungsprozess berücksichtigt.

Zunächst sollen die Digitalisate der Landtagsprotokolle und der Sammlung Nicolai in die LZA überführt werden. Diese digitalen Bestände liegen bereits in einem guten, stabilen Zustand vor, so dass aufwendige LZA-Updates kaum vorkommen werden.

In der ersten Phase wird die WLB nur einen Workflow mit TIFF- und METS/MODS-Daten realisieren. Mittelfristig könnten durchaus auch weitere Workflows hinzukommen (Bilddatenbank, Audiodatei, usw.).

Die Vorarbeiten der WLB sind abgeschlossen. Starten soll die LZA mit der Rosetta-Version 5.2, die im Sommer 2017 im SLUBArchiv eingeführt wurde. Für diese neue Version wurde für die WLB ein Setup eingerichtet. Mit der Freigabe der Konfiguration wird die WLB beginnen, Daten in das Langzeitarchiv einzubringen. Dies erfolgt voraussichtlich ab dem 3. Quartal 2017.

Für jede „Intellektuelle Einheit IE“ wird dazu ein SIP-Paket (= Ablieferungspaket) erstellt, das an einen wohldefinierten Ort über eine verschlüsselte Verbindung in einen SLUBArchiv- Austauschordner hochgeladen wird. Dort wird das Paket geprüft, als Archivierungspaket (AIP) in die LZA überführt und es geht in den Verantwortungsbereich des SLUBArchivs über. Von den Archivpaketen werden jeweils drei Kopien erstellt und an zwei räumlich getrennten Standorten abgelegt.

Um eine IE wieder auslesen zu können, stellt das SLUBArchiv einen Webservice zur Verfügung. An diesen kann von der WLB ein Request gerichtet werden, den der Service mit der Auslieferung der DIPs (Auslieferungspakete) beantwortet.

Aufwand und Kosten – für konventionelle und digitale Medien

Die Ressourcen, die wir für die Verwaltung der klassischen Papiermedien aufwenden, der Stellplatz und die Bücherregale, die wir bereitstellen, die Betreuung der Magazine, die Maßnahmen der Bestandserhaltung – all dies müssen wir auch äquivalent zusätzlich für die digitalen Objekte einplanen. Das „digitale Magazin“ braucht die gleiche Beachtung und Sorgfalt wie der klassische konventionelle Bestand. Die umfangreichen Dienstleistungen, die der Betrieb einer guten LZA erfordert, gibt es nicht zum Nulltarif. Eine wachsende digitale Bibliothek wird daher auch ansteigende zusätzliche Kosten verursachen und weitere Ressourcen an Personal binden.

Wie bereits erwähnt: Äußerlich betrachtet ist die LZA unspektakulär. Auf den zweiten Blick eröffnen sich umfangreiche, unterschiedlichste und anspruchsvolle Aufgabengebiete. Dies verlangt auch ein stetiges am Ball bleiben und eine entsprechende Fortbildung.

Weitere vertiefende Informationen können auch auf der Homepage des SLUB-Archivs² nachgelesen werden.

Langzeitarchivierung in Baden-Württemberg

Auch in vielen anderen Einrichtungen Baden-Württembergs geht man die Thematik der Langzeitarchivierung an. In einem zweijährigen Projekt BW-Data-Bib sollen bis 2019 verschiedene Lösungskonzepte für die digitale Langzeitarchivierung evaluiert und anschließend in einem zweiten Projekt nach Möglichkeit eine Lösung für die bibliothekarischen Einrichtungen des Landes Baden-Württemberg realisiert werden. Mit einer produktiven Nutzung ist daher frühestens ab dem Jahr 2021 zu rechnen. Der WLB war es jedoch wichtig, bereits jetzt die relevanten digitalen Daten in die LZA zu überführen. Mittelfristig bedeutet diese Entwicklung allerdings, dass es zwei interessante Dienstleister für die LZA gibt.

Langzeitarchivierung bedeutet viel Arbeit – aber es ist ein beruhigendes Gefühl, unsere Daten im Langzeitarchiv zu wissen.

Stephan Abele / Jörg Oberfell