

Entwicklung in den Bereichen Records Management / Vorarchiv – Übernahme – Langzeitarchivierung



Dreizehnte Tagung des Arbeitskreises
"Archivierung von Unterlagen aus digitalen Systemen"
vom 27./28. April 2009
ausgerichtet vom Staatsarchiv St.Gallen

Veröffentlichungen des Staatsarchivs St.Gallen

Staatsarchiv St.Gallen (Hrsg.)

Entwicklung in den Bereichen Records Management / Vorarchiv – Übernahme – Langzeitarchivierung

Dreizehnte Tagung des Arbeitskreises
"Archivierung von Unterlagen aus digitalen Systemen"
vom 27./28. April 2009
ausgerichtet vom Staatsarchiv St.Gallen

Veröffentlichungen des Staatsarchivs St.Gallen

Impressum:

Titelbild: © 2009 Staatsarchiv St.Gallen

Herausgeber und Copyright: © 2009 Staatsarchiv St.Gallen

Redaktion: Martin Lüthi, Felix Stadler

ISBN 978-3-908048-51-0

Inhaltsverzeichnis

Vorwort.....	5
--------------	---

Records Management / Vorarchiv

Normen und Best Practices im Records Management. Arbeit und Ziele des Arbeitskreises <i>Schriftgutverwaltung (Records Management)</i> im Deutschen Institut für Normung e.V.	7
Die Einführung des DMS/VBS EL.DOK (Prodea) in der Ministerialverwaltung Brandenburg und die Konzipierung eines digitalen Zwischenarchivs unter Mitwirkung des Brandenburgischen Landeshauptarchivs (BLHA)	11

Übernahme

Das "Digitale Archiv " des Bundesarchivs.....	18
Normung und Best Practices im Records Management. Mehrwert der ISO-15489.....	27
Das Digitale Archiv des Bundesarchivs: Fachliche Anforderungen für die Übernahme elektronischer Akten aus Vorgangsbearbeitungssystemen.....	31
Ingest von Fachverfahren im Landesarchiv Baden-Württemberg. Anmerkungen zu Authentizität, Prozessen und Softwareentwicklung.....	35
Von ELO nach FAUST - Konzeptuelle Überlegungen zur Langzeitarchivierung digitaler Überlieferungen und Anwendung im Archiv der sozialen Demokratie der Friedrich-Ebert- Stiftung	46
Wege ins Archiv - Der nestor-Leitfaden für die Informationsübernahme	50

Langzeitarchivierung

Handreichungen zur Webarchivierung des Arbeitskreises Dokumentation und Archivierung von Webpräsenzen	57
Planung, Ausbau und Betrieb des brandenburgischen revisionssicheren digitalen Langzeitarchivs nach OAIS	67
EL.NORM. Die geplante elektronische Normverkündung im Land Brandenburg (Stand: Mai 2009)	74
Das Archivierungssystem AMphora der Aktion Mensch	78
PADUA: Produktivsystem für die Archivierung digitaler Unterlagen im (Stadt)-Archiv Stuttgart	89
<i>arcun</i> . Überlegungen zur digitalen Langzeitarchivierung als Dienstleistungsangebot	98
Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (GEVER): Projekt AUGev und Pilotprojekt LARIS I	109
Abkürzungsverzeichnis	125

Vorwort

Die 13. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" fand vom 27. und 28. April 2009 statt. Sie wurde vom Staatsarchiv des Kantons St.Gallen ausgerichtet. Damit wurde das jährliche Fachtreffen erstmals nicht in Deutschland durchgeführt. Dieser Umstand verdeutlicht sicher auch den Stellenwert, den die Tagung unter den Archiven des deutschen Sprachraums genießt.

Die Veranstaltung widmete sich den vielfältigen Themen aus den Bereichen Records Management / Vorarchiv – Übernahme – Langzeitarchivierung. Wiederum ergab sich ein intensiver, lehrreicher und vor allem praxisnaher Austausch von Erfahrungen aus Archiven und Behörden.

Die auf den ersten Seiten der vorliegenden Publikation veröffentlichten Vorträge verdeutlichen die Wichtigkeit von Normen und Standards. Die Einbindung der Archive bereits bei der Schriftgutverwaltung / Aktenführung - respektive ihre Mitwirkung bei der Einführung von DMS¹ oder ähnlichen Systemen - ist dabei ein entscheidendes Element. Eine frühe Mitwirkung, bestenfalls sogar in Form einer Mitbestimmung, hat sicherlich einen lenkenden Einfluss und kann die künftige Archivierung erleichtern.

Bei der Thematik Übernahme wurde den Teilnehmenden vor Augen geführt, wie gross die Vielfaltigkeit der elektronischen Unterlagen ist. Das gilt nicht nur im Bezug auf die Dateiformate, sondern auch mit Blick auf die Strukturierung. Dabei richtet sich der Fokus bei der Übernahme auf einfache und klare Schnittstellen.

Vermehrt werden auch Lösungen und Hilfstools im Bereich der digitalen Langzeitarchivierung, welche in Archiven selber entwickelt wurden, der archivischen Welt frei zur Verfügung gestellt.

Eine bedeutende Rolle fällt immer mehr der Webarchivierung zu, weil grundlegende Informationen über Dienststellen und Ämter oft nur noch im Intranet bzw. Internet aktuell gehalten werden.

Auch bei den Beiträgen zur Tagung von 2009 werden Lösungen zur Langzeitarchivierung - aus kleinen Einrichtungen bis zu ganz grossen Archiven – präsentiert. Es ist erfreulich, dass sich Archive mit unterschiedlichem Hintergrund dem Thema der elektronischen Langzeitarchivierung widmen. Dabei geht es von ersten Ansätzen im Pilotbetrieb bis zu Lösungen im produktiven Betrieb.

¹ Dokumenten-Management Systemen

Allen Referentinnen und Referenten möchte ich für Ihre Beiträge herzlich danken.

Martin Lüthi

Normen und Best Practices im Records Management. Arbeit und Ziele des Arbeitskreises *Schriftgutverwaltung (Records Management)* im Deutschen Institut für Normung e.V.

Irmgard Mummenthey, Alexandra Lutz, Joachim Kemper²

Dieser Beitrag basiert auf einem kurzen Vortrag, der anlässlich der 13. Tagung des *Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen* in St.Gallen gehalten wurde. Er konzentriert sich auf die Vorstellung des Arbeitskreises im Deutschen Institut für Normung e.V. (DIN) und die Nutzen der zentralen Norm zum Records Management, der DIN ISO-15489, vor allem aus der Sicht der Archive. Der Beitrag von Steffen Schwalm, ebenfalls Mitarbeitender im Arbeitskreis, widmet sich dann der Sicht des Organisationsberaters von aussen.

Im Jahr 1998 hatte das Technical Committee (TC 46) der ISO die Arbeit an der ISO-Norm 15489 *Information and documentation - Records Management* aufgenommen. Innerhalb des DIN gründete man daraufhin den Arbeitsausschuss *Archiv- und Schriftgutverwaltung*, der dem Normenausschuss 15 *Bibliotheks- und Dokumentationswesen* (DIN NABD 15) zugeordnet wurde. In den folgenden Jahren beteiligte sich dieses sogenannte Spiegelgremium intensiv an der Erstellung der Norm, die 2001 fertig gestellt und 2002 in das deutsche Normenwerk übernommen wurde – als DIN ISO-15489 *Information und Dokumentation – Schriftgutverwaltung*.³

Angesichts der zunehmenden Bedeutung digitaler Unterlagen wurde der DIN NABD 15 im Sommer 2008 umstrukturiert und trägt nunmehr die Bezeichnung Arbeitsausschuss *Schriftgutverwaltung und Langzeitverfügbarkeit digitaler Informationsobjekte*. Innerhalb des Arbeitsausschusses gibt es nun mehrere Arbeitskreise. Dazu gehört der Arbeitskreis *Schriftgutverwaltung*, der mittlerweile auch den Begriff "Records Management" im Titel führt, da der Begriff "Schriftgut" bzw. "Schriftgutverwaltung" möglicherweise nicht mehr ausreicht, um die Verwaltungsrealität abzubilden.⁴ Zu den weiteren Arbeitskreisen zählen beispielsweise diejenigen für *Vertrauenswürdige digitale Archive* und *Rechtssichere Aufbewahrung von digital signierten Dokumenten*.

² Dr. Alexandra Lutz (Archivschule Marburg) ist Sprecherin des *Arbeitsausschusses Schriftgutverwaltung (Records Management)*, Dr. Joachim Kemper (Generaldirektion der staatlichen Archive Bayerns) und Irmgard Mummenthey (Staatsarchiv Hamburg) sind Mitarbeitende.

³ Vgl. zur Vorgeschichte und zur Beteiligung deutscher Archivarinnen und Archivare Nils Brübach, *Records oder Akten – Internationale Normung für die Schriftgutverwaltung*, in: Andreas Metzing (Hrsg.), *Digitale Archive – Ein neues Paradigma?* Marburg 2000, S. 221-237 sowie Michael Wettengel, *Normierung der Schriftgutverwaltung - Zu einer neuen ISO-Norm aus staatlicher Sicht*, *Der Archivar* 4 (2003), S. 307-310.

⁴ Vgl. auch Irmgard Mummenthey, *Schriftgutverwaltung als strategische Aufgabe – Selbstverständnis und Politik des Staatsarchivs der Freien und Hansestadt Hamburg*, in: Alexandra Lutz (Hrsg.), *Zwischen analog und digital. Schriftgutverwaltung als Herausforderung für Archive*, Marburg 2009 (im Erscheinen) und zum Verständnis von Records und Akten Nils Brübach, wie Fussnote. 2, S. 226ff.

Den grundlegenden Nutzen einer Norm für die Schriftgutverwaltung hat u.a. Michael Wettengel formuliert:

- Betonung der Gemeinsamkeiten in der Schriftgutverwaltung
- Stärkung des Qualitätsbewusstseins in der Schriftgutverwaltung
- Grundlage für die Qualifizierung des Personals.⁵ Die Wechselwirkungen zwischen der Qualität der Schriftgutverwaltung bei den Registraturbildnern und der Überlieferungsbildung springen jeder Archivarin und jedem Archivar ins Auge, weshalb sie ein unmittelbares Interesse daran haben müssten, die Norm in ihrem Zuständigkeitsbereich bekannt werden zu lassen und, sofern sie ihren Träger bei seiner Schriftgutverwaltung beraten, als Grundlage zu nutzen. Auf der anderen Seite hat eine solche Beratung möglicherweise mehr Durchsetzungskraft, wenn sie sich auf eine Norm stützen kann.

Die DIN ISO-15489 gilt für die analoge und digitale Schriftgutverwaltung bei öffentlichen und privaten Organisationen, sie gilt ausdrücklich nicht für die Verwaltung von Archivgut.⁶ Sie definiert eingangs den Nutzen von Schriftgutverwaltung aus der Sicht der betroffenen Organisation und bietet Argumente, die auch beratenden Archiven in der Diskussion mit den Trägern von Nutzen sein können. An erster Stelle steht dabei die Gewährleistung der Rechenschaftsfähigkeit, weshalb die Norm dann auch folgerichtig die Schriftgutverwaltung als Führungsaufgabe beschreibt.⁷ Konkrete Anforderungen liefert die Norm, soweit es um die Merkmale von Schriftgut und Schriftgutverwaltungssystemen geht. Erfüllt ein Träger diese Anforderungen, kann er sicher sein, sich im Rahmen derjenigen Vorgaben zu bewegen, die für ihn gelten – für die öffentliche Verwaltung wäre das beispielsweise das aus Artikel 20 Absatz 3 Grundgesetz (Bindung der Verwaltung an Recht und Gesetz) hergeleitete Gebot, Verwaltungshandeln jederzeit nachvollziehbar und transparent zu gestalten.⁸ Weiterhin definiert die Norm die Prozesse der Schriftgutverwaltung. Wer schlagkräftige Argumente für die Existenzberechtigung von Aktenplänen und Aufbewahrungsfristen sucht, findet sie in diesem Abschnitt.

Neben dem eigentlichen Normtext im Teil 1 (DIN ISO-15489-1) bietet der so genannte Fachbericht im Teil 2⁹ eine Beschreibung der Arbeitsschritte bei der Implementierung von Schriftgutverwaltungssystemen und damit eine Folie für die Analyse und Planung derartiger Prozesse.¹⁰

⁵ Michael Wettengel, wie Fussnote. 2, S. 309.

⁶ Vgl. DIN ISO 15489-1:2002-12, Abschnitt 1.

⁷ Ebd., Abschnitte 3, 4 und 6.3.

⁸ Vgl. dazu u.a. den Beschluss des Bundesverfassungsgerichts vom 6.6.1983 - 2 BvR 244, 310/83 (Entfernung und Vernichtung von Vermerken aus der Ausländerakte), in: Neue Juristische Wochenschrift 1983, S. 2135.

⁹ DIN-Fachbericht ISO TR 15489-2 (2004).

¹⁰ Vgl. auch die Vorgehensweise bei Thomas Brakmann, Schriftgutverwaltung im Umbruch: Domea-Pilotierung in einer Mittelbehörde, in: Alexandra Lutz (Hrsg.), Zwischen analog und digital. Schriftgutverwaltung als Herausforderung für Archive, Marburg 2009 (im Erscheinen).

Beide Teile der Norm sind an manchen Stellen eher allgemein gehalten. Das ergibt sich aus der Tatsache, dass sie für unterschiedlichste Träger Gültigkeit behalten müssen. Deshalb arbeitet der Arbeitskreis *Schriftgutverwaltung (Records Management)* zurzeit an einem Handbuch, in dem die verschiedenen Abschnitte der Norm und auch des Fachberichts sehr praxisbezogen aufbereitet werden. Im Vordergrund steht dabei zunächst, bedingt durch die aktuelle Zusammensetzung des Arbeitskreises, die öffentliche Verwaltung. Weitere Bereiche sollen folgen, wobei dann die Mitwirkung von entsprechenden Expertinnen und Experten – beispielsweise von Records Managern in Unternehmen – sinnvoll wäre.¹¹

Weiterhin beabsichtigt der Arbeitskreis den Aufbau eines Online-Angebotes, welches neben Häufig gestellte Fragen (FAQ) auch eine Zusammenstellung von Normen, Standards und Arbeitskreisen rund um das Thema Schriftgutverwaltung bzw. Records Management beinhalten soll. Handbuch und Online-Angebot richten sich sowohl an Einrichtungen, die ihre Schriftgutverwaltung optimieren möchten, als auch an die Archive, die sie dabei beraten. Letztlich ist es Aufgabe der Träger und der sie beratenden Archive, die Norm für ihren jeweiligen Bereich "herunterzubrechen". Der Arbeitskreis will mit den erwähnten Vorhaben Unterstützung bieten dabei unterstützen und auch dafür sorgen, dass Doppelarbeit vermieden wird.

Der Einsatz der ISO-15489 hat sich aus Sicht des Staatsarchivs Hamburg bewährt. Die Norm konnte in den letzten Jahren einerseits als strategisches Instrument, andererseits als Grundlage für die Erarbeitung von Schulungsunterlagen genutzt werden. In Hamburg wurde mittels einer Umfrage die Notwendigkeit festgestellt, die Schriftgutverwaltung in den Behörden und Ämtern zu optimieren.¹² Prozesse, an denen Menschen beteiligt sind, können nur dann optimiert werden, wenn die beteiligten Menschen gestärkt werden. Die Norm hebt daher neben den bereits erwähnten Führungskräften auch weitere Verantwortliche hervor, darunter selbstverständlich auch die Mitarbeitenden in Registraturen, und betont die Notwendigkeit der ständigen Aus- und Fortbildung.¹³ Deshalb hat das Staatsarchiv die Angebote für verschiedene Zielgruppen in der zentralen Fortbildung ausgebaut, gemeinsam mit den an der Ausbildung der Nachwuchskräfte beteiligten Stellen die Lehrpläne überarbeitet und Multiplikatoren geschult. Für die Inhalte der Aus- und Fortbildung wiederum dient die DIN ISO-15489 als Orientierung.

Auf der internationalen Ebene befindet sich zurzeit eine Familie von verschiedenen Normen zum Records Management in der Entstehung. Sie setzt sich aus schon bestehenden Normen wie der ISO-23081 (Metadata for Records) und ISO-26122 (Work Process Analysis for Records) und noch im Entwurf befindlichen Normen (z.B. Risk Assessment for Records) zusammen. Die ISO-15489 wird überarbeitet und geht möglicherweise in einer neuen Norm auf. Der

¹¹ Interessierte aus allen Bereichen wenden sich gern an Frau Dr. Alexandra Lutz (Archivschule Marburg).

¹² Vgl. Julia Brüdegam, Hendrik Eder und Irmgard Mummenthey: Die gefühlte Misere in greifbaren Zahlen. Schriftgutverwaltung in der Freien und Hansestadt Hamburg. Ausgangssituation – Standards – Perspektiven, in: Der Archivar 1 (2007), S. 29-33.

¹³ Vgl. DIN ISO 15489-1:2002-12, 6.3 und 11.

Arbeitskreis ist hier wieder in der Rolle eines Spiegelgremiums zum TC 46 beim ISO und beteiligt sich intensiv an der Diskussion. Die Schaffung dieser Normenfamilie ist sehr zu begrüßen, da sie den Stellenwert des Records Management bzw. der Schriftgutverwaltung ein weiteres Mal steigern könnte – dies aber nur, wenn alle international beteiligten Ausschüsse in den Normentexten den Balanceakt zwischen Allgemeingültigkeitsanspruch und Praxisnähe schaffen.

Die Einführung des DMS/VBS EL.DOK (Prodea) in der Ministerialverwaltung Brandenburg und die Konzipierung eines digitalen Zwischenarchivs unter Mitwirkung des Brandenburgischen Landeshauptarchivs (BLHA)

Ilka Stahlberg

Rahmenbedingungen

Anknüpfend an meinen Vortrag "Ausschreibung/Fachkonzept Landes-DMS/VBS – Von der Durchsetzung archivischer Interessen bis zur Implementierung einer Aussonderungskomponente" auf der 12. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen in Koblenz 2008"¹⁴ möchte ich über meine Erfahrung bei der Begleitung der DMS/VBS-Einführung aus archivischer Sicht berichten und zugleich neue Konzeptionen vorstellen.

"Türöffner" für die enge Projekteinbindung des BLHA waren einerseits das Brandenburgische Archivgesetz¹⁵ und andererseits die Gemeinsame Geschäftsordnung für die Ministerien des Landes Brandenburg (GGO) mit ihrer Anlage 1: Rahmenrichtlinie für das Bearbeiten und Verwalten von Schriftgut in der Landesverwaltung Brandenburg (Registraturrichtlinie – RegR)¹⁶. Während das Archivgesetz dem Archiv die Aufgabe überträgt, die Behörden bei der Schriftgutverwaltung zu beraten und die Möglichkeit gibt, ein Zwischenarchiv zu unterhalten, wird in der GGO das Motiv für die Schriftgutverwaltung definiert: der Verfassungsgrundsatz der Nachvollziehbarkeit des Verwaltungshandelns. Die Registraturrichtlinie (RegR) bindet die obersten brandenburgischen Landesbehörden auf einen Standard für die Erstellung, Verwaltung und Aussonderung des Schriftgutes in Papier und elektronischer Form. Motivation für das Engagement des BLHA in dieser Angelegenheit ist das Ziel am Ende der Verwaltungstätigkeit Schriftgut in archivierbarer Form zu erhalten.

Wie bereits in Koblenz berichtet, ist das Fachkonzept zur DMS/VBS-Einführung auf die Umsetzung der Regelungen der GGO und RegR gerichtet.

In der RegR ist z.B. nach "§ 7 Abs. 4 geregelt, dass bei elektronischer Vorgangsbearbeitung sicherzustellen ist, dass die Dokumente, der Laufweg und die Aufzeichnungen aus der Bearbeitung (z.B. Geschäftsablaufmerke, Verfügungen, Aktenvermerke, Zeichnungen, Mitzeichnungen, Kenntnisaufnahmen) in Protokoll- und Bearbeitungsinformationen nachgewiesen und der elektronischen Akte zugeordnet werden. Nach Abs. 5 sind bei elektronischer Aktenführung aktenrelevante Dokumente 'schreibgeschützt' zu speichern. Unter schreibgeschützt wird verstanden, dass ein elektronisches Dokument nicht mit geringerer Energie und geringerem Aufwand zu verändern wäre, als dass für die Verfälschung eines Papierdokumentes erforderlich wäre (Wahrung des Sicherheitsstandards)."

¹⁴ <http://www.bundesarchiv.de/aktuelles/fachinformation/00080/index.html>.

¹⁵ Gesetz über die Sicherung und Nutzung von öffentlichem Archivgut im Land Brandenburg vom 7. April 1994 (BbgArchivG), in: GVBl I 1994 S. 94-100.

¹⁶ http://www.landesrecht.brandenburg.de/sixcms/detail.php?gsid=land_bb_bravors_01.c.23725.de.

Im Ergebnis der zitierten Regelungen ist "elektronische Aktenführung" in Brandenburg nur mit Nutzung eines Domea-konformen Dokumentenmanagementsystems möglich.

In § 11 Abs. 2 der RegR wird darauf verwiesen, "dass elektronisch gespeichertes Schriftgut der laufenden Pflege bedarf und jeweils rechtzeitig ohne inhaltliche Veränderung auf Formate und Datenträger übertragen werden muss, die dem aktuellen Stand der Technik entsprechen." Diese Regelung sensibilisiert nach unserer Erfahrung bei der Festlegung von Aufbewahrungsfristen. Zur Aussonderung ist nach §14 festgeschrieben, "dass nach Abs.1 nach Ablauf der Aufbewahrungsfrist dem BLHA elektronisch gespeichertes Schriftgut gemäss der Anlage 7 anzubieten und sofern es als archivwürdig bewertet wurde, vollständig zu übergeben ist. Über die Form der Abgabe entscheidet das BLHA im Benehmen mit der zuständigen Stelle. Nach Abs. 2 ist bei der Übergabe elektronisch signierter Unterlagen von der abgebenden Stelle zu bestätigen, dass diese nicht nachträglich verändert wurden und die elektronischen Signaturen zum Zeitpunkt der Übergabe gültig waren. Und nach Abs. 3 ist zur Gewährleistung einer ordnungsgemässen Aussonderung bei Einführung von Systemen zur elektronischen Schriftgutverwaltung und Vorgangsbearbeitung in Abstimmung mit dem BLHA eine Schnittstelle vorzusehen und sicherzustellen, dass die Schliessung von elektronischen Akten und die Bildung von Bänden möglich ist."

Diese Regelungen der RegR bilden die Grundlage dafür, dass das BLHA in die derzeitige konzeptionelle Vorbereitung eines digitalen Langzeitspeichers/Zwischenarchivs eingebunden ist und die archivischen Anforderungen geltend machen kann. Die Grundlage für die archivischen Anforderungen bildet das Erweiterungsmodul 5 zum Domea Organisationskonzept 2.1.: "Aussonderung und Archivierung elektronischer Akten"¹⁷. Im Hinblick auf den zitierten § 14 der RegR ist hier v.a. auf die Normensetzung des BLHA im Bereich der Archivierungsformate und auf die Pflicht zur Entwicklung der Schnittstellen von brandenburgischen IT-Projekten zum BLHA aufmerksam zu machen.

EL.DOK-Einführung

Die Einführung des landeseinheitlichen DMS/VBS EL.DOK (auf der Basis des Produktes Prodea von SER) erfolgt auf der Grundlage des Kabinettsbeschlusses 734/08 und sieht die Einführung von EL.DOK in allen Ministerien, dem zentralen IT-Dienstleister ZITBB und in Teilen des BLHA bis Ende 2012 vor. Unter Projektleitung des Innenministeriums erfolgt der zentrale Betrieb durch den ZITBB. Das Projekt umfasst auch die Einführung eines Kabinetttinformationssystems (EL.KIS) unter Projektleitung der Staatskanzlei. Die Kommunikation zum DMS/VBS des Landtags (Projekt ELVIS) erfolgt über eine erweiterte XDomea-Schnittstelle. Das Projekt EL.DOK umfasst ebenso die Schnittstelle zum elektronischen Archiv des BLHA (Projekt EL.Archiv).

¹⁷ http://www.verwaltung-innovativ.de/cln_047/nn_684678/DE/Organisation/domea__konzept/domea__konzept__node.html?__nnn=true Schriftenreihe der KBST Bd. 66, 2004.

Der Projektauftrag vom 29.5.2008 beauftragte das Innenministerium mit der Entwicklung eines Landesbasisplatzes einschliesslich der Schnittstellen zu Landtag und BLHA, dem Aufbau eines Kompetenzzentrums und der Begleitung der Einführung des landeseinheitlichen DMS. Diese Vorhaben sind zentral finanziert und werden durch den Lenkungsausschuss begleitet. Häuser-spezifische Anpassungen und Einbindung von Fachverfahren werden im Rahmen des Customizing durch die Ressorts finanziert. Ein Ausrollen auf den nachgeordneten Ressortbereich soll anschliessend erfolgen, wobei die Finanzierung hier Gegenstand späterer Haushaltsverhandlungen (Landeslizenz) sein wird.

Nach der Erteilung des Auftrages an SER wurde durch das Innenministerium mit der Projektumsetzung begonnen.

Das Gesamtprojekt EL.DOK konstituierte sich im Herbst 2008. Ihm obliegt die Koordinierung der EL.DOK-Einführung in allen obersten Landesbehörden. Schwerpunktmässig wird hier das im Rahmen des EL.DOK-Projektes aufgelegte umfangreiche Akzeptanzmanagementprogramm durchgeführt, welches einen regelmässigen Erfahrungsaustausch zu den in den Ministerien laufenden Vorbereitungsarbeiten an den häuserspezifischen Anpassungskonzepten ermöglicht. Diese umfassen u.a. Scan-, Sicherungs- und Rechtekonzepte, die Aktenplanrevision und Formularanpassungen. Im Rahmen des Gesamtprojektes werden regelmässige Statusberichte zum Vorbereitungsstand und zu Problemen der einzelnen Ressorts gegeben, es erfolgt die Erarbeitung eines Schulungskonzeptes, der Aufbau des Kompetenzzentrums und der Aufbau der Infrastruktur durch den ZITBB.

Das BLHA wirkt seit Herbst 2008 auch im Teilprojekt DMS-Basis mit, das die korrekte Umsetzung des Fachkonzeptes, der GGO und RegR im Landesbasisplatz einrichtet und abstimmt. Begleitend fanden für alle Ressorts Schulungen der Teilprojektteams und Fachadministratoren statt, die seit Januar die jeweiligen Projektstände an eigenen Testmandanten testen und ihre Erfahrungen und Anforderungen in das Teilprojekt DMS-Basis einbringen. Vorgesehen ist im Mai 2009 ein Praxistest im Innenministerium, an den sich zeitnah Performancetests, das Verfahren der Freigabe und schliesslich die voraussichtliche Einführung im 2. Halbjahr in der ersten Abteilung des Innenministeriums anschliessen sollen.

Das Teil-Projekt EL.KIS unter Federführung der Staatskanzlei ist ein behördenübergreifendes Vorgangsbearbeitungssystem für alle Kabinettreferate, welches für das Dokumentenmanagement auf dem Teilprojekt DMS-Basis aufsetzt. Geplant ist noch das Teilprojekt EL.ZA (Elektronische Zusammenarbeit), mit dem eine organisationsübergreifende Projektarbeit unterstützt werden soll. Die Arbeiten an diesem Projekt werden erst nach Abschluss des Teilprojektes DMS-Basis beginnen.

Motor des Gesamtprojektes ist der Auftrag des Landtages, der ab Herbst 2009 ausschliesslich elektronisch mit der Landesverwaltung kommunizieren will. Im Augenblick ist noch nicht absehbar, ob sich der Projektzeitplan ohne grosse Verschiebungen umsetzen lässt, es ist aber hervorzuheben, dass bedingt durch die Projektstruktur alle Ressorts in den Prozess der Vorbereitung und Abstimmung eingebunden sind, um künftig Synergien bei der gemeinsamen Vorgangsbearbeitung nutzen zu können. Die gemeinsame Vorgehensweise ermöglicht bereits in

der Vorbereitungs- und Anpassungsphase Synergien durch die Nachnutzung von Testszenarien, Konfigurationen, Formularen usw.

Schriftgutverwaltung

In der derzeitigen Entwicklungsphase stehen im Zusammenhang mit der Erarbeitung der Anpassungskonzepte in allen Ministerien, aber auch bei der Konfiguration des Landesbasisplatzes Fragen der Schriftgutverwaltung auf der Tagesordnung.

Einige dieser Fragen möchte ich hier vorstellen:

- **Dokument – Vorgang – Akte:** Verbunden mit diesem "Dreigestirn" ist zu allererst immer wieder die Diskussion um die Definitionen: Die Sichtweise der Sachbearbeiter in den Behörden ist generell auf den Vorgang als "Verwaltungshandeln" orientiert. Die Bearbeiterablagen (als vorherrschende Registraturform) werden selten als Teil der Schriftgutverwaltung wahrgenommen, so dass Verfügungstechniken wie z.d.A. selten sachgerecht angewendet werden. Es fällt den Sachbearbeitern daher schwer, den "Vorgang" im Sinne der Schriftgutverwaltung von der "Akte" zu unterscheiden. Es erweist sich daher als notwendig, die Sachbearbeiter im Zusammenhang mit der DMS-Einführung nicht nur in der Handhabung des Programms, sondern auch in GGO, RegR und Schriftgutverwaltung zu schulen. Die Unterscheidung von Dokumenten, Vorgängen und Akten spielt in den Schulungen eine fundamentale Rolle, weil dieses "Dreigestirn" in Brandenburg als unverzichtbar für eine strukturierte Schriftgutverwaltung angesehen wird. EL.DOK wird daher entsprechend dem auf dem Domea-Organisationskonzept fussenden Fachkonzept so installiert und administriert, dass kein Dokument ohne Vorgang und kein Vorgang ohne Akte angelegt werden kann. Lediglich beim Posteingang dürfen Dokumente in Postmappen verschickt werden, die vor der Bearbeitung zwingend aufgelöst werden müssen.
- **Federführung/Mitwirkung:** In der in Brandenburg vorherrschenden Bearbeiterablage stellte sich die Frage der Federführung bisher für den einzelnen Bearbeiter kaum. Verschleiert wurde das Problem auch durch organisationsbezogene Aktenpläne. Durch die Nutzung eines einheitlichen, sachbezogenen Aktenplans je Behörde im Projekt EL.DOK fiel zunehmend auf, dass unterschiedliche Organisationseinheiten auf ein Aktenplankennzeichen zugreifen. In den Diskussionen um das Rechtekonzept der Behörde ist zu entscheiden, ob alle Strukturteile auf eine Akten zugreifen oder ob mehrere Akten angelegt werden und ob es nicht sinnvoll wäre, im System zu hinterlegen, ob eine Akte als Mitwirkungs- oder federführende Akten geführt wird. Das BLHA unterstützt diese Unterscheidung, da die Aufbewahrungsfristen für Mitwirkungsakten herabgesetzt und diese Akten bei der Hinterlegung von Bewertungskatalogen im System (zweistufiges Aussonderungsverfahren) auf der Grundlage von Vereinbarungen über die Ausnahme aus der Anbietungspflicht nach § 4 Abs. 6 BbgArchivG automatisiert ausgesondert werden können.
- In den Diskussionen um die Anpassungskonzepte spielte die Handakte ebenfalls eine grosse Rolle. Einerseits wird darunter die Ablage der Informationen verstanden, die für die Erledigung der Arbeitsabläufe erforderlich, aber nicht direkt Gegenstand des Verwaltungs-

handelns sind ("Wissensmanagement"). Vielfach verbirgt sich darunter auch der Wunsch nach einer Ablagemöglichkeit für nicht aktenrelevante Notizen und für Vorentwürfen, die nicht im DMS/VBS gespeichert werden sollen, da sie noch nicht "reif" für den Geschäftsgang, d.h. auch für die Sichtung durch Kollegen und Vorgesetzte sind. Für die Akzeptanz des Projektes EL.DOK ist eine Lösung dieser Frage zwingend notwendig. Es wird daher für jeden Nutzer ein Handaktenbereich angelegt, der individuell strukturierbar ist. Er kann vom Bearbeiter zur Ablage nicht oder noch nicht aktenrelevanter Dokumente benutzt werden. Das System wird so konfiguriert, dass Dokumente mit Geschäftsgangvermerken nicht im Handaktenbereich gespeichert werden können.

- In allen Ministerien und natürlich auch im ZITBB und im BLHA als Projektteilnehmer erfolgte eine Revision oder Neuerarbeitung des Aktenplanes. Dieses wurde z.T. mit Hilfe externer Dienstleister oder zusätzlicher Kräfte realisiert. Rückblickend ist leider einschätzen, dass die Aktenplanrevision z.T. mit einer Verminderung der Gliederungstiefe verbunden war, was mit Zeitnot und auch der Möglichkeit der Volltextrecherche im System begründet wurde. Aus der Sicht des BLHA wurde oft für eine stärkere Untergliederung plädiert, um die Anlage von zu vielen Akten pro Betreffseinheit zu vermeiden und nicht den Weg zur Erarbeitung von Bewertungskatalogen zu verbauen. Da es, bedingt durch die vorherrschende Bearbeiterablage, z.T. an Erfahrungswerten über das tatsächliche Aktenaufkommen fehlte, fanden diese Argumente nicht immer Gehör. Nach der Einführungsphase, wenn die Zahl der Akten je Betreffseinheit zu gross und eine Strukturierung gewünscht wird, rechne ich allerdings wieder mit einer Rückkehr zur alten Gliederungstiefe.
- Gegenstand der Diskussion war auch die Registraturorganisation. Nachdem zuerst in der Mehrheit der Ministerien eine Fortsetzung der Bearbeiterablagen favorisiert wurde, besteht jetzt – auch auf Empfehlung des BLHA hin – weitgehend die Auffassung, die Rechte für die Anlage und Aussonderung von Akten an eine Registraturkraft zu vergeben. Dadurch können das bereits angesprochene Problem der Mitwirkungs-/Federführungsakten und die strukturübergreifende Vergabe eindeutiger Betreffe und damit eine vernünftige Recherchemöglichkeit gesteuert werden. Die Anlage von Vorgängen erfolgt durch die Bearbeiter.
- Die Aussonderung wird in der Regel nach dem vierstufigen Aussonderungsverfahren ablaufen, der Übergang zum zweistufigen Verfahren ist vom Fortschreiten der Erarbeitung von Katalogen und dem bereits erwähnten Abschluss von Vereinbarungen über die Ausnahme aus der Anbietungspflicht abhängig.
- Die Aussonderung erfolgt auf der Aktenebene, d.h., das System wird entsprechend der Fachkonzept so konfiguriert, dass alle geschlossenen Vorgänge eines Jahrgangs einer Akte zu einem Aktenschnitt zusammengefasst und nach Ablauf der Aufbewahrungsfrist dem Aussonderungsverfahren zugeführt werden.
- Einigkeit besteht im Projekt darüber, dass die Metadaten von Hybridakten im System mit verwaltet werden. Das BLHA gab die Empfehlung, auch die Registraturdaten der Altakten in das System mit zu übernehmen, um ein vollständiges Aktenverzeichnis der Behörde in einem System zu verwalten.

- Nach dem Fachkonzept ist im Augenblick mit Hybridaktenführung aus zwei Gründen zu rechnen: 1. bei Überformaten und 2. bei Dokumenten mit Urkundscharakter. Da diese von den Behörden identifiziert werden müssen, wird angestrebt, dieses Problem in einem der künftigen Workshops des Gesamtprojekts (siehe oben) zu diskutieren.
- Für die Anbindung der zahlreichen Fachverfahren ist es erforderlich, die ins Projekt EL.DOK zu übernehmenden Daten zu klären und darüber hinaus festzulegen, ob in EL.DOK oder im Fachverfahren die "Originale" aufbewahrt werden ("Führende Akte"). Dies ist Voraussetzung für die Festlegung von Aufbewahrungsfristen.

EL.Projekte und EL.Infrastruktur

Wie aus den bisherigen Ausführungen deutlich wurde, wird das Projekt EL.DOK in Brandenburg im Zusammenwirken mit den anderen EL-Projekten realisiert, um Synergieeffekte zu erzielen und den Datenaustausch über die XDomea-Schnittstelle zu ermöglichen. Zusammenfassend sind diese in anliegender Grafik dargestellt.

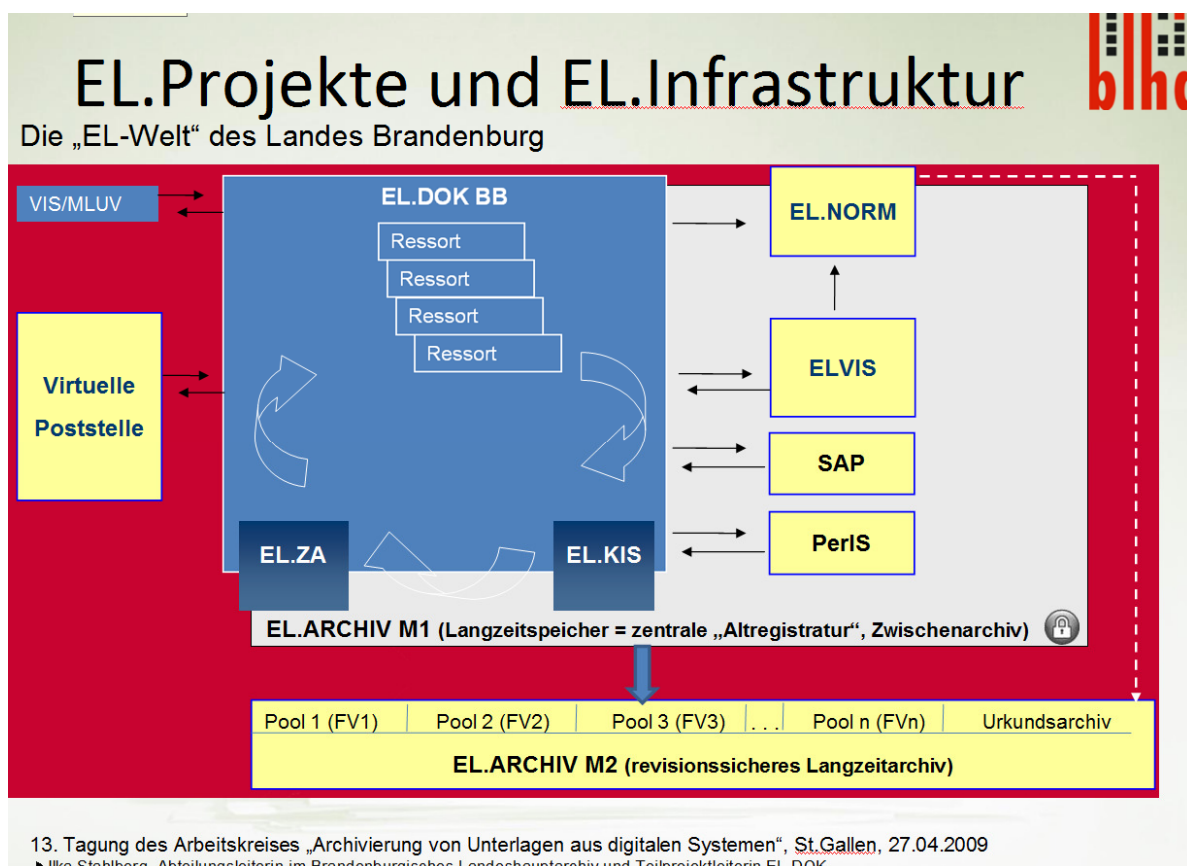


Abbildung 1: EL-Welt im Land Brandenburg

Auf dem zentralen DMS/VBS EL.DOK wird das elektronische Kabinettsinformationssystem (EL.KIS) aufgesetzt, in welchem Kabinettsitzungen vorbereitet werden sowie die Kommunikation mit dem DMS/VBS des Landtags (ELVIS) realisiert wird. Über EL.NORM erfolgt die elektronische Normenverkündung. Die Fachverfahren NFM (Neues Finanzmanagement – SAP) und PerIS (Personalinformationssystem) werden zentral in EL.DOK eingebunden. Ein Ministerium bleibt aus Bestandsschutzgründen bei seinem bisherigen System DMS/VBS VISkompakt und

kommuniziert über die XDomea-Schnittstelle mit EL.DOK.

Über die gleiche Schnittstelle erfolgt die Anbindung von EL.Archiv des BLHA. EL.Norm wird die verkündeten Normen unmittelbar nach Verkündung in das Urkundszwischenarchiv überführen.

Digitales Zwischenarchiv

Wie bereits in der Grafik zur EL-Infrastruktur in Brandenburg sichtbar, wird angestrebt zum Lückenschluss zwischen EL.DOK und EL.Archiv, ein Digitales Zwischenarchiv aufzubauen, in dem die "Aktenschnitte" in der Transferfrist (nach Domea) gespeichert und somit die Hochverfügbarkeitsspeicher der aktiven Systeme entlastet werden. Die Konzipierung erfolgt durch das Innenministerium unter Beteiligung des BLHA. Es sind Absprachen mit dem Datenschutz, der IT-Leitstelle und potentiellen Nutzern erfolgt und die Finanzierung ist in die Haushaltsverhandlungen 2010 eingebracht worden. In einer ersten Phase kann die Infrastruktur von EL.Archiv genutzt werden.

Über folgende Punkte besteht Einvernehmen: Die Speicherung erfolgt mandantengetrennt, verschlüsselt und im Transferformat, d.h. EL.DOK und die Fachanwendungen übergeben ihre digitalen Objekte im definierten Transferformat. Das System muss übersignieren können (Archisafe-konform). Für EL.DOK ist eine Transferfrist von 3 Jahren nach Bildung des Aktenschnitts (siehe oben) angedacht. Die Metadaten verbleiben in der Anwendungsdatenbank (für die Recherche und Aussonderungssteuerung). Dem Benutzer wird der Aktenschnitt im Transferformat zur Verfügung gestellt. Der Aktenschnitt bleibt geschlossen. Bei Bedarf muss deshalb ein Folgevorgang für die Bearbeitung angelegt werden. Die Aufbewahrungsfrist eines benutzten Aktenschnittes allerdings beginnt wieder von neuem zu laufen.

Fazit

Ziel des Beitrags war es, über den Sachstand eines mehrjährigen Projektes zu berichten, dabei auf Schwer- und Streitpunkte bei der Beratungstätigkeit hinzuweisen und zu zeigen, dass ein Archiv bei rechtzeitiger – wenn auch aufwendiger – Mitwirkung durchaus steuernd wirken und damit künftige Archivierungen erleichtern kann.

Das "Digitale Archiv" des Bundesarchivs

Kathrin Schroeder, Karsten Huth, Niels Hoppe, Ekkehard Fertig-Bilger

Ausgangssituation

Im Jahr 2006 erfolgten die ersten Aussonderungen elektronischer Akten an das Bundesarchiv. Der Datenbestand an originärem digitalen Archivgut umfasst über 9 Millionen Dateien, die zu 203 digitalen Archivobjekten, also inhaltlich zusammengehörenden Datenkomplexen, zusammengestellt wurden.

Die zu diesem Zeitpunkt vorhandene Infrastruktur war nicht für die Übernahme elektronischer Akten ausgelegt und stiess zudem an die Grenzen ihrer Kapazität. Deshalb wurde ein Projekt initiiert, welches das Ziel hatte, ein "Digitales Archiv" aufzubauen, um dadurch in der Lage zu sein, der gesetzlichen Verpflichtung in effizienter Weise nachzukommen. Gefragt war ein neuer Ansatz für die digitale Langzeitarchivierung.

Seit Mitte Oktober 2008 konnte ein OAIS-konformes Archivinformationssystem im Bundesarchiv in den Echtbetrieb gehen. Diese Lösung wurde mit HP als Generalunternehmer und dem Partner SER geschaffen. Damit ist es möglich, digitale Dokumente zu bewerten, ins Archivsystem zu übernehmen sowie dauerhaft zu sichern und zu nutzen.

Um in Zukunft regelmässig aus verschiedenen Quellen elektronische Daten in hohem Umfang übernehmen zu können, ist das System von Grund auf so angelegt, dass es eine hohe Skalierbarkeit bietet – sowohl in Bezug auf die Datenmengen als auch auf die Einbindung weiterer Behörden sowie unterschiedlicher Ausgangsformate wie z.B. der elektronischen Dokumente aus der Bundeswehr.

Unser Beitrag stellt das Produktivsystem unter verschiedenen Aspekten vor, wie z.B. den Gesamtprozess "Von der Behörde ins Archiv", die technische Lösung, Erfahrungen während der Projektphase und den Umgang mit Daten aus Filesystemen.

Der Gesamtprozess "Von der Behörde ins Archiv"

Daten aufbereiten, Daten archivieren, Datenzugriff sichern: Dies ist der wesentliche Ablauf, den die Architektur des Systems mit einer Workflow- und Archivkomponente abbildet. Eine hohe Qualität der Daten kann nur sichergestellt werden, wenn diese durch einen definierten Eingangsprozess aufbereitet werden. Ein weiteres Ziel ist, die Bewertung durch Trennung von Bewertung und Erschliessung zu effektivieren. Weitere Synergieeffekte werden durch die Verzahnung von etablierten mit neuen Verfahren zur Übernahme von Schriftgut in Papier- oder elektronischer Form erwartet.

Endet die Aufbewahrungsfrist der einzelnen Bundesbehörden, bieten diese ihre Daten dem Bundesarchiv an. Gesteuert werden beide Prozesse durch das Standard-Archivierungsmodul (SAM), dessen Entwicklung als EfA-Projekt (Einer-für-Alle)¹⁸ finanziell gefördert wurde. Das

¹⁸ EfA-Dienste: SAM http://www.cio.bund.de/cln_093/DE/IT-Angebot/EfA-Dienste/SAM/sam_node.html letzter Zugriff am 19.12.2008.

Datenangebot wird nach einer Aufbereitung der Daten automatisch in den Workflow übernommen und dem zuständigen Referat zur Bewertung zugesandt. Welche Daten archivwürdig und welche sofort vernichtet werden können, prüfen die zuständigen Kolleginnen und Kollegen aus den Fachreferaten auf Basis der Metadaten¹⁹. Abbildung 2 zeigt die Darstellung der Daten für die Bewertung nach der Aufbereitung für den Workflow. Diese Daten lagen ursprünglich in einer XML-Struktur vor.

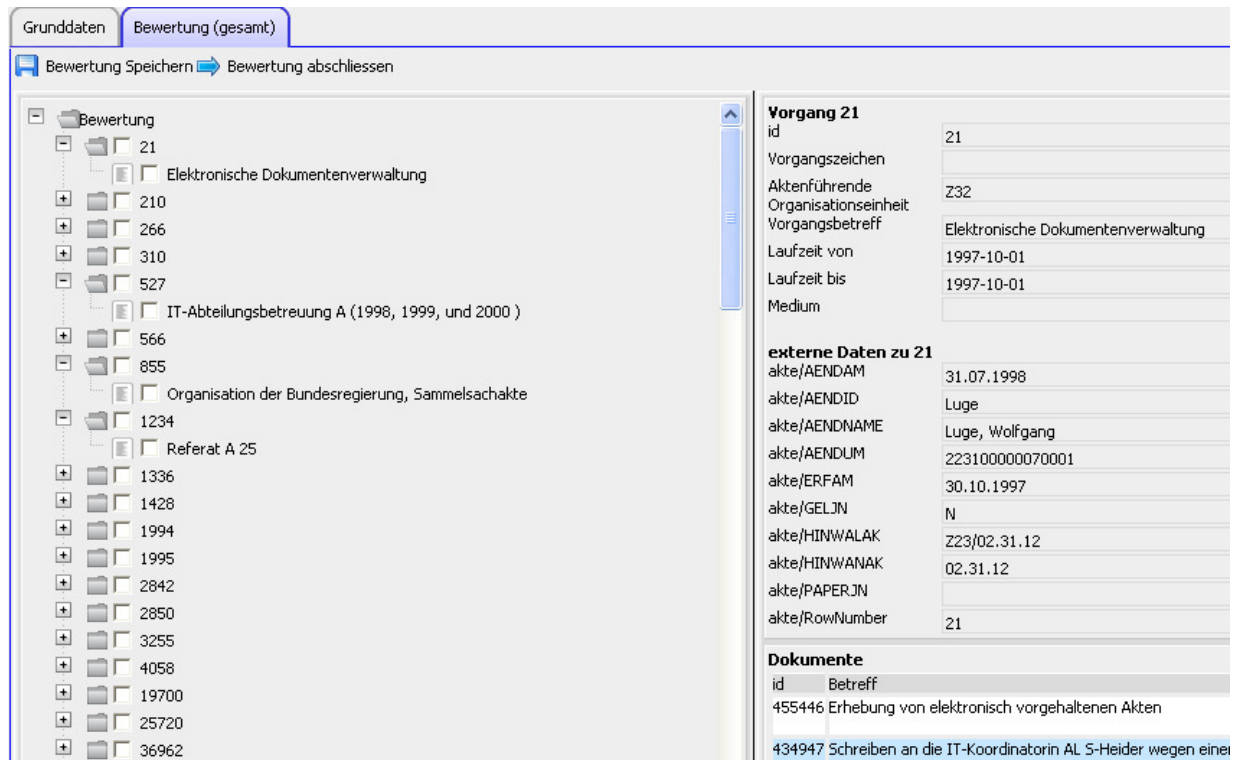


Abbildung 2: Aufbereitete Daten für die Bewertung

Die abgebende Behörde liefert auf der Grundlage der archivischen Vorbewertung daraufhin die angeforderten Daten (Aussonderung), die in den Workflow übernommen werden und dort den Ingest durchlaufen. Die Aussonderung besteht aus Metadaten im XML- oder CSV-Format und Dokumenten (Primärdaten). Die Primärdaten liegen im Idealfall im Format PDF/A vor. Sollten die Primärdaten in einem nicht archivfähigen Format übergeben werden, wird automatisch ein Konvertierungsprozess ausgelöst, in dessen Rahmen diese Daten nach PDF/A konvertiert werden²⁰. Abbildung 3 zeigt die Struktur eines SIPs. Der String "a1c75d9afc684003b924453399a14770c01b5dd6" ist die Transaktions-ID.

¹⁹ Damit wurde das 4-stufige Verfahren gemäss dem Erweiterungsmodul zum Organisationskonzept 2.0 Aussonderung und Archivierung elektronischer Akten (Oktober 2004) umgesetzt.

²⁰ Kathrin Schroeder: Format fürs Archiv. In: move – moderne Verwaltung. – 1/2009. – URL: http://www.pdfa.org/lib/exe/fetch.php?id=links&cache=cache&media=links:move_01_2009_s40_sch.pdf (letzter Zugriff: 9.7.2009).



Abbildung 3: Beispiel eines SIPs

Sind die Daten auf diese Weise vorbereitet und von den Kolleginnen und Kollegen aus den Fachreferaten auf deren Qualität geprüft worden, erfolgt die Übergabe an die Archivkomponente. Dort wird auf Basis der Metadaten der Index für die Recherche aufgebaut. Die AIPs werden auf Speicherplatten und parallel auf Band gespeichert. Die abgebende Behörde erhält eine Archivierungsbestätigung über die erfolgreiche Archivierung. Weitere Einzelheiten zur Aussonderung digitaler Unterlagen und deren Archivierung im Bundesarchiv für Behörden sind im Internet veröffentlicht.²¹ Nach einem definierten Zeitraum verbleiben die Pakete lediglich auf dem Band, dem kostengünstigeren Speichermedium. Von dort lassen sie sich allerdings bei Bedarf jederzeit auf schnellere Disk-Medien zurückholen.

Übernahme von Daten aus Filesystemen

Elektronische Akten, die in einem Vorgangsbearbeitungssystem erstellt und verwaltet wurden, sind trotz aller Heterogenität der Daten und Metadaten noch vergleichsweise gut strukturiert. Ganz anders verhält es sich, wenn Dokumente zusammen mit sonstigen Daten innerhalb der Behörde ohne Zuhilfenahme eines DMS oder VBS in einfachen Verzeichnisstrukturen eines Filesystems abgelegt und dann ausgesondert werden. Die Eins-zu-eins-Übernahme der Daten in der vorgegebenen Filesystemstruktur empfiehlt sich für das Archiv nicht, da die Formate für Filesysteme nicht Plattform-unabhängig sind und auch mit der Zeit technisch obsolet werden. Für diese Systeme können keine Schnittstellen zum Digitalen Archiv implementiert werden. Die Daten liegen in besonders unregelmässigen Verzeichnisstrukturen vor, wenn es innerhalb einer Behörde keine Regeln für Namenskonventionen bei den Verzeichnissen, den Primärdokumenten sowie zur Ablage gibt. Ohne Festlegung auf bestimmte Datenformate ist die Vielfalt der

²¹ Aussonderung digitaler Unterlagen und deren Archivierung im Bundesarchiv – Ein Leitfaden. – URL: <http://www.bundesarchiv.de/imperia/md/content/abteilungen/abtb/bbea/behoerdenleitfaden-v1.0-2009-05-08-internet.pdf> (letzter Zugriff: 9.7.2009).

Primärdaten erheblich. Insgesamt wurden bei einer Lieferung 65 verschiedene Dateiformate identifiziert. Noch ungünstiger für das Archiv ist der Mangel an beschreibenden Metadaten. Lediglich der Dateiname und der Verzeichnisname lassen – nur unzureichende – Rückschlüsse auf die Daten zu. Das Bundesarchiv hat in diesem Fall entschieden, dass diese Daten als elektronische Akten mit einer festen Akte-Vorgang-Dokument-Struktur überführt und entsprechende Metadaten vergeben werden. Dazu wurde im Projekt ein Strukturierungseditor entwickelt. Dieser ist in der Lage, strukturierte Daten aus Filesystemen in eine Akte-Vorgang-Dokument-Struktur zu überführen. Komprimierte Verzeichnisse (z.B. Dokumente, die in ZIP-Dateien abgelegt wurden) werden dabei entpackt. Der Editor verfügt über Eingabemasken für die nachträgliche Erfassung von Metadaten. Am Ende der Strukturierung erzeugt der Editor ein SIP, das dem Workflow (Ingest) übergeben werden kann. Damit wird das 2-stufige Verfahren gemäss dem Erweiterungsmodul zum Organisationskonzept 2.0 Aussonderung und Archivierung elektronischer Akten umgesetzt. Abbildung 4 zeigt die Oberfläche des Strukturierungseditors. Dieser kann auch in Behörden eingesetzt werden.

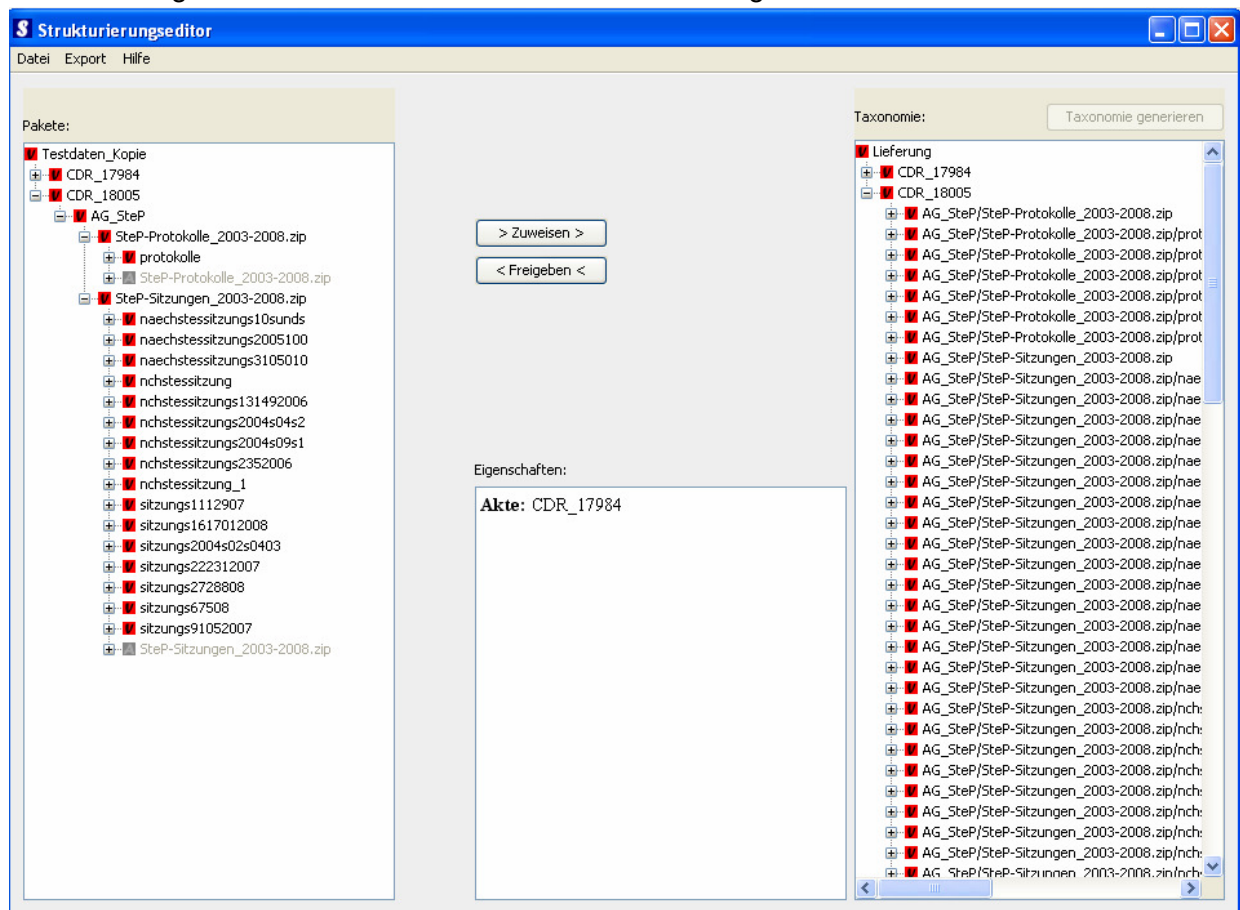


Abbildung 4: Benutzeroberfläche des Strukturierungseditors

In gewissen Grenzen ist die Übernahme von Daten aus Filesystemen als elektronische Akten möglich. Dies erfordert jedoch einen zusätzlichen Aufwand. Insgesamt wird aber durch solche Lieferungen die Bildung nachvollziehbarer Strukturen im Digitalen Archiv des Bundesarchivs erschwert.

Lösungsarchitektur

Herausforderungen

Die wesentliche Herausforderung beim Design einer Lösung, welche einen unbegrenzt langen Zeitraum überdauern soll, ist der Vorgang der Veränderung.

Die Lösung muss insbesondere im Hinblick auf technologische Veränderungen so gestaltet sein, dass möglichst effizient auf Veränderungen reagiert werden kann und dies in einem ungewöhnlich grossen Zeitrahmen.

Ein wesentliches Grundprinzip der Architektur ist deshalb die Modularität und Erweiterbarkeit der Lösung. Dies wird beispielsweise erreicht durch die Verwendung einfacher und klarer Schnittstellen und eine deutliche Abgrenzung von Teilkomponenten der Lösung. Die Ausrichtung am OAIS-Modell liefert hierbei eine gute Grundlage.

Weiterhin erfolgt die Abstraktion der physikalischen Speicherebene gegenüber den Applikationszugriffen und damit die Möglichkeit, neue Speichermedien verwenden und die bestehenden Daten effizient migrieren zu können.

Für Skalierbarkeit und Ausfallsicherheit werden etablierte Standardmechanismen genutzt, die einen möglichst niedrigen Komplexitätsgrad aufweisen.

Ausrichtung am OAIS-Modell

Die Lösung ist ausgerichtet an den funktionalen Blöcken des OAIS-Modells.

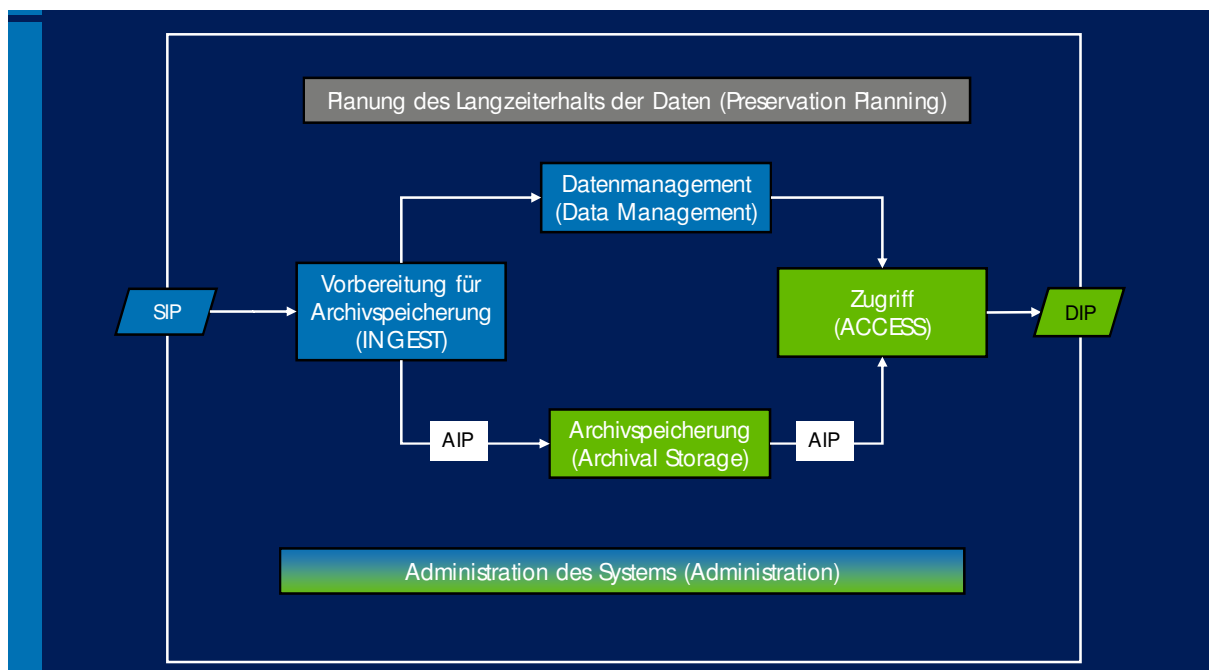


Abbildung 5: OAIS-Modell

In den folgenden Abbildungen wird deutlich, wie die einzelnen funktionalen Blöcke auf der Basis von Standardsoftware implementiert wurden.

Die Bilder geben ausschliesslich einen groben Überblick über die wesentlichen Komponenten, damit soll das Prinzip der Umsetzung aufgezeigt werden.

Workflowkomponente – Ingest

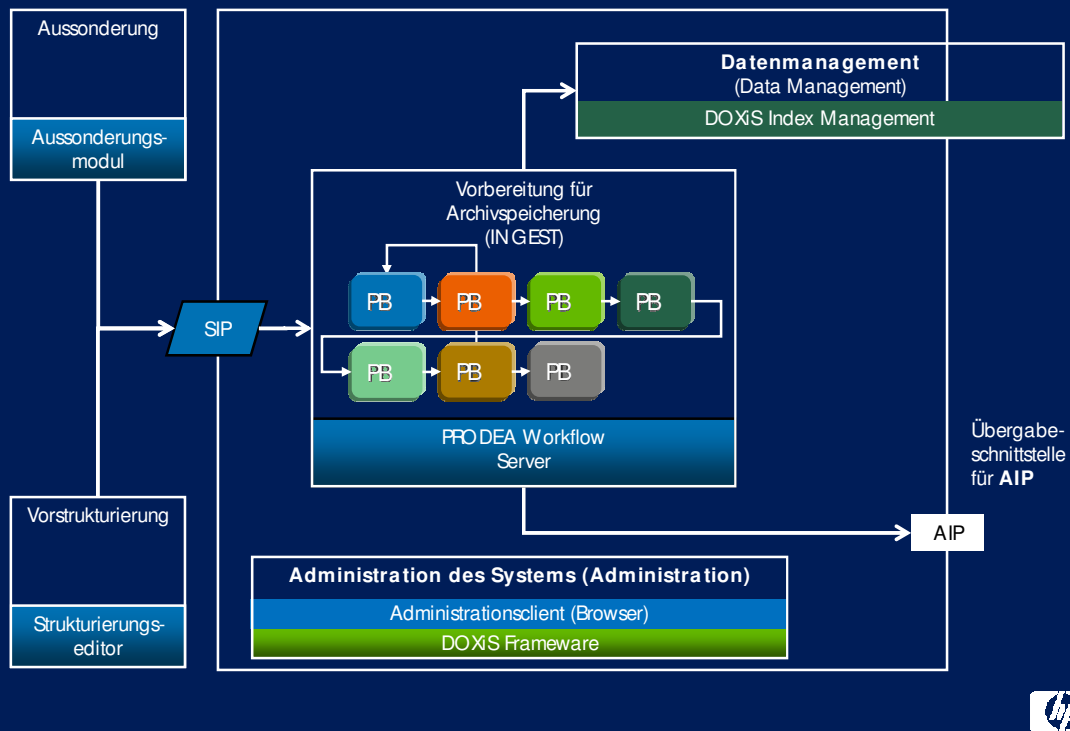


Abbildung 6: Funktionale Architektur – Workflowkomponente (Ingest)

Archivkomponente

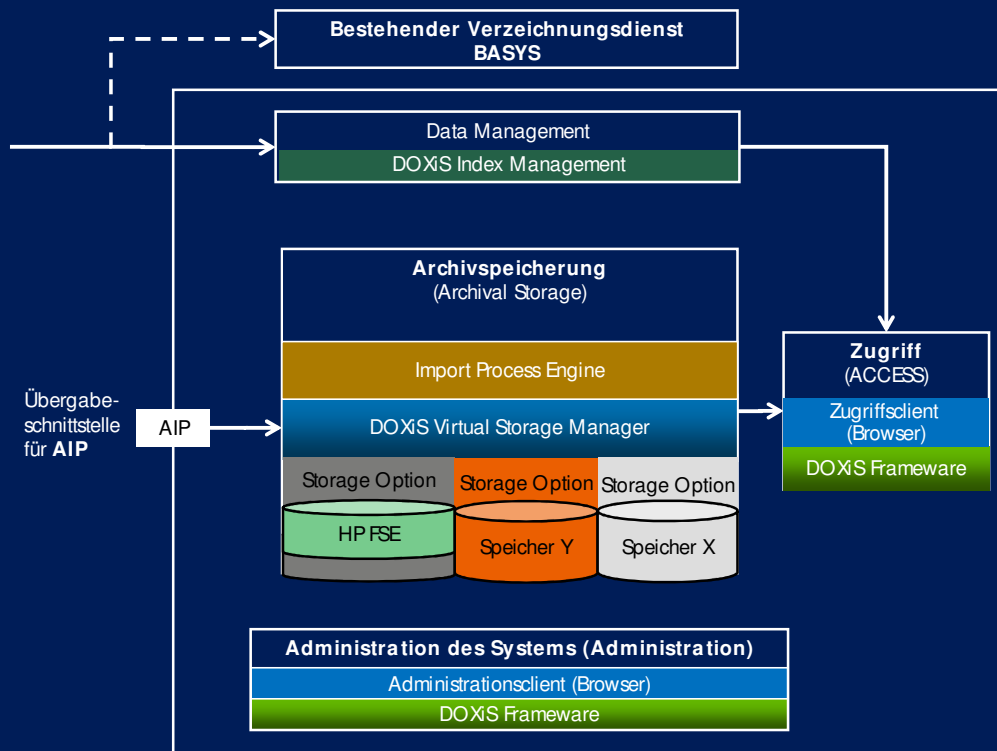


Abbildung 7: Funktionale Architektur – Archivkomponente

Software-Architektur

Die Softwarearchitektur ist komplexer als in den vorhergehenden Abbildungen dargestellt. Es kommen neben den Standardsoftwarekomponenten von SER Eigenentwicklungen und weitere Softwareprodukte anderer Hersteller bzw. auf Opensource Basis zum Einsatz. Dabei wurde immer nach dem "Best of Breed " Ansatz verfahren und darauf geachtet, dass die Komponenten möglichst sauber von einander abgegrenzt sind und klar definierte Schnittstellen besitzen.

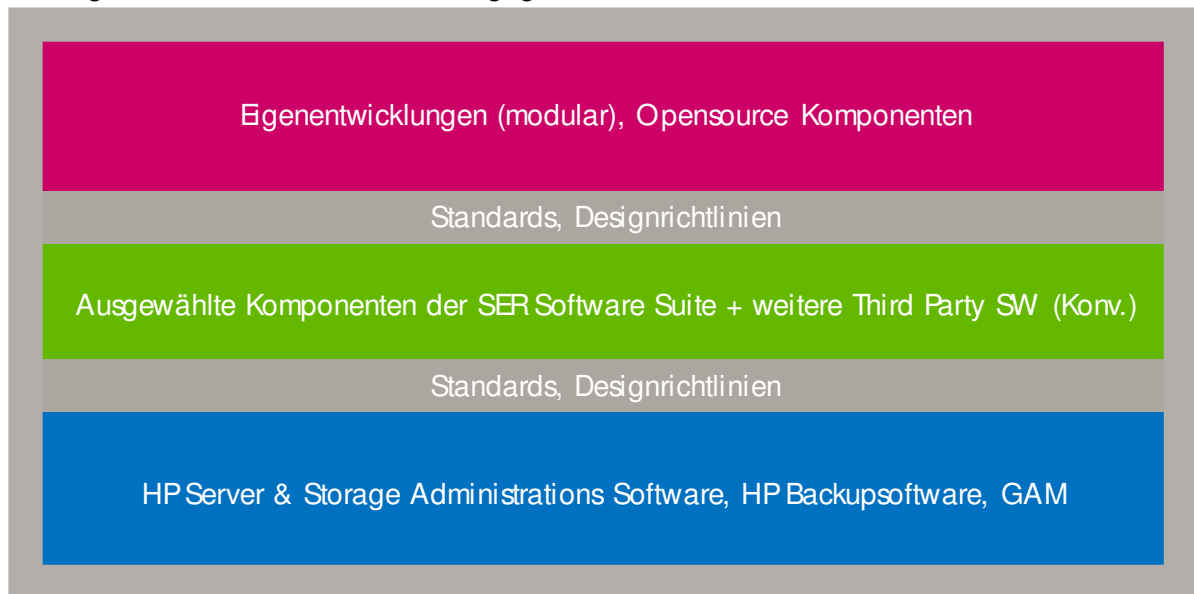


Abbildung 8: Software-Architektur

Datenmodell

Ein wesentlicher Teil der Lösung ist das Datenmodell. Dieses lehnt sich ebenfalls an das OAIS-Modell an.

Als wesentlicher Bestandteil des Datenmodells ist das Metadatenschema im Archiv zu sehen. Dieses beruht auf den folgenden vorhandenen Standards:

- PREMIS
Core preservation metadata set (Fokus auf technische Beschreibung der Objekte und Beschreibung von Veränderungen an diesen)
Strategien für die Enkodierung, Speicherung und das Management der Preservation Metadaten
- DC
Standardisiertes Set von Konventionen zur (fachlichen) Beschreibung von Dokumenten und anderen Objekten
DCMI
- XDOMEA
- Schnittstellendefinition zum Austausch von Dokumenten, Vorgängen und Akten zwischen unterschiedlichen Systemen in der Verwaltung
Da keiner der Standards für sich alle Zwecke für die Langzeitarchivierung im Bundesarchiv erfüllte, hat das Bundesarchiv auf Basis dieser Standards ein eigenes Metadatenschema (XBARCH) entwickelt, dieses wurde im Projekt an die zusätzlichen Erfordernisse ange-

passt.

Hardware-Architektur

Hardwareseitig besteht die Lösung aus den entsprechenden Servern, die in Bladetechnologie realisiert sind. Dies sorgt einerseits für einen möglichst niedrigen Stromverbrauch und weniger Abwärme und ermöglicht andererseits auch eine komfortablere Administration und eine schnellere Wiederherstellungen mit Hilfe von Images.

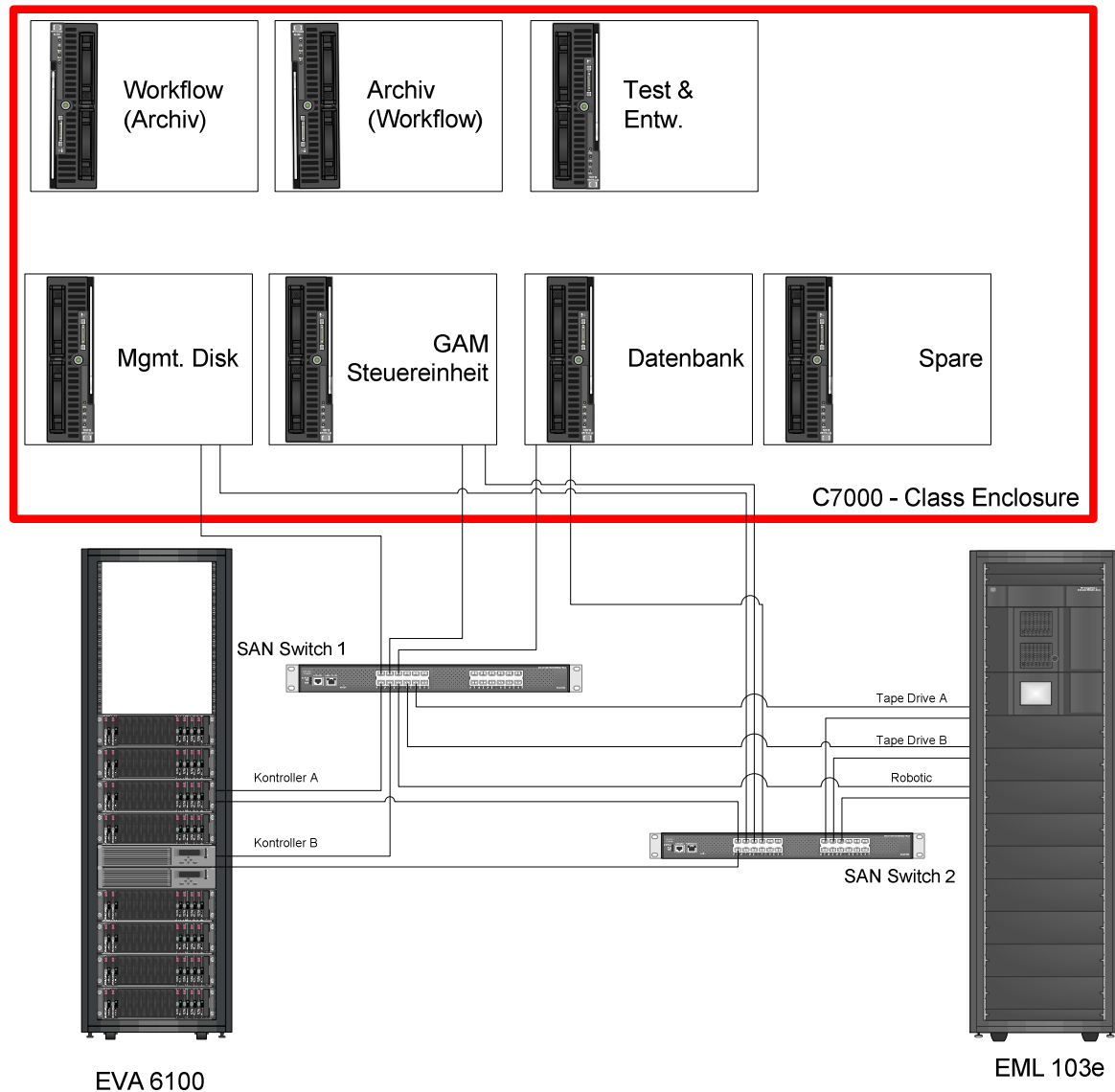


Abbildung 9: Hardware-Architektur

Projektvorgehen

Zeitplan

Was niemand so recht für möglich gehalten hatte, nämlich ein Digitales Langzeitarchiv innerhalb von acht Monaten produktiv zu setzen, wurde erreicht.

Dazu beigetragen haben insbesondere folgende Faktoren:

- Einsatz erprobter Software-Komponenten als Basis

- Klar strukturiertes Vorgehen bei der Erstellung der Spezifikation. In 6 Schritten wurden die Details zur Umsetzung der Lösung gemeinsam erarbeitet: Datenmodell, Archivsignatur; Prozesse/Masken für Anbieter/INGEST; Metadaten (Archiv XML); Prozesse/Masken für Archiv/Recherche; IT-Infrastruktur; Finalisierung Spezifikation
- Vor-Ort-Präsenz des HP Projektteams
- Flexibilität innerhalb des Projektzeitrahmens, aber mit
- Einer gemeinsamen Verpflichtung auf den Zieltermin
- Ein erprobtes Vorgehensmodell für Tests und Qualitätssicherung: Die Erstellung der Abnahmetest Spezifikation erfolgte sukzessive während der Spezifikationsphase, ebenso wurden die zugehörigen Abnahmetestpakete schrittweise erstellt. Alle "fachlichen" Abnahmetestpakete erstellte das Bundesarchiv, alle "technischen" Abnahmetestpakete erstellte HP. Die Testpakete wurden gemeinsam verabschiedet. Ein dedizierter HP-Testmanager betreute die Tests. Bestimmte Tests wurden gemeinsam mit dem Bundesarchiv durchgeführt, um schnellstmöglich auf die Ergebnisse zu reagieren.

Das Digitale Archiv als lebendes System

Dass ein Digitales Archiv kein statisches Gebilde sein kann, war dem Bundesarchiv schon in der Vorphase der Ausschreibung klar. Denn es war nicht nur ein Wachstum des Datenaufkommens zu erwarten, sondern es waren auch folgende Faktoren zu berücksichtigen:

- Konvertierung von Dokumenten in PDF/A
- Erweiterung um Archivierung aus diversen Fachverfahren
- Erweiterung des Nutzerkreises, auch ausserhalb des Bundesarchivs
- Anbindung des Erschliessungssystems BASYS

Dem wurde durch einen 3-jährigen Rahmenvertrag mit dem Realisierungspartner HP Rechnung getragen. Dadurch entstand ein schlagkräftiges und gut eingespieltes Projektteam, das nach der Übernahme unstrukturierter Daten aus Filesystemen (siehe oben) mittlerweile auch eine Konvertierungsplattform realisiert hat und sich in einem Folgeschritt der Anbindung des Digitalen Archivs an das Erschliessungssystem BASYS widmen wird.

Ausblick

Das Projekt startete im Februar 2008 und nahm im Oktober 2008 mit der ersten Ausbaustufe des Digitalen Archivs (40 TB) seinen regulären Betrieb auf. Das Speichersystem ist bis zu 400 TB skalierbar. Bis 2011 wird das Bundesarchiv in Zusammenarbeit mit HP und SER das System stufenweise erweitern. So ist gewährleistet, dass auf zukünftige Änderungen reagiert werden kann.

Normung und Best Practices im Records Management. Mehrwert der ISO-15489

Steffen Schwalm

Einführung

Die ISO-15489 definiert Records Management als:

"Oberbegriff für Methoden und Technologien zu Erstellung, Verwaltung, Beschreibung, Austausch, Organisation, Speicherung, Aussonderung, Abgabe und Vernichtung von Aufzeichnungen im Rahmen der Geschäftsführung öffentlicher wie privater Institutionen".

Records Management ist insofern system- und medienunabhängig. Die Anforderungen an ein ordnungsgemässes Records Management gelten sowohl für elektronische wie analoge Unterlagen, einschliesslich deren beschreibender Informationen, unabhängig davon, mit welchen Systemen diese verwaltet werden. Der Terminus schliesst den vollständigen Lebenszyklus geschäfts- bzw. rechtsrelevanter Unterlagen ein, von der Entstehung und Bearbeitung bis hin zur Langzeitspeicherung und Aussonderung. Records Management liefert damit strukturierte und aufbereitete, wiederauffindbare Aufzeichnungen für die Aufgabenerfüllung in Behörden wie die Geschäftsführung in privaten Unternehmen und stellt somit das informationelle Herzstück effizienter, nachvollziehbarer und damit rechtssicherer Geschäftsprozesse dar.

So impliziert das aus Art. 20 Abs. 3 Grundgesetz abgeleitete Rechtsstaatsprinzip, das Gebot der Aktenmässigkeit, wonach der Stand einer Sache jederzeit aus den Akten aller drei Staatsgewalten Legislative, Exekutive und Judikative erkennbar sein muss. Dies gilt medien- wie systemunabhängig sowie für den gesamten Lebenszyklus behördlicher Aufzeichnungen. Daraus schlussfolgert, dass alle drei Staatsgewalten gezwungen sind Massnahmen zur Erfüllung dieser Anforderungen zu treffen. Für Unternehmen ist insbesondere § 147 Abgabenordnung einschlägig, wonach steuerlich relevante Unterlagen geordnet aufzubewahren sind. Daneben gelten weitere Regelungen, z.B. §§ 238 und 257 HGB oder branchenspezifische Bedingungen.

Die Erfüllung der Forderung nach Nachvollziehbarkeit und Transparenz des Verwaltungshandelns bedingt Methoden und Technologien sowie organisatorische Regelungen zur "Verwaltung, Beschreibung, Austausch, Organisation, Speicherung, Aussonderung, Abgabe und Vernichtung von Aufzeichnungen im Rahmen der Geschäftsführung öffentlicher wie privater Institutionen", kurzum Records Management. Ohne eine regelbasierte Zusammenführung logisch zusammengehöriger Dokumente (Akten- und Vorgangsbildung) sowie die Registrierung und Bereitstellung rechts- wie geschäftsrelevanter Unterlagen ist deren zielgerichtete Nutzung sowohl im Rahmen des behördlichen Geschäftsgangs wie auch innerhalb eines Privatunternehmens kaum möglich. Dies gilt sowohl für die Erfüllung rechtlicher Anforderungen als auch für die Beantwortung komplexer Fragestellungen auf Basis der in den entsprechenden Aufzeich-

nungen befindlichen Informationen, so im Rahmen eines Wissensmanagements²². Dabei ist im konkreten Projekt zu entscheiden, welche Methoden und Technologien als Grundlage zur Strukturierung und Ordnung des Schriftguts, der unmittelbaren Akten- und Vorgangsbildung oder der Registrierung zum Einsatz kommen, sei es bspw. ein einheitlicher Aktenplan oder eine Klassifikation, die zur Registrierung genutzt wird, sei es die qualifizierte elektronische Signatur zur Beweissicherung oder ausschliesslich ein umfassendes Rechtekonzept für den Zugriff auf die betroffenen Unterlagen – um nur einige Beispiele zu nennen.

Status quo in Deutschland

In der deutschen Terminologie ist der Begriff Records Management mehrheitlich mit Schriftgutverwaltung übersetzt. Mit letzterem Terminus werden häufig DMS/VBS für öffentliche Institutionen nach dem DOMEA®-Konzept verbunden. In der Praxis werden jedoch rechtsrelevante Unterlagen nicht ausschliesslich in DMS/VBS verwaltet, sondern in spezifischen Fachverfahren. Diese Lösungen dienen vordringlich der Erfüllung besonderer Fachaufgaben so bspw. im Ausländerwesen oder im Umweltschutz. Deren Nutzung erzeugt aktenrelevante Unterlagen, die in der Folge jedoch in den Fachverfahren verbleiben und nicht in einem DMS/VBS gespeichert werden. Daneben bestehen vielfach Unterlagen, die auf Gruppen- wie persönlichen Laufwerken oder E-Mail-Postkörben abgelegt werden neben einer vollständig papiernen Aktenführung. Dies gilt sowohl für Unternehmen als auch für öffentliche Institutionen. Die Erhaltung der Rechts- und Beweissicherheit sowie effiziente Geschäftsprozesse erfordern die Registrierung, Erfassung und regelbasierte Aufbewahrung der gespeicherten rechtsrelevanten elektronischen Dokumente (verbindliche Registraturrichtlinien, Aktenpläne, elektronische Signaturen). Liegen bspw. rechts- und geschäftsrelevante Unterlagen unstrukturiert in analoger oder digitaler Form undifferenziert auf verschiedenen Laufwerken, Datenträgern etc. vor, so ist die Nachvollziehbarkeit und damit die Rechts- und Beweissicherheit zumindest gefährdet. Werden darüber hinaus Formerfordernisse nicht erfüllt, steht auch die Nichtigkeit behördlicher Entscheidungen und in der Folge eine deutliche Rechtsunsicherheit im Raum. Der international übliche Terminus technicus *Records Management* ist branchen- wie systemunabhängig und könnte hier der umfassendere Begriff sein, um mit den entsprechenden Methoden und Technologien alle Verfahren, in denen rechtsrelevante Unterlagen gespeichert werden, abzudecken.

Bedeutung und Mehrwert der ISO-15489

Die ISO-15489 bilden, wie in der Definition von Records Management angedeutet, derzeit die faktisch einzige internationale Norm, die grundlegende Anforderungen an ein ordnungsgemässes Records Management definiert. Dazu zählen neben fachlich-organisatorischen Bedingungen an die Verwaltung, Speicherung oder Nutzung rechtsrelevanter Aufzeichnungen auch Anforderungen an entsprechende IT-Verfahren sowie Hinweise zu deren Planung und Implementierung. Die Norm beschreibt allgemeingültige Rahmenbedingungen für alle Branchen und Nut-

²² Vgl. auch Toebak, Peter M.: Records Management. Ein Handbuch. Baden 2007.

zer. So sind als Grundbedingungen durch Records Management und entsprechende Systeme (RMS) die

- Authentizität,
- Verfügbarkeit,
- Integrität und
- Nutzbarkeit

der verwalteten rechts- wie geschäftsrelevanten Unterlagen sicherzustellen. Dabei inkludieren die ISO-15489 den gesamten Lebenszyklus analoger wie elektronischer Unterlagen bis hin zur Aussonderung gegenüber dem zuständigen Archiv. Insofern stellen die ISO-15489 faktisch das "Grundgesetz" einer rechtssicheren Entstehung, Bearbeitung, Verwaltung, Langzeitspeicherung und Archivierung rechts- und geschäftsrelevanter Unterlagen dar.

Genau in diesem grundlegenden und allgemeingültigen Charakter ist der Mehrwert der ISO-15489 zu sehen. Die Norm definiert übergreifend und international gültige Anforderungen, die auf die jeweiligen Branchen übertragbar sind. Diese allgemeinen Inhalte sind in der Folge für den entsprechenden Anwender und das Anwendungsgebiet zu spezifizieren.

So ist im konkreten Projekt zu definieren wie bspw. die Authentizität sichergestellt wird, sei es durch Nutzung einer qualifizierten elektronischen Signatur und/oder ein umfassendes Rechtskonzept des eingesetzten DMS einschliesslich einer fundierten Verfahrensdokumentation. Hier bestehen vielfältige Möglichkeiten. Gleiches gilt für den Schutz der Integrität, sei es durch Hashwert oder datenträgerbezogene Massnahmen.

Die ISO-15489 vermeiden durch den übergreifenden Charakter ausdrücklich Redundanz und das "Übersehen" oder "Ignorieren" essentieller Bedingungen wie auch Verantwortlichkeiten. Letztere werden von den ISO-15489 klar definiert:

- Records Management Professionals
- Entscheider
- Systemadministratoren
- Alle Angestellten der jeweiligen Institution

Hieraus lassen sich Aufgaben, Pflichten und Rechte explizit ableiten – ein nicht zu unterschätzender Mehrwert vor dem Hintergrund, dass eine Vielzahl an Projekten im Kontext von Records Management aufgrund nicht klar definierter Verantwortlichkeiten stocken oder gar scheitern. Abgesehen davon lassen sich aus definierten Verantwortlichkeiten auch Ansprüche personeller wie finanzieller Art ableiten (betrifft Entscheidungsebenen in Verwaltung und Unternehmen). Im Sinne einer vollständigen archivischen Überlieferung, deren Grundlage ein ordnungsgemässes Records Management bildet, sind solcherart definierte Verantwortlichkeiten und Ansprüche nicht zu unterschätzen.

Die ISO-15489 ermöglicht insofern auch eine Qualitätssicherung des Records Managements, so bspw. durch regelmässige Audits auf Basis der Norm oder daraus abgeleiteter nationaler Standards. In den Niederlanden werden solche Audits, an denen u.a. das Nationalarchiv beteiligt ist und deren Ergebnis die Budgetierung der betroffenen Behörden massgeblich beeinflusst, bereits durchgeführt.

Die ISO-15489 umschreibt den grundlegenden Rahmen, in dem sich aus der Norm abgeleitete Standards bewegen können. Die unmittelbare Ausgestaltung muss national resp. im Projekt erfolgen. Diese Aufgabe kann und soll eine internationale Norm nicht übernehmen, es widerspräche dem übergreifenden Anspruch und Charakter. Abbildung 10 zeigt beispielhaft zahlreiche anerkannte internationale Standards, die auf der ISO-15489 basieren:

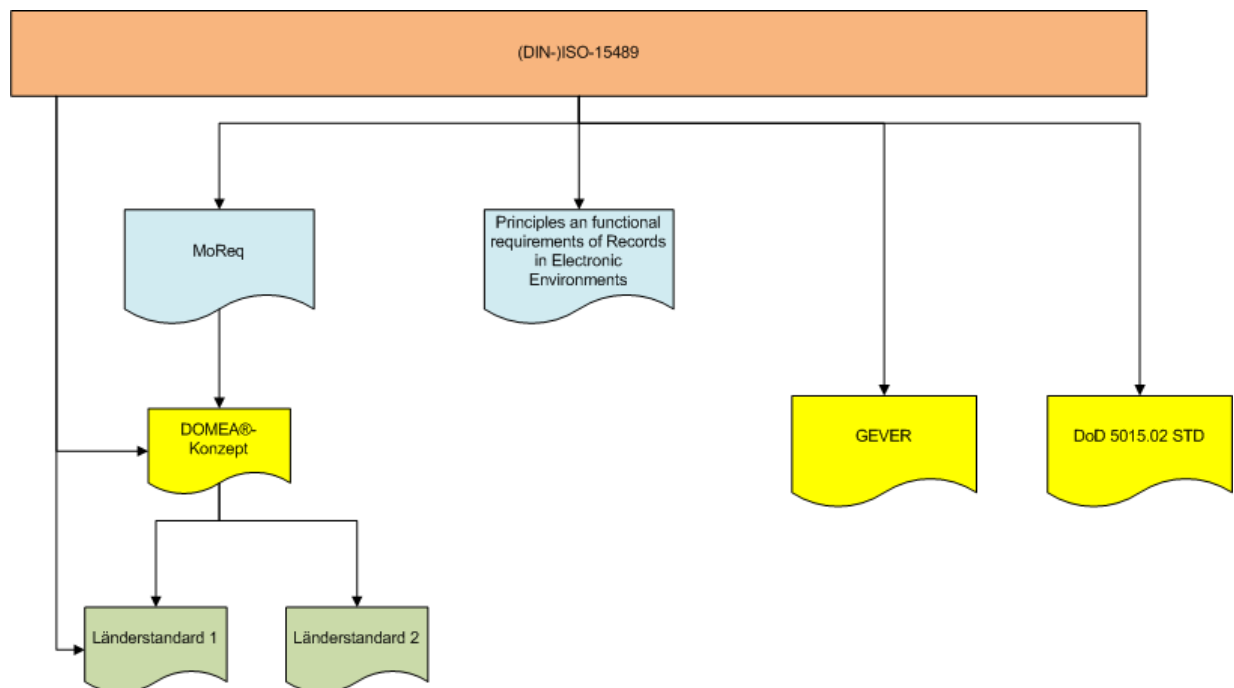


Abbildung 10: Übersicht Standards basierend auf ISO-15489

Aktuell erfolgt eine Überarbeitung der ISO-15489. Es steht zu hoffen, dass der beschriebene Mehrwert auch nach der Revision erhalten bleibt.

Das Digitale Archiv des Bundesarchivs: Fachliche Anforderungen für die Übernahme elektronischer Akten aus Vorgangsbearbeitungssystemen

Dr. Claudia Zenker-Oertel

Ein Bericht über das Digitale Archiv des Bundesarchivs, das seit dem vergangenen Herbst fertiggestellt und aufnahmebereit ist, wäre ohne das Eingehen auf die flankierenden rechtlichen Rahmenbedingungen unvollständig. Somit erscheint es sinnvoll und geboten, zunächst kurz auf das Bundesarchivgesetz (BArchG) und seine entsprechenden Paragraphen einzugehen, die sich unter anderem mit der Archivierung von Unterlagen aus digitalen Systemen beschäftigen.

Dieser rechtliche Zusammenhang, die grundlegenden Voraussetzungen für das Digitale Archiv und die Perspektiven desselben sollen im Folgenden grundsätzlich und knapp betrachtet werden.²³

Rechtliche Grundlagen

Das Bundesarchivgesetz vom 6. Januar 1988 ist im Vergleich mit den alten Bundesländern und ihrer Archivgesetzgebung (mit Ausnahme des Archivgesetzes für Baden-Württemberg von 1987) ein frühes und wegweisendes – auch für die Archivgesetze, die ab 1990 im wiedervereinigten Deutschland entstanden. Das Bundesarchivgesetz sieht in seinen Paragraphen 1 und 2 eine Reihe von heute selbstverständlich klingenden Aufgaben und Zuständigkeiten vor, die sich allerdings für den Aufbau und Betrieb des Digitalen Archivs als zentral erweisen:

§ 1 Das Archivgut des Bundes ist durch das Bundesarchiv auf Dauer zu sichern, nutzbar zu machen und wissenschaftlich zu verwerten.

§ 2, 1 Die Verfassungsorgane, Behörden und Gerichte des Bundes, die bundesunmittelbaren Körperschaften, Anstalten und Stiftungen des öffentlichen Rechts und die sonstigen Stellen des Bundes haben alle Unterlagen, die sie zur Erfüllung ihrer öffentlichen Aufgaben einschliesslich der Wahrung der Sicherheit der Bundesrepublik Deutschland oder eines ihrer Länder nicht mehr benötigen, dem Bundesarchiv oder in Fällen des Absatzes 3 dem zuständigen Landesarchiv zur Übernahme anzubieten und, wenn es sich um Unterlagen von bleibendem Wert im Sinne des § 3 handelt, als Archivgut des Bundes zu übergeben.

§ 2, 8 Unterlagen im Sinne dieses Gesetzes sind Akten, Schriftstücke, Karten, Pläne sowie Träger von Daten-, Bild-, Film-, Ton- und sonstigen Aufzeichnungen, die bei den in Absatz 1 genannten Stellen des Bundes, bei Stellen der DDR, bei Stellen der Besatzungszonen, des Deutschen Reiches oder des Deutschen Bundes erwachsen oder in deren Eigentum überge-

²³ Der Aufsatz basiert auf dem am 27.04.2009 in St.Gallen gehaltenen einführenden Vortrag. Der fünfteilige Vortrag wurde ferner von Kathrin Schroeder, Niels Hoppe, Karsten Huth und Ekkehard Fertig-Bilger vorgetragen.

gangen oder diesen zur Nutzung überlassen worden sind.²⁴

Das Bundesarchiv hat somit den gesetzlich festgeschriebenen Auftrag, auch elektronische Unterlagen zu übernehmen und dauerhaft zu sichern. Diese elektronischen Unterlagen sind in der Regel im Rahmen der Aufgabenerledigung in der Bundesverwaltung entstanden: Es handelt sich dabei sowohl um Akten aus elektronischen Schriftgutverwaltungs- oder Dokumentenmanagementsystemen als auch um Datenbanken und andere Objekte aus elektronischen Fachverfahren. Die Formulierungen im Bundesarchivgesetz klingen 1988 im Hinblick auf die Archivierung digitaler Unterlagen zwar noch vage, da sie explizit auf "Datenträger" abheben, bilden aber gleichzeitig bis heute die Voraussetzung für die Arbeit des Digitalen Archivs.²⁵

Vorarbeiten im Bundesarchiv

Konsequenterweise muss das Bundesarchiv neue Verfahren zur Übernahme elektronischer Akten, Fachverfahren u.ä. von erheblicher technischer Komplexität schaffen, um auch weiterhin die gesetzliche Vorgabe des Bundesarchivgesetzes erfüllen und um Überlieferungsverluste vermeiden zu können.

Diese Verfahren müssen folglich einsatzbereit sein, bevor die grosse Masse elektronischer Objekte dem Archiv angeboten wird.

Gleichzeitig verfügt das Bundesarchiv bereits über jahrelange Erfahrungen und umfangreiche Daten, da schon nach 1990 Dateien aus Anwendungen zentraler staatlicher Stellen der früheren DDR übernommen wurden. Diese für die Forschung wichtigsten Dateien der DDR wurden gesichert, die Findmittel fertiggestellt und der Öffentlichkeit zugänglich gemacht, nachdem kopiert, konvertiert, anonymisiert und dokumentiert wurde. Hierzu gehören historisch bedeutsame Datenbestände wie der Zentrale Kaderdatenspeicher, das Eingabenprojekt des Staatsrates, die Datei Grenzzwischenfälle oder die Strafgefangenen- und Verhaftetendateien ebenso wie zentrale Statistiken aus dem Gesundheitsbereich, der Wirtschaft, dem Bildungswesen oder zur Kriminalität. Die enthaltenen Informationen sind nicht nur für die Wissenschaft, sondern zurzeit vor allem für die Betroffenen von grosser Wichtigkeit und bilden einen der Schwerpunkte bei den Anfragen an das Bundesarchiv.²⁶

Der Datenbestand an originärem digitalen Archivgut, das heisst an Archivgut, das bereits in elektronischer Form entstanden ist, umfasst 9,2 Millionen Dateien, die zu 203 Digitalen Archiv-

²⁴ Vgl. dazu das Bundesarchivgesetz (BArchG) vom 06.01.1988 (BGBl. I S. 62), zuletzt geändert durch das Informationsfreiheitsgesetz vom 05.09.2005 (BGBl. I S. 2722). Siehe dazu www.bundesarchiv.de/benutzung/rechtsgrundlagen/bundesarchivgesetz/index.html sowie Siegfried Becker/Klaus Oldenhage, Bundesarchivgesetz. Handkommentar, Baden-Baden 2006.

²⁵ Die Formulierung "Datenträger" ist der Entstehungszeit des Gesetzes geschuldet und widerspiegelt den damaligen Stand der Technik. Auch wenn der Begriff "Dateien" nicht erwähnt wird, sind diese jedoch genauso gemeint wie die "Datenträger".

²⁶ Siehe zum Projekt der Archivierung von DDR-Daten im Bundesarchiv auch den Bericht und das dahinterstehende technische Konzept: www.bundesarchiv.de/aktuelles/projekte/00015/index.html.

objekten in den Beständen BRD und DDR zusammengestellt wurden und deren älteste Daten aus den 1970er Jahren stammen.²⁷

Aufbau des Digitalen Archivs

Seit Herbst 2008 wird in Zusammenarbeit mit den Firmen HP und SER ein Digitales Archiv im Bundesarchiv aufgebaut, um die bereits vorhandene Menge digitaler Unterlagen im Sinne des Bundesarchivgesetzes dauerhaft zu sichern und nach Ablauf der entsprechenden Schutzfristen für die Öffentlichkeit zugänglich zu machen.

In der nächsten Zeit werden die elektronischen Unterlagen aus der Bundesverwaltung noch zahlenmässig geringer an das Bundesarchiv abgegeben werden. Durch zahlreiche und jahrelange Kontakte zu den abgebenden Stellen im Rahmen der regulären Behördenberatung zeigt sich, dass insbesondere ab 2012 die Zahl der elektronischen Unterlagen stetig steigen wird. Diese abzugebenden Unterlagen stammen aus technisch heterogenen Systemen und werden erst im Zuge der Archivierung mit Hilfe der Konvertierungsplattform in ein dauerhaftes Archivformat überführt.²⁸

Abschliessend ist vom Bundesarchiv geplant, ein Digitales Zwischenarchiv nach dem bisherigen Zwischenarchivmodell des Bundes als sogenannten Shared Service für die obersten Bundesbehörden aufzubauen. Dieses Digitale Zwischenarchiv wäre vom Prinzip her an das allgemein geschätzte und praktikable Zwischenarchivverfahren beim konventionellen Schriftgut in St. Augustin bei Bonn, Hoppegarten bei Berlin sowie Freiburg i. Br. angelehnt, bei dem sich die Behörden zu einem denkbar frühen Zeitpunkt von den elektronischen Unterlagen und ihrem weiteren Erhalt entlasten.²⁹

Zur Vorbereitung der automatisierten Aussonderung und Übernahme elektronischer Akten kann ein Beratungs- und Informationsgespräch im Vorfeld helfen, in dem technische und organisatorische Fragen abgestimmt werden, mindestens ist jedoch das Übersenden von Testdaten im Vorfeld erforderlich. Für das Anbieten, Bewerten und die Übernahme elektronischer Daten kommen alternativ das 2- oder 4-stufige Verfahren des DOMEA-Aussonderungskonzeptes zum Einsatz. Voraussetzung für das 2-stufige Verfahren ist das Vorliegen eines Bewertungskataloges, der auf der Basis des Aktenplans durch die zuständigen Archivare im Benehmen mit der abgebenden Stelle erarbeitet wird. Das heisst, dass von der abgebenden Stelle nur diejenigen SIPs zu erzeugen und zu übermitteln sind, die dauerhaft im Archiv als AIPs gesichert werden

²⁷ Vgl. Kathrin Schroeder, Das "Digitale Archiv". Möglichkeiten für den individuellen Umgang mit Massendaten, in: Mitteilungen aus dem Bundesarchiv. Heft 1/2008, S. 52-57.

²⁸ Im Bundesarchiv wird die Behördenberatung auch im Hinblick auf die Aussonderung von elektronischen Akten durch ein entsprechendes Team aus den Fachreferaten gewährleistet:
www.bundesarchiv.de/service/behoerdenberatung/index.html.

²⁹ Der Vortrag von Dr. Michael Hollmann vom 13.05.2009 betont noch einmal die Attraktivität des Zwischenarchivmodells für die abgebenden Stellen und den Unterschied zum Endarchiv:
www.bundesarchiv.de/imperia/md/content/abteilungen/abtb/bbea/01_hollmann__digzwarch_2009-05-13.pdf.

sollen. Im 4-stufigen Verfahren wird zusätzlich im Vorfeld das Anbieterverzeichnis an das Archiv gesandt, das eingelesen und bestätigt wird, bevor es zur Aussonderung kommt.³⁰

³⁰ Vgl. dazu auch die Präsentationen zum Digitalen Archiv des Bundesarchivs im Rahmen des Informationsforums der Behördenberatung vom 13.05.2009: www.bundesarchiv.de/service/behoerdenberatung/01716/index.html.

Ingest von Fachverfahren im Landesarchiv Baden-Württemberg. Anmerkungen zu Authentizität, Prozessen und Softwareentwicklung

Christian Keitel, Rolf Lang

Wie gelangen digitale Objekte ins Archiv? Während Archivare gerne von "Übernahme" sprechen, kennt der einschlägige ISO-Standard OAIS (ISO-14721) den Terminus "Ingest".³¹ Schon in der Abgrenzung des relevanten Zeitraums offenbaren sich hier erste Unterschiede. Noch grösser sind die Differenzen, wenn man sich bei den Produkten der digitalen Welt selbst umsieht. Angesichts der Vielgestaltigkeit digitaler Formen sollte nach den grossen Objektgruppen unterschieden werden. Einheitlich in einem DMS abgelegte elektronische Akten können nicht nach denselben Abläufen wie Fachverfahren oder Dateisammlungen übernommen werden.

Fachverfahren

In der Landesverwaltung Baden-Württembergs steht – ebenso wie in vielen anderen Archiven und Archivverwaltungen – in den nächsten Jahren vor allem die Übernahme von Fachverfahren an. Dem Landesarchiv Baden-Württemberg stellten sich in diesem Zusammenhang zwei Fragen:

- Wie kann der Übernahmeprozess möglichst standardisiert und zugleich den Eigenarten der Fachverfahren angepasst werden?
- Wie kann der mit einer Übernahme digitaler Unterlagen automatisch verbundenen Gefährdung der Glaubwürdigkeit entgegengewirkt werden?

Fachverfahren basieren in aller Regel auf einer relationalen Datenbank, sie bestehen aus Tabellen, manchmal sind zusätzlich weitere digitale Objekte wie z.B. Bild- oder Textdateien eingebunden. Das IT-System setzt die an verschiedenen Orten jeweils einmal abgelegten Informationen zu einem einheitlichen Bild am Monitor zusammen. Das Bild des einen Benutzers muss dabei dem Bild, das einem anderen Benutzer gegeben wird, nicht entsprechen. Das IT-System ist also konstitutiv für die Datenpräsentation, erst durch seine Zusammenstellung werden die gespeicherten Daten so wiedergegeben, dass der Anwender aus ihnen die gewünschten Informationen gewinnen kann. Derartige Systeme können nach dem derzeitigen Kenntnisstand nicht zuverlässig archiviert werden. Die Daten müssen daher aus den Fachverfahren exportiert und in eine neue Form gebracht werden. Zusatzinformationen (Metadaten) sollen einen Teil der Zusammenhänge herstellen, die bislang von den Fachverfahren selbst geleistet wurden. Die hier kurz skizzierte Migrationsstrategie wird derzeit von nahezu all den Archiven eingesetzt, die bereits Fachverfahren übernommen haben. Das alternativ diskutierte Emulationsverfahren sieht den Erhalt von Daten in ihren ursprünglichen proprietären Formaten vor. Aufgrund verschiedener Unwägbarkeiten und des damit verbundenen Risikos eines Totalverlusts der Informationen wird die Emulationsstrategie derzeit kaum zur Archivierung von Fachverfahren eingesetzt.

³¹ <http://public.ccsds.org/publications/archive/650x0b1.pdf>, alle Abrufe am 4.6.2009.

Prozesse

Die grosse Flexibilität relationaler Datenbanken ermöglicht den Benutzern in den Behörden, sehr unterschiedliche Wünsche an die Systeme heranzutragen. Für die Archive heisst dies im Umkehrschluss, dass fast jedes Fachverfahren eigene Anforderungen an den Export der Daten stellt. Wie können diese wenig normierten Unterlagen dennoch möglichst effizient übernommen werden? Der Königsweg bestünde sicherlich auch bei Fachverfahren darin, gleich bei der Einführung eines Systems dessen Exportschnittstelle programmieren zu lassen. Dieser Weg wurde für die Übernahme elektronischer Akten bereits vom DOMEA-Konzept skizziert. Bei Fachverfahren kann er derzeit aus verschiedenen Gründen nur selten begangen werden:

- Die blossе Zahl der jährlichen Neueinführungen dürfte die personellen Ressourcen der meisten Archive überfordern.
- Den Behörden ist zumeist unbekannt, dass auch Fachverfahren anzubieten sind. Archive werden daher bei der Systemеinführung kaum beteiligt.
- Die Fachverfahren ändern sich nicht selten kontinuierlich. Eine Exportschnittstelle müsste daher über alle Anpassungen hinweg nachgezogen werden. Bei Übernahmezyklen, die in sehr grossen Zeitabständen erfolgen, ist unter Umständen eine Neuprogrammierung der Schnittstelle günstiger als deren regelmässige Anpassung.

Auch das Landesarchiv Baden-Württemberg musste in den letzten Jahren die Erfahrung machen, dass seine Einflussmöglichkeiten bei der Systemеinführung sehr gering sind. In der Regel gestaltete sich daher die spätere Übernahme eines Fachverfahrens als Projekt, das auf die jeweiligen Rahmenbedingungen besonders eingehen musste. Einerseits erschien es sinnvoll, den damit verbundenen erhöhten Aufwand zu betreiben, da Fachverfahren mittlerweile in vielen Behörden einen wesentlichen Anteil bei der Aufgabenerledigung spielen. Auf die Archivierung von Fachverfahren generell zu verzichten hiesse, den archivgesetzlichen Aufgaben nicht nachzukommen. Andererseits müssen trotz allen Schwierigkeiten Wege gefunden werden, wie die Aufwände möglichst weit reduziert werden können. Eine Möglichkeit dazu besteht in der Standardisierung der im Ingest-Prozess enthaltenen Teilprozesse.

Im ISO-Standard OAIS ist der Ingest-Prozess auf der obersten Abstraktionsebene eine von sechs Kern-Entitäten des Digitalen Archivs. Er gliedert diesen in fünf Teilprozesse auf: Die Übernahme selbst, deren Überprüfung, die Herstellung eines Archivierungspakets, die inhaltliche Beschreibung des Objekts (Erschliessung) sowie die Koordination von Updates. Zur weiteren Beschreibung haben die Autoren des OAIS-Standards, also Vertreter der Nationalen Agenturen für Luft- und Raumfahrt, einen weiteren Unterstandard veröffentlicht: PAIMAS steht für Producer-Archive Interface Methodology Abstract Standard (ISO-20652).³² PAIMAS unterscheidet vier Phasen (Vorbereitungsphase, Definitionsphase, Transferphase und Validierungsphase) und benennt innerhalb dieser Phasen 88 unterschiedliche Schritte. Dabei beschreibt PAIMAS nur einen Teil des OAIS-Ingest-Prozesses (die Aufbereitung des Objekts im Archiv wird ausgeblendet). Die Übernahme digitaler Objekte wurde von dem Standard jedoch

³² <http://public.ccsds.org/publications/archive/651x0b1.pdf>.

erstmalig umfassend dargestellt. Allerdings führt das methodische Herangehen zu einer Vielzahl von Redundanzen und damit einhergehend zu einer Schwerfälligkeit, die bei der Übernahme eines konkreten Fachverfahrens zumeist eher hinderlich erscheint. Die Auseinandersetzung mit PAIMAS machte somit zweierlei deutlich: Erstens sollte die Zahl der notwendigen Schritte und mit ihnen die inhärente Komplexität erheblich reduziert werden. Zweitens sollten diese Schritte nicht in einem festen Ablauf verhaftet sein, sie sollten also an die spezifischen Eigenschaften einzelner Fachverfahren anpassbar sein.³³

Authentizität

Neben der Komplexität und Flexibilität der erforderlichen Prozesse stellen Fragen nach der Glaubwürdigkeit der übernommenen Unterlagen die zweite grosse Herausforderung dar, auf die die Archive zu reagieren haben. Derartige Fragen kommen nicht ohne ein Konzept der Authentizität aus. Der von Nestor herausgegebene "Kriterienkatalog vertrauenswürdige digitale Langzeitarchive" definiert Authentizität sehr kurz: "Das Objekt stellt das dar, was es vorgibt darzustellen."³⁴ Etwas konkreter könnte formuliert werden: Ein überlieferter Text ist dann authentisch, wenn er von dem in ihm genannten Autor und zu dem von ihm ausgewiesenen Zeitpunkt und mit den in ihm enthaltenen Inhalten verfasst wurde. Bei Fragen der Glaubwürdigkeit muss also eine Übereinstimmung des ursprünglichen Textes mit dem heutigen Text festgestellt werden. Diese Übereinstimmung muss nur in wesentlichen Punkten vorhanden sein. Beispielsweise wird die Glaubwürdigkeit frühmittelalterlicher Urkunden nicht allein deshalb angezweifelt, weil diese nicht mehr im Original, sondern nur in Abschrift und vielleicht in einer anderen Schriftart vorliegen. Auch wird bei der Frage nach Authentizität nicht danach gefragt, ob der Autor in seinem Manuskript unwahre Sachverhalte als Realität ausgegeben hat. Die Inhalte des Textes müssen also nicht zwingend "wahr" sein. Die hier am Beispiel von Texten entwickelten Überlegungen können auch auf andere analoge und digitale Objekte wie z.B. Ton- und Bildaufzeichnungen übertragen werden.

³³ Neuere Ansätze wie der Nestor-Katalog "Wege ins Archiv" kommen diesen Anforderungen weitgehend nach, Wege ins Archiv – Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv, hrsg. v. der nestor-Arbeitsgruppe Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff, 2009, <http://nbn-resolving.de/urn:nbn:de:0008-2008103009>. Allerdings müssen die standardtypisch abstrakt formulierten einzelnen Punkte noch in die alltäglichen Anforderungen der Archive übersetzt werden. Der vorliegende Text beschreibt einen derartigen Versuch.

³⁴ Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version, hrsg. von der nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung, 2008, <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>.

Die Glaubwürdigkeit konventioneller Unterlagen wird häufig an der Beschaffenheit des Trägermaterials, also an Papier oder Pergament, festgemacht. Bei digitalen Objekten ist dies so nicht möglich. Zunächst existieren derzeit keine dauerhaften digitalen Datenträger.³⁵ Ausserdem kann digitale Information in identischer Form auf andere Datenträger transferiert werden. Glaubwürdigkeitsprüfungen können daher weder in der anbieterpflichtigen Behörde noch im Archiv am einzelnen Datenträger festgemacht werden. Das hier beschriebene Glaubwürdigkeitsproblem stellt sich während des gesamten Erhaltungsprozesses: Auch im Archiv müssen Datenträger und Softwareformate migriert werden. Es wird ein wenig durch den Umstand entschärft, dass das Archiv an den von ihm verwahrten Inhalten keinerlei inhaltliches Interesse hat. Bewusste Fälschungen durch Archivarinnen und Archivare sind folglich unwahrscheinlich, ein Umstand, der juristisch im *ius archivi* bzw. international in der These der *unbroken custody* ausgeführt wurde. Für die Authentizität der Objekte bei der Behörde ist diese selbst zuständig. Innerhalb der archivischen Einflussmöglichkeiten liegt dagegen die Phase der Übernahme digitaler Objekte. Sie stellt einen für die spätere Glaubwürdigkeit besonders problematischen Zeitraum dar. Dabei stellt sich das digitalen Unterlagen ganz allgemein inhärente Glaubwürdigkeitsproblem bei Fachverfahren in besonderem Masse. Wie oben beschrieben werden bei der Migrationsstrategie die Daten ihrer ursprünglichen Präsentationsumgebung (dem Fachverfahren) entnommen. Sie werden also nicht nur physisch auf einen anderen Datenträger, sondern auch logisch in eine andere Umgebung überführt. Sie werden serialisiert, also in einzelne Dateien überführt. Damit ist zu fragen, wie an diesen Dateien Glaubwürdigkeitsuntersuchungen vorgenommen werden können.

Das hier beschriebene Konzept der Authentizität beruht auf dem Gedanken der Übereinstimmung von ursprünglichen und gegenwärtig vorliegenden Angaben. Diese Übereinstimmung kann logisch gesehen durch einen Vergleich vorgenommen werden. Wenn die ursprünglichen und die heutigen Angaben übereinstimmen, sind Letztere authentisch. Im Archiv stellt sich diese Frage erstmals nach erfolgter Übernahme, dann bei der Benutzung, also zu Zeitpunkten, an denen das ursprüngliche Fachverfahren nicht mehr greifbar ist. An die Stelle des ursprünglichen Objekts muss daher etwas anderes treten. Metadaten zum ursprünglichen Objekt können eine solche stellvertretende Funktion ausüben. Die Erhebung geeigneter Metadaten und ihr späterer Vergleich (Validierung) sind daher als zwei wesentliche Teilprozesse des Ingest-Prozesses anzusehen.

³⁵ Eine Ausnahme gilt vielleicht für die Verwendung des Mikrofilms als digitalen Datenträger, d.h. die Ablage der Bytes auf Film (zu unterscheiden von der Ablage eines mit einer Lupe mit dem menschlichen Auge sichtbaren Bilds auf Film). Jedoch müssen die gespeicherten Datei- und Zeichenformate auch hier in regelmässigen Abständen migriert oder emuliert werden. Auch müssen die notwendigen Lesegeräte dauerhaft vorgehalten werden.

Übernahmeverfahren

Wie kann die Übernahme von Fachverfahren so gestaltet werden, dass sie a) möglichst wenig komplex, b) flexibel, c) möglichst standardisiert und d) den Anforderungen der Authentizität Rechnung tragend erfolgt? Jeder Ansatz, sämtliche bei der Übernahme zu beachtenden Punkte aufzunehmen, muss bei Fachverfahren zwangsläufig zu sehr umfangreichen und komplexen Anweisungen führen, die dann in der Praxis nicht mehr umgesetzt werden können. Sinnvoller erscheint es, sich auf die wesentlichen, bei der Übernahme von Fachverfahren unabdingbaren Prozesse zu konzentrieren. Andere Prozesse, wie z.B. das Abklären der Rechtsverhältnisse, unterscheiden sich nicht essentiell von der Übernahme konventioneller Unterlagen. Sie adressieren häufig das in den Archiven vorhandene implizite Wissen, können also wie gewohnt umgesetzt werden. Für die Übernahme von Fachverfahren erscheinen danach nur vier Prozesse unabdingbar:

- **Transfer:** Ingest ist im Kern die Übertragung von einem Datenträger auf einen anderen, vom einen Datei- oder Zeichenformat in ein anderes und von der Behörde zum Archiv. Alle diese Einzelprozesse können als Transfer begriffen werden.
- **Validierung:** Transferierte Informationen müssen auf ihre Authentizität überprüft werden (Vergleich).
- **Bestandsaufnahme:** Grundlage des Vergleichs sind Metadaten, die vor und nach dem Transfer erhoben wurden.
- **Archivische Bewertung:** Am Anfang des Prozesses steht die Frage der Archivwürdigkeit. Zwar stellt sich diese Frage bei allen angebotenen Unterlagen. Bei digitalen Objekten muss zu diesem Zeitpunkt allerdings schon definiert werden, in welchen Aspekten die Informationen unverändert erhalten werden sollen (signifikante Eigenschaften). Zu diesem Zeitpunkt wird also bereits festgelegt, welche Punkte bei einer Bestandsaufnahme ermittelt und bei der Validierung überprüft werden sollen.

Diese Prozesse ereignen sich zumeist in folgender Reihenfolge: Archivische Bewertung – Bestandsaufnahme (1) – Transfer – Bestandsaufnahme (2) – Validierung (Vergleich der Bestandsaufnahmen (1) und (2)). Allerdings sieht das Konzept keine feste Reihenfolge vor. Beispielsweise kann ein Teil der Bewertung auch später nachgeholt werden. Auch können mehrere Bestandsaufnahmen unmittelbar aufeinander folgen, da weitere Metadaten erst nach einer ersten Bestandsaufnahme zugänglich werden.

Transfer und Archivische Bewertung sind zwei Prozesse, die sich einer verallgemeinerbaren informationstechnischen Unterstützung weitgehend entziehen. Mit der Software IngestList versucht das Landesarchiv Baden-Württemberg die anderen beiden Prozesse zu unterstützen. Das Programm wurde vom Landesarchiv selbst entwickelt. Es ermöglicht die Bestandsaufnahme bei Fachverfahren schon in der Behörde. Automatisiert werden dabei zentrale Daten (z.B. übliche Feldtrenner, Datensatztrenner, Formateigenschaften) von den aus den Fachverfahren exportierten Dateien erhoben. Darüber hinaus ist es möglich, zentrale Kenndaten der Fachverfahren in die von IngestList geführte XML-Datei selbst einzutragen. Mit diesen Angaben ermöglicht das Tool die Validierung der Daten unmittelbar nach der Ankunft im Archiv. Damit wird der

Übernahmeprozess insgesamt nachvollziehbar. Demselben Ziel dient auch die Protokollfunktion von IngestList. Bereits in der abgebenden Behörde können zu einem abzugebenden Objekt erste Protokolleinträge angefertigt werden. Diese Einträge bilden nach der Übernahme den Grundstock für das Protokoll, das im Landesarchiv Baden-Württemberg jedem einzelnen digitalen Archivalie mitgegeben wird. Es gibt künftigen Benutzern vom Beginn der Übernahme an Auskunft über den Weg des Objekts. Protokolle, Bestandsaufnahmen und Validierungsergebnisse sind unterschiedliche Gruppen von Metadaten, die auf die eigentlich interessierenden Primärdaten verweisen. Primär- und Metadaten bilden ein Geflecht gegenseitiger Verweise, die kaum umfassend gefälscht werden können. Damit stellen diese Daten ein praktisches Werkzeug zur künftigen Überprüfung der Authentizität und Integrität der Daten dar.

Neben der vom Archiv oder einem künftigen Benutzer durchgeführten Validierung können die bei der Bestandsaufnahme erhobenen Daten aber auch für die digitale Bestandserhaltung verwendet werden. Gerade die durch IngestList ermittelten Formatangaben ermöglichen es, den Zeitpunkt und die Umsetzung einer künftigen Migration zu bestimmen.

Zwar eignet sich IngestList besonders für die Übernahme von Fachverfahren, es kann aber auch für die Übernahme beliebiger anderer Dateitypen und -formate eingesetzt werden. IngestList folgt dem Metadatenkonzept des Landesarchivs Baden-Württemberg. Das Programm ermöglicht den Import in die ebenfalls vom Landesarchiv entwickelte Archivierungssoftware DIMAG.³⁶ Allerdings kann IngestList auch unabhängig von Konzept und Archivierungssoftware angewandt werden. Das Landesarchiv stellt das Werkzeug kostenlos zur Verfügung, um andere Archive bei der digitalen Archivierung zu unterstützen.

Eigenschaften von IngestList

IngestList kann Daten unterhalb eines Verzeichnisses erfassen, charakterisieren, protokollieren und transferieren. Darüber hinaus werden das Formieren und die Datenmigration unterstützt. Parallel zum Verzeichnisnamen mit den Primärdaten wird eine gleichlautende XML-Datei erzeugt. Typischerweise gibt es pro Abgabe viele Primärdateien und genau eine XML-Datei. Liegt das Abgabe-Verzeichnis mit der XML-Datei vor, kann dieses entweder per Datenträger oder auch mittels IngestList transferiert werden. IngestList verwendet hierzu das FTP-Protokoll, um die Daten an einen frei definierbaren Server zu übertragen. So gelingt im Landesarchiv Baden-Württemberg auch der Transfer der Daten nach DIMAG.

Sind die Primärdaten angekommen, wird zunächst die XML-Datei auf Unversehrtheit kontrolliert. Hierzu wird ein md5-Wert über die XML-Datei gebildet und mit dem erwarteten Wert verglichen. Unstimmigkeiten werden erkannt und automatisch protokolliert. Weitere Aktionen wie beispielsweise die Überprüfung auf Vollständigkeit und Unversehrtheit aller Primärdaten, ebenfalls auf Basis von md5-Werten der Primärdaten, führen zu weiteren Protokoll-Einträgen in der XML-Datei. Bisher erfasste Protokolle werden dabei nicht überschrieben, sondern ergänzt.

³⁶ Metadatenkonzept und Ausführungen zu DIMAG s. www.landesarchiv-bw.de unter "Digitale Unterlagen".

IngestList kann auch ohne grafische Oberfläche im Kommandozeilen-Modus verwendet werden. Zum einen kann damit die Erstellung der XML-Datei vollständig automatisiert werden, was aber nur bei gleichförmigen Abgabestrukturen Sinn macht. Zum anderen kann auch genau eine Datei charakterisiert werden. In DIMAG kommt dies für manuell eingestellte Einzeldateien zum Einsatz, um die technische Metadaten zu erfassen.

IngestList verwendet intern openSource-Klassen und -Methoden von DROID³⁷ bereitgestellt von 'The National Archives' und JHove³⁸ von JSTOR und 'Harvard University Library'.

Installation von IngestList

IngestList wurde in JAVA entwickelt und ist damit Plattform-übergreifend verwendbar. Alles, was für den eigenen Einsatz benötigt wird, beschränkt sich auf:

- **IngestList Jar**³⁹, eine Java Archiv-Datei mit allen benötigten Modulen
- Eine aktuelle **Java Runtime**⁴⁰-Umgebung, bereitgestellt unter **Rechner** mit mindestens 512 MB RAM
- **Internet-Zugriff** während des Ingest-Vorgangs und natürlich zum Zeitpunkt der Installation.

IngestList ist im Landesarchiv Baden-Württemberg produktiv im Einsatz auf folgenden Betriebssystemen:

- Windows XP Professional
- Suse 10, eine Linux-Distribution

Getestet wurden die Funktionen von IngestList für weitere Betriebssysteme. Dabei lief das Betriebssystem in einer VM-ware Box unter Windows XP:

- Windows7
- Windows Vista
- Windows 2003
- Ubuntu, eine Linux-Distribution
- SUSE Linux Enterprise Server 64-bit

³⁷ <http://droid.sourceforge.net/wiki/index.php/Introduction>.

³⁸ <http://hul.harvard.edu/jhove/index.html>.

³⁹ <http://sourceforge.net/projects/ingestList>.

⁴⁰ <http://www.java.com/de/download/manual.jsp>.

IngestList ist als openSource frei verfügbar. Für die Nutzung reicht es, die Jar-Datei zu laden. Für die Weiterentwicklung bzw. den Einblick ist die src-Datei gedacht.

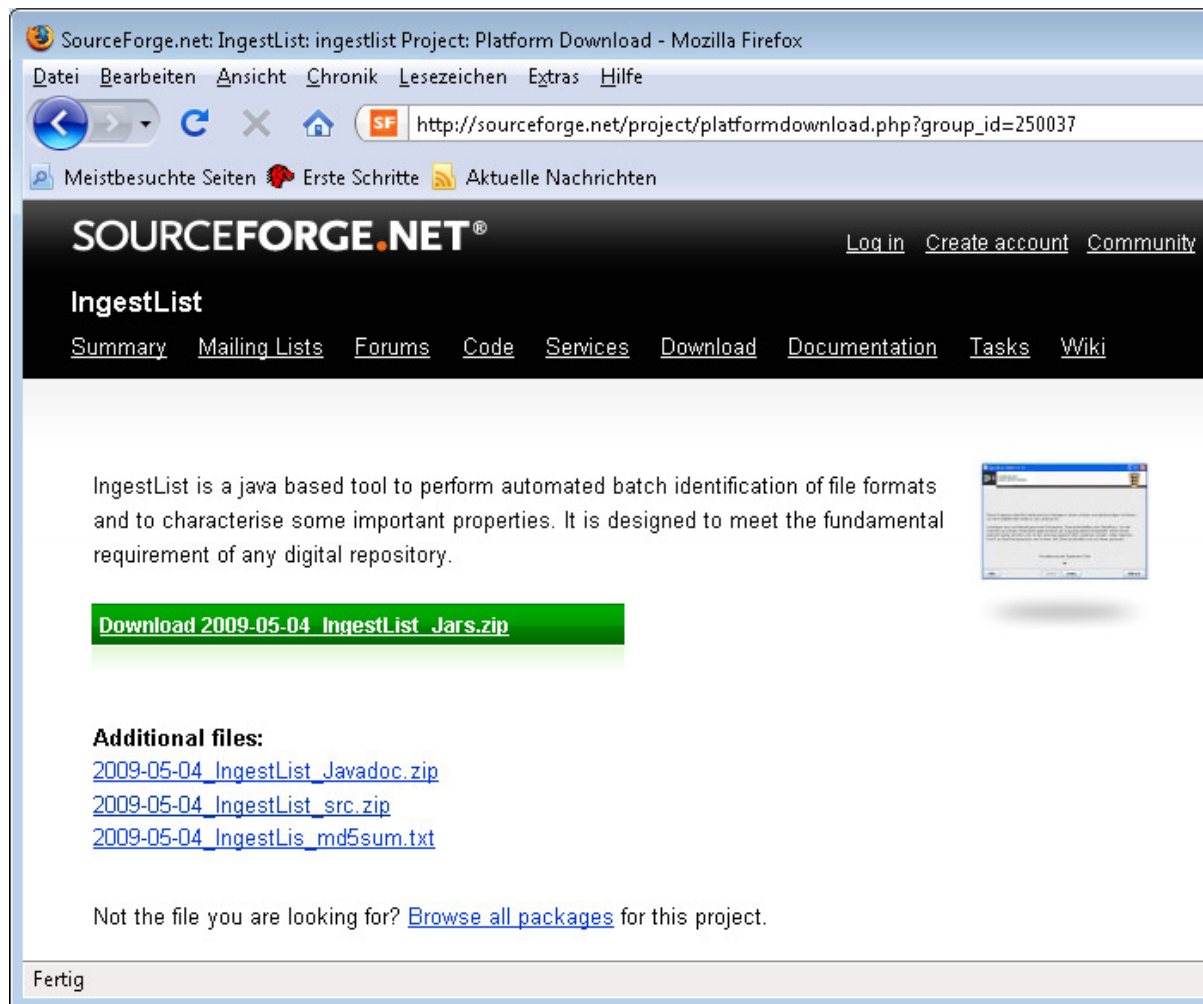


Abbildung 11: IngestList jar-Datei

Die IngestList jar-Datei wird hier auf dem Desktop entpackt, kann aber überall kopiert werden, auch auf einen USB-Stick. Das Tool kann somit sehr einfach auch von der abgebenden Behörde eingesetzt werden.

Bestandsaufnahme

IngestList ist aufgebaut wie ein Assistent, welcher schrittweise den gesamten Prozess begleitet und in der Regel ein Vor- und Zurücknavigieren ermöglicht.

Zuerst werden die ausführende Person und dann die Angaben zur abgebenden Stelle erfasst:

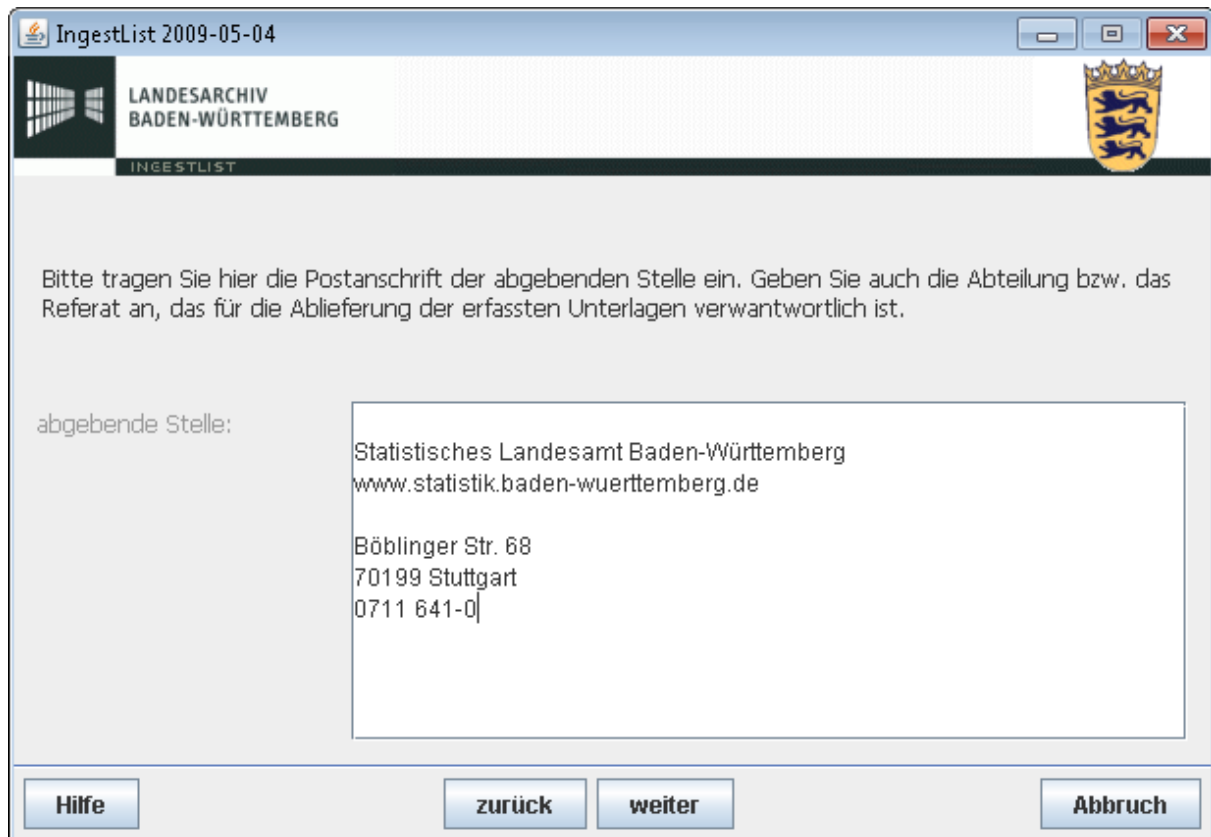


Abbildung 12: Angaben zur abgebenden Stelle

Weitere Eingabefelder sind:

- **Inhalt der Abgabe:** In diesem Textfenster soll ein inhaltlicher Überblick über die abzugebenden Dateien vorgenommen werden. Die Eingaben werden ist zwingend.
- **Quirks:** Sofern vorhanden sollten hier Unzulänglichkeiten, welche schon zum Zeitpunkt der Abgabe aufgefallen sind, notiert werden. Diese Angabe ist optional.

Nun sind die notwendigen Angaben gemacht und das Abgabeverzeichnis wird angewählt. Das Tool liest nun alle darin befindlichen Dateien rekursiv ein und analysiert bzw. charakterisiert diese.

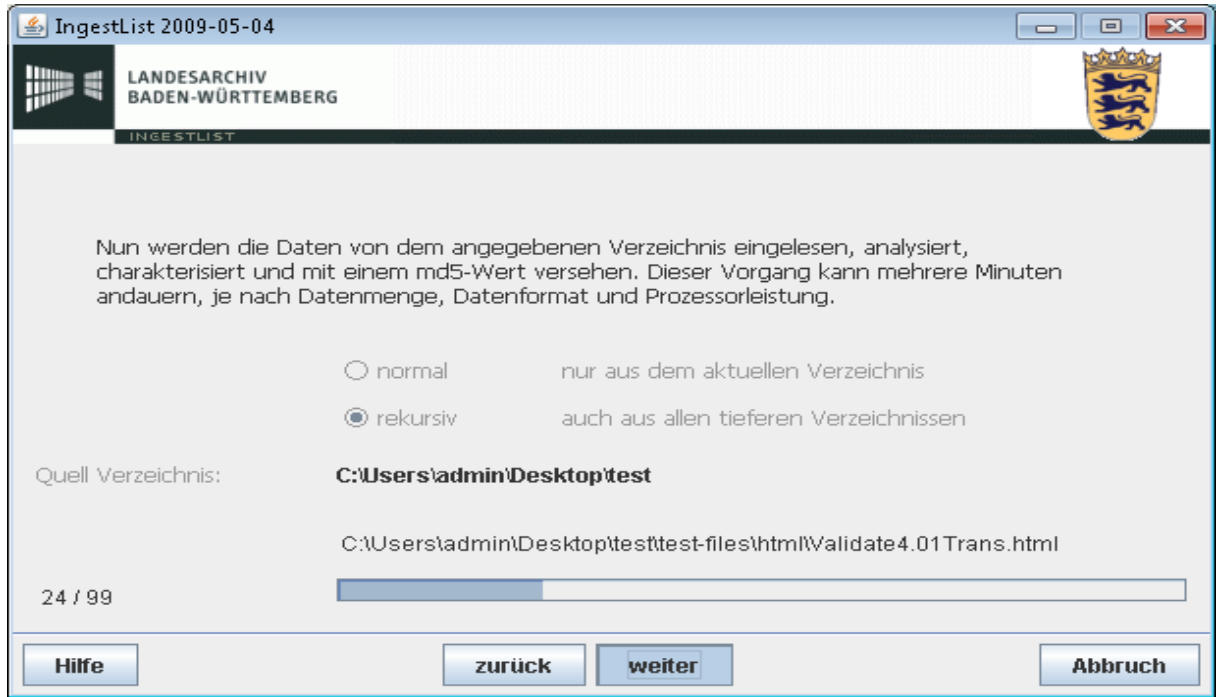


Abbildung 13: Daten einlesen etc.

Das Ergebnis wird in einem Vorschaufenster angezeigt. Darin sind alle Dateien und deren erkannte Metadaten erfasst. Die Anzeige der Metadaten ist gruppiert in reine Anzeigefelder, welche automatisch erfasst wurden und somit nicht mehr veränderbar sind, und in änderbare Felder. Diese können in zwei Bereiche eingeteilt werden:

Feldname		Änderbar?
Beschreibung		
beschreibung	Beschreibung für diese Datei	Immer
quirks	Unzulänglichkeiten	
Datenbank Angaben		
view-tabellen-name	Tabellen-Name	Einmalig beim ersten Erfassen.
anzahl-zeilen	Anzahl der Zeilen	
anzahl-spalten	Anzahl der Spalten	
export-sql	SQL für Export	Später nur lesbar.
describe-sql	Describe-Tabelle	

Die Tabelle ist sortierbar mit einem Mausklick auf die Spaltenüberschrift. Alle Spalten lassen sich in ihrer Grösse variieren und mit drag & drop auch an eine andere Position nach links und rechts bewegen.

name	jhove-format	jhove-version	jhove-wellformed	jhove-valid	jhove-mime
monochrome.jp2	JPEG 2000		TRUE	TRUE	image/jp2
ROItest.jpg	JPEG 2000		TRUE	TRUE	image/jp2
AA_Banner.pdf	PDF	1.2	TRUE	FALSE: Improperly formed date	application/pdf
AA_Banner-single.pdf	PDF	1.3	TRUE	TRUE	application/pdf
bedfordcompressed.pdf	PDF	1.5	TRUE	TRUE	application/pdf
DDAP_Single3.pdf	PDF	1.3	TRUE	FALSE: Outline dictionary missing required entry	application/pdf
DDAP_Spread3.pdf	PDF	1.3	TRUE	FALSE: Outline dictionary missing required entry	application/pdf
fallforum03.pdf	PDF	1.3	TRUE	TRUE	application/pdf
lmd.pdf	PDF	1.2	TRUE	TRUE	application/pdf
6mp_soft.tif	TIFF	6.0	TRUE	TRUE	image/tiff
AA_Banner.tif	TIFF	6.0	TRUE	TRUE	image/tiff
chase-tiff.tif	TIFF		FALSE	FALSE: Count mismatch for tag 34025; expecting ...	image/tiff

Abbildung 14: Tabelle zur Erfassung der Metadaten

Abschliessend wird die gesamte Information in eine XML-Datei exportiert, welche den Namen des Verzeichnisses trägt und parallel zu diesem liegt. Jede Bestandsaufnahme gliedert sich dabei in mindestens drei Bereiche:

- **Entstehung:** Enthält gemachte Angaben zur Entstehung
- **Datei-Liste:** Enthält pro Datei detaillierte Informationen
- **Protokolle:** Diverse automatisch/manuell erstellte Protokolle

XML Structure	XML Code
xml	<code><?xml version="1.0" encoding="utf-8" /></code>
bestandsaufnahme	
entstehung	
datei-liste	
datei	
name	<code>6mp_soft.tif</code>
pfad	<code>\\test-files\6mp_soft.tif</code>
anzahl-bytes	<code>18145024</code>
aenderung-datum	<code>2009-02-11 09:44:00</code>
md5	<code>cf5d73482ad29774974ee4513a4cdd7b</code>
pronom-file-format	<code>Exchangeable Image File Format (Uncompressed) V2.2</code>
pronom-format-id	<code>x-fmt/387</code>
pronom-classification	<code>Positive (Specific Format)</code>
pronom-mime-type	<code>image/tiff</code>
quicks	
beschreibung	
zeichenformat	
anzahl-byte-werte	
anzahl-semicolon-ex	<code>2</code>
anzahl-lf	<code>348835</code>
anzahl-cr	<code>374844</code>
anzahl-tab	<code>326844</code>
anzahl-pipe	<code>47180</code>
anzahl-komma	<code>131837</code>
anzahl-semi-colon	<code>92219</code>
anzahl-single-quote	<code>154267</code>
anzahl-double-quote	<code>185193</code>
kenngrößen	
jhove	
jhove-format	<code>TIFF</code>
jhove-version	<code>6.0</code>
jhove-wellformed	<code>TRUE</code>
jhove-valid	<code>TRUE</code>
jhove-mime	<code>image/tiff</code>
jhove-compression	<code>uncompressed</code>
jhove-image-length	<code>2128</code>
jhove-image-width	<code>2832</code>
jhove-color	<code>RGB</code>

Abbildung 15: Export in XML-Datei

Weitere Informationen finden sich in der Online-Hilfe. Bei Problemen helfen wir gerne weiter.⁴¹

⁴¹ E-Mail-Kontakt: rolf.lang@la-bw.de (Software), christian.keitel@la-bw.de

Von ELO nach FAUST – Konzeptuelle Überlegungen zur Langzeitarchivierung digitaler Überlieferungen und Anwendung im Archiv der sozialen Demokratie der Friedrich-Ebert-Stiftung

Mike Zuchet

Im März 2009 hat der Landesbezirk Niedersachsen der Vereinten Dienstleistungsgewerkschaft (ver.di) dem Archiv der sozialen Demokratie (AdsD) der Friedrich-Ebert-Stiftung (FES) digitale Überlieferungen angeboten, die von der ehemaligen Gewerkschaft Öffentliche Dienste, Transport und Verkehr (ÖTV), Bezirk Weser-Ems stammten. Bei diesen digitalen Überlieferungen handelte es sich sowohl um Metadaten als auch um die dazugehörigen Dokumente, die im Rahmen des DMS ELO office 6.0 entstanden sind. Dabei konnte das AdsD auf zahlreiche Erfahrungen in technischer und struktureller Hinsicht zurückgreifen, die es durch eine enge Kooperation mit dem Bundesvorstand der Gewerkschaft der Polizei sammeln konnte, der ebenfalls mit diesem System arbeitet. Vor diesem Hintergrund hat das AdsD bereits vor einiger Zeit ein detailliertes Konzept entwickelt, um im konkreten Fall die Datenübernahme aus Dokumentenmanagementsystemen zuverlässig und den Transfer in das hiesige Archivspeichersystem FAUST durchführen zu können.

Mit dem Hinterleger wurde vereinbart, dass das Archiv die digitalen Überlieferungen direkt vor Ort in Bremen übernimmt. Dabei übernahm das AdsD den Datentransfer aus dem DMS und die Gewährleistung für die Wahrung der Datenintegrität. Nach der Konfiguration der Exportschnittstelle und dem Export der in Frage kommenden Dokumente auf den hiesigen ver.di-Server wurden von sämtlichen Dokumenten und den dazugehörigen Metadaten-Dateien Prüfsummen generiert, die fortan als Referenz-Prüfsummen dienten. Anschliessend wurden die Metadaten-Dateien und Dokumente – nach dem OAIS-Referenzmodell, dem sogenannten SIP – auf einen mobilen AdsD-Datenträger übertragen. Um Datenverluste und/oder Dateibeschädigungen bei diesem Transfer nachzuhalten und dann den Arbeitsschritt ggf. zu wiederholen, generierte das AdsD erneut von sämtlichen Dateien Prüfsummen (2) und glich diese mit der Referenz-Prüfsumme ab. Der Abgleich wies keine Unterschiede auf, was den Schluss zuliess, dass das SIP weiterhin unverändert vorlag. Der letzte Transfer bestand darin, das SIP vom mobilen AdsD-Datenträger auf den Archivserver zu übertragen. Die oben beschriebene Prozedur wurde wiederholt – Abgleich Prüfsumme (3) mit Referenz-Prüfsumme –, mit dem Ergebnis, dass auch der zweite Transferschritt erfolgreich verlaufen und damit der Bitstream erhalten geblieben ist. Dem OAIS-Sprachgebrauch entsprechend wurden die digitalen Dokumente nach dem Transfer auf den Archivserver (den Archival Storage), zum AIP.

Anschliessend fand im AdsD die Eingangsbearbeitung, der Ingest der digitalen Überlieferungen statt, wobei der Focus zunächst auf die vorliegenden Metadaten-Dateien, die Descriptive Information, gelegt wurde.

Zwecks eines besseren Verständnisses seien vorweg einige Strukturmerkmale des DMS ELO office 6.0 angeführt. Um Strukturen abbilden zu können, bietet das System bis zu drei Gliederungsebenen, die seitens des Anwenders vergeben werden können: Aktenschrank – Aktenordner – Aktenregister. Dabei ist der Anwender nicht gezwungen, sämtliche Gliederungsebenen in dieser Reihenfolge zu verwenden, denn das Dokument ist auch direkt der Gliederungsebene Aktenschrank zuzuordnen.

Ein Umstand, der die Aufbereitung der Metadaten-Dateien für den späteren Transfer in das Archivdatenbanksystem FAUST 6.0 professional erschwerte, war die Tatsache, dass ELO zu jeder Gliederungsebene, in der das Dokument verortet war und zum Dokument selbst jeweils eine Metadaten-Datei generierte. Da die übernommenen Dokumente in der Struktur Aktenschrank – Aktenordner – Aktenregister vorlagen und aus dieser exportiert wurden, gab es zu jedem Dokument vier separate Metadaten-Dateien, die in einer komplizierten und unübersichtlichen Ordnerstruktur abgelegt waren. Nur in den Metadaten selbst, die als ASCII-Dateien vorlagen und daher mit üblichen Texteditoren zu öffnen waren, fanden sich Verweise auf die nächst niedere Gliederungsebene. Im Hinblick auf die derzeitigen Importmöglichkeiten von FAUST boten sich nun zwei Varianten an, wie mit der Aufbereitung der Metadaten-Dateien weiter zu verfahren war: a) hierarchische Zusammenfassung der Metadaten-Dateien des jeweiligen Dokuments zu einer Metadaten-Datei und anschliessend Zusammenfassung sämtlicher Metadaten-Dateien zu einer *Gesamt*-Metadaten-Datei als FAUST-Importdatei, b) jeweilige Zusammenfassung der Metadaten-Dateien einer Gliederungsebene zu einer Metadaten-Datei und Verknüpfung der zusammengehörigen Datensätze im Archivdatenbanksystem FAUST 6.0 professional.

Bei der Auseinandersetzung mit beiden Modellen und dem Abwägen der einsetzbaren Ressourcen *Mensch – Technik – Zeit* fiel die Wahl auf die Variante B. Nachdem die Metadaten-Dateien jeweils einer Gliederungsebene und der einzelnen Dokumente zusammengefasst und in das Strukturformat CSV überführt waren, wurden sie sukzessive über den Importfilter in die FAUST-Datenbank überführt. Dort bildeten sie einzelne Datensätze, die über speziell eingerichtete und konfigurierte Assoziativ-Referenzfelder miteinander verknüpft wurden und sogenannte horizontale Referenzketten bildeten. Am Ende einer jeden Referenzkette befand sich das entsprechende Dokument, das von den Metadaten beschrieben wurde. An dieser Stelle sei anzumerken, dass sich die Zusammenfassung und die dazu notwendige Identifizierung der Metadaten-Dateien, von welcher Gliederungsebene sie stammten, als recht aufwendig erwiesen. Letztendlich war dazu ein umfassender und eingehender Blick in die eigentliche Datenbasis des DMS ELO office 6.0 der übernommenen Überlieferungen notwendig, die eine ausserordentlich hohe Datengranularität aufwies. Nun war es möglich, Angaben in den Metadaten-Dateien zu interpretieren und die entsprechenden Metadaten-Dateien der verschiedenen Gliederungsebenen zu identifizieren und zusammenzufassen.

Wie oben bereits erwähnt, lagen in den Metadaten-Dateien Verweise auf die nächst niedere Gliederungsebene vor, jedoch gab es keine Verweise in die entgegengesetzte Richtung. Um sich jedoch in beide Richtungen entlang der besagten Referenzkette bewegen zu können, waren die FAUST-Datensätze um weitere Metadaten in einem definierten Assoziativ-Referenzfeld anzureichern, die zwecks Unterscheidung intern fortan als Sekundär-Metadaten firmierten. Dieses Verfahren erwies sich ebenfalls als aufwendig und kompliziert, zumal genau darauf geachtet werden musste, keine fehlerhaften Verknüpfungen herbeizuführen. Dieser Umstand wurde bereits bei der ersten archivischen Bewertung der ELO-Metadatenstruktur deutlich. Deshalb ist ein automatisiertes und zuverlässiges Verfahren, das die skizzierte Aufbereitung nach entsprechenden Vorgaben vornimmt, vom Autor entwickelt worden, der im AdsD für die Langzeitarchivierung verantwortlich ist.

Nachdem die Metadaten erfolgreich übernommen, aufbereitet und in das Archivdatenbanksystem des AdsD überführt worden waren, lag der Fokus der weiteren Arbeit auf der Vorbereitung der digitalen Dokumente für die Langzeitarchivierung. Da es für die langfristige Erhaltung digitaler Dokumente erforderlich ist, zu wissen, welches Dateiformat vorliegt und ob es valide ist im Hinblick auf die Format-Spezifikationen, wurden sämtliche Dokumente einer Überprüfung durch das Tool JHOVE unterzogen. Überdies extrahierte JHOVE die technischen Metadaten, die ebenfalls von FAUST aufgenommen werden sollten. Dieser Arbeitsschritt war die Voraussetzung für die zwingend erforderliche Formatkonvertierung der digitalen Dokumente, die ausschliesslich in verschiedenen proprietären Dateiformaten vorlagen. Ziel war es, Letztere in ein Dateiformat zu überführen, das nach dem derzeitigen Stand der Diskussion und Technik auch noch in Zukunft einen Zugriff ermöglicht und sowohl Authentizität und Integrität gewährleistet. Die Wahl fiel dabei auf das Dateiformat PDF/A (1a), in das sämtliche digitalen Dokumente konvertiert wurden. Die Originaldateien wurden ebenfalls aufbewahrt.

Bereits bei der ersten Auseinandersetzung mit dem DMS ELO office 6.0 und der Überprüfung der archivischen Anforderungen an dieses System wurde deutlich, dass das zugrundeliegende Metadaten-Konzept keinen bekannten Standards entsprach und deshalb eine Aufbereitung und Übernahme der Metadaten nur unter hohem zeitlichen und technischen Aufwand durchzuführen wäre. Neben diesem Aspekt machten sich auch weitere Gegebenheiten negativ bemerkbar. Es existiert kein Konzept bei ELO über Aufbau und Struktur der digitalen Akte, wie z.B. das DOMEA-Aktenkonzept, und es gibt kein Modul zur Aussonderung von Überlieferungen in ein DMS-Zwischenarchiv, in dem Bewertungen und ggf. Übernahmen in eine digitale 'Endlagerstätte' stattfinden können. Die Gewährleistung der Authentizität mittels elektronischer Signaturen ist ebenfalls nicht gegeben. Ein Aktenplan, der in jedem DMS für eine sinnvolle archivische Sicherung und Nutzung umgesetzt sein sollte, ist seitens der ÖTV Bezirk Weser-Ems nicht eingerichtet worden, wenngleich die angeführten Gliederungsebenen eine detaillierte Abbildung zulassen. Ebenso können im DMS Verwaltungsvorgänge angelegt werden, die sich auch in

entsprechenden Wertebereichen der dazugehörigen Metadaten niederschlagen und im Rahmen des Ingest-Prozesses ausgelesen werden können. Allerdings hat die hinterlegende Organisation weder vom Aktenplan noch von den Möglichkeiten der Vorgangsabbildung Gebrauch gemacht, so dass diese wichtigen Metainformationen nicht übernommen werden konnten.

Ausgehend von der eingehenden Analyse verschiedener Dokumentenmanagementsysteme und in Anlehnung an die Empfehlungen der Bundeskonferenz der Kommunalarchive beim Deutschen Städtetag hat das AdsD ein archivisches Anforderungsprofil mit folgenden Komponenten entwickelt: Aktenplan, Aussonderung, Digitale Akte, Authentizität, Bewertung, Metadaten, Vorgang und Exportfunktionen. Im Rahmen bereits bestehender Kooperationen mit nahezu sämtlichen deutschen Gewerkschaften ist das AdsD ein gefragter Ansprechpartner bei der Einführung von Dokumentenmanagementsystemen und der Implementierung der genannten archivischen Anforderungen. Ziel ist es, die digitalen Überlieferungen und die dazugehörigen Metadaten bereits bei ihrer Entstehung auf die zukünftige Langzeitarchivierung vorzubereiten.

Wege ins Archiv - Der nestor-Leitfaden für die Informationsübernahme

Jens Ludwig

Hintergrund des Leitfadens

Die Informationsübernahme in ein dLZA⁴² ist ein zentraler Prozess für die langfristige Erhaltung digitaler Daten. Aus mehreren Gründen ist er von besonderer Wichtigkeit. Zum einen ist er sehr kostenintensiv: So wird die Informationsübernahme, wie Beagrie et al 2008⁴³ festgestellt haben, von allen Datenzentren in Grossbritannien als der grösste Kostenfaktor betrachtet. Für das sozialwissenschaftlich orientierte UK Data Archive macht sie 42% aller Kosten aus, der Zugriff hat einen Anteil von 35% und der Bereich "Archival Storage and Preservation" sogar nur 23%.⁴⁴ Zum anderen ist er besonders wichtig für die Qualität. Ein dLZA wird Anfangsfehler selten verbessern können bzw. nur mit einem deutlich höheren Aufwand.⁴⁵

Die Informationsübernahme digitaler Objekte für die Langzeitarchivierung ist zudem besonders schwierig. Digitale Objekte sind voraussetzungsreicher als analoge Materialien. Wenn digitale Objekte in ein dLZA überführt werden, kommen sie aus sehr voraussetzungsreichen und meist unterschiedlichen technischen Umgebungen und sie sollen später in anderen, noch nicht vorhersehbaren technischen Umgebungen verwendbar sein. (Dass auch die ursprünglichen organisatorischen Kontexte und Prozesse zum Verständnis dieser Objekte bekannt sein müssen, klammern wir an dieser Stelle aus, da das prinzipiell auch für analoge Objekte gilt.) Um am Anfang der Archivierung die richtigen Massnahmen zu ergreifen und um zu entscheiden, welche Aspekte eines Objekts erhalten werden sollen und welche Metadaten dafür notwendig sind, müsste man idealerweise in die Zukunft schauen können.

Als Vertreter einer Gedächtnisinstitution wird man häufig gefragt, wie Objekte für die digitale Langzeitarchivierung "abgegeben" werden können. In diesen Nachfragen zeigt sich zwar ein klares Interesse an der Informationsübernahme, aber oft ist den Fragestellenden nicht klar, dass es nicht mit einem einfachen Kopieren der Objekte getan ist, sondern dass das dLZA und

⁴² "dLZA – Digitales Langzeitarchiv" wird dabei im Sinne des nestor Kriterienkatalog vertrauenswürdige digitale Langzeitarchive verstanden: "Eine Organisation (bestehend aus Personen und technischen Systemen), die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit digitaler Daten sowie die Bereitstellung für eine bestimmte Zielgruppe übernommen hat. Dabei bedeutet "Langzeit": "über Veränderungen in der Technik (Soft- und Hardware) und in der Zielgruppe hinweg (z.B. für künftige Generationen, auf unbegrenzte Zeit)." S. 44 in: nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (Version 2). Frankfurt am Main : nestor c/o Deutsche Nationalbibliothek, 2008. URN: urn:nbn:de:0008-2008021802.

⁴³ Beagrie, Neil; Chruszcz, Julia; Lavoie, Brian: Keeping Research Data Safe – A Cost Model and Guidance for UK Universities. JISC 2008. <http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>.

⁴⁴ Beagrie et al 2008, S. 4f.

⁴⁵ Vergleiche die Abschätzung zur nachträglichen Erstellung von Metadaten in: Nationaal Archief, Costs of Digital Preservation, the Hague, 2005, <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf>, Seite 15.

die Produzenten⁴⁶ eine Reihe von Aufgaben gemeinsam lösen müssen. Da die Informationsübernahme wichtig, schwierig und nachgefragt ist, hat die nestor-Arbeitsgruppe "Standards für Metadaten, Transfer von Objekten in digitale Langzeitarchive und Objektzugriff" den Leitfaden "Wege ins Archiv"⁴⁷ für die Informationsübernahme entworfen, der als Einführung und als gemeinsame Arbeitsbasis von dLZA und Produzenten gedacht ist, die eine solche Übernahme planen. In der nestor-Arbeitsgruppe waren hauptsächlich Personen aus Bibliotheken und Archiven vertreten, der Leitfaden richtet sich aber nicht nur an diese Institutionen. Unterstützt wurde die Arbeit unter anderem durch die Initiative "Innovation mit Normen und Standards" des BMWI und DIN.

Bei der Entwicklung des Leitfadens war eine Reihe von Überlegungen besonders wichtig. Eine Entscheidung war, nicht die Terminologie des OAIS-Referenzmodells zu nutzen, da diese abstrakt und voraussetzungsreich ist und dem einführenden Charakter im Weg stehen würde. Da das OAIS-Referenzmodell aber natürlich eines der massgeblichen Dokumente im Bereich der Langzeitarchivierung digitaler Daten ist, wurde darauf geachtet, zu ihm kompatibel zu bleiben. So war auch PAIMAS⁴⁸, ein ergänzender Standard zum OAIS, der die Informationsübernahme detailliert behandelt, einer der Ausgangspunkte der Arbeit der Arbeitsgruppe. Aber neben der OAIS-Terminologie weist er auch eine sehr hohe Granularität der Arbeitsschritte auf, die der Arbeitsgruppe praxisfern schien.

Besonders erwähnenswert ist, dass der Leitfaden unter "Informationsübernahme" nicht das versteht, was das OAIS unter "Ingest" fasst. Der Ingest im OAIS ist mehr oder weniger funktional definiert. Er ist abgeschlossen, wenn die Daten im dLZA an ihrem endgültigen Platz sind. Die Informationsübernahme, die der Leitfaden behandelt, zielt aber auf die Aufgaben, an denen beide Akteure, sowohl das dLZA als auch die Produzenten, beteiligt sind. Ingest und Informationsübernahme enthalten jeweils Aufgaben, die im anderen nicht enthalten sind. Die Informationsübernahme ist primär eine Verantwortungsübernahme, und wenn das dLZA gegenüber dem Produzenten die Verantwortung für ein digitales Objekt übernommen hat, dann dürfte es für den Produzenten wohl meist unerheblich sein, wenn das dLZA intern noch weitere Arbeitsschritte als Teil des Ingest durchzuführen hat. Aspekte, die nur für das dLZA relevant sind, wurden deshalb im Leitfaden weitgehend ausgeklammert. Umgekehrt sind einige, insbesondere administrative Aufgaben der Informationsübernahme, über die aber Informationen notwendig sind, um Produzenten bündig in das Thema einzuführen, im funktional definierten Ingest nicht enthalten.

⁴⁶ Natürlich muss es sich nicht immer um die ursprünglichen Ersteller der digitalen Objekte handeln, die die zu archivierende Information an das dLZA übergeben. Der Terminus Produzent wurde aber aus dem OAIS-Referenzmodell übernommen und steht einfach für den Akteur, der die Daten an das dLZA übergibt. Siehe Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS). CCSDS 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>, Seite 1-12.

⁴⁷ nestor AG Standards: Wege ins Archiv. Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv. nestor 2008. URN: urn:nbn:de:0008-2008103009.

⁴⁸ Consultative Committee for Space Data Systems: Producer Archive Interface Methodology Abstract Standard (PAIMAS). CCSDS 2004. <http://public.ccsds.org/publications/archive/651x0b1.pdf>.

Aufbau des Leitfadens

Es mag verwundern, dass die Aufgaben der Informationsübernahme im Leitfaden nicht in einer Reihenfolge dargestellt werden, in der sie auch in der Praxis behandelt werden müssen. Das liegt darin begründet, dass sich für die Praxis nur wenige Vorgaben für die Reihenfolge machen lassen, da das optimale Vorgehen von vielen Faktoren abhängt. Komplexe Abhängigkeiten machen es üblicherweise notwendig, verschiedene Aufgaben zugleich, im Zusammenhang und wiederholt zu behandeln – wie z.B. die Aspekte der zu archivierenden Informationsmenge, die dafür möglichen und notwendigen Transfermethoden und die sich daraus ergebenden Kosten (vs. vorhandene Mittel) –, bis die für den Einzelfall optimale Lösung gefunden wurde. Statt eine Reihenfolge zu beschreiben, behandelt der Leitfaden die Aufgaben der Informationsübernahme in drei thematischen Blöcken: Aufgaben mit Bezug auf die Objekte, die Prozesse und das Management. Er gibt keine Lösungswege vor, die in der Praxis ohnehin ganz unterschiedlich ausfallen können, sondern versucht, die vom dLZA und den Produzenten zu erreichenden Ziele prägnant zu beschreiben und mit Beispielen zu erläutern.

Objekte

Es gibt drei zentrale Aufgaben, die sich direkt auf die digitalen Objekte beziehen: Die zu archivierenden Informationen müssen ausgewählt, die benötigten Metadaten definiert und die wesentlichen Eigenschaften der Objekte festgehalten werden.

Für die Auswahl der zu archivierenden Information ist wichtig, dass digitale Objekte intellektuelle und technische Entitäten zugleich sind. Beide Aspekte müssen separat untersucht werden. Zuerst: Welche intellektuellen Entitäten sollen archiviert werden? Üblicherweise sollen nicht Dateien als solche archiviert werden, sondern das, was die Dateien repräsentieren: Dokumente, Fotos, Messergebnisse etc. Ausgehend von dieser Auswahl muss dann in einem zweiten Schritt analysiert werden, in welcher technischen Form das Produzentensystem die Objekte liefern kann. Fragen, die hier auftauchen können, sind z.B., ob das Format für die Langzeitarchivierung akzeptabel ist und ob Änderungen des Formats oder der beteiligten technischen Systeme durchgeführt werden müssen.

Was für Metadaten benötigt werden, richtet sich danach, was bekannt sein muss, um die Objekte in einem anderen, späteren Kontext wieder benutzen zu können. Eine Orientierung geben hier die Metadatenunterscheidungen des OAIS und detaillierter der Standard PREMIS⁴⁹ für Langzeitarchivierungsmetadaten. Neben den üblichen deskriptiven Metadaten, die die Objekte inhaltlich beschreiben und zum Auffinden dienen können, sind meist auch technische Metadaten, um die spätere technische Nutzbarkeit zu unterstützen, persistente Identifikatoren und Provenienzmatakten über die Geschichte des Objekts, notwendig. Nicht alle diese Metadaten müssen zwingend vom Produzenten bereitgestellt werden und viele können auch automatisiert

⁴⁹ PREMIS Editorial Committee: PREMIS Data Dictionary for Preservation Metadata (Version 2.0). 2008.
<http://www.loc.gov/premis/v2/premis-2-0.pdf>.

gewonnen werden. Wer aber jeweils für die Bereitstellung der Metadaten verantwortlich ist, ist zu klären.

Für die digitale Langzeitarchivierung ist es unverzichtbar, dass man sich bereits ganz am Anfang klar darüber ist, was eigentlich an den digitalen Objekten erhalten werden soll. Diejenigen Eigenschaften, die über die technischen Veränderungen hinweg unbedingt erhalten werden sollen, werden als "wesentliche Eigenschaften" (engl. significant properties) bezeichnet. Aus der Papierwelt kommend denkt man dabei zuerst häufig nur an das Erscheinungsbild. In der digitalen Welt kann aber z.B. der Textinhalt einer Internetseite von ihrem Erscheinungsbild getrennt sein, sodass man den Textinhalt getrennt vom Erscheinungsbild erhalten kann. Zudem kann es auch um die Erhaltung von interaktiven oder dynamischen Objekten gehen, für die nur das Erhalten des Erscheinungsbild als nicht hinreichend betrachtet wird, wie z.B. für Computerspiele. Allerdings sind Eigenschaften nie an sich wesentlich, was als wesentliche Eigenschaft zu betrachten ist, hängt immer von der Zielgruppe ab. So kann es selbst im Fall der Computerspiele für ein dLZA, deren Zielgruppe nur aus Grafikern besteht, hinreichend sein, nur die Bildschirmgrafiken zu erhalten und auf die wesentlich aufwändiger zu erhaltende Interaktivität zu verzichten. "Alles" zu erhalten, ist üblicherweise aus finanziellen oder technischen Gründen nicht möglich. Wie die Entscheidung auch immer ausfällt, innerhalb der Informationsübernahme muss dokumentiert werden, was als wesentlich betrachtet wird. Eine gängige und nützliche Kategorisierung von wesentlichen Eigenschaften digitaler Objekte stellt u.a. das IN-SPECT-Projekt⁵⁰ vor.

Prozesse

Die drei Aufgaben, die Definition der Transferpakete, der Validierungsschritte und des Transfers, lassen sich als Aufgaben beschreiben, die in den Bereich der Prozesse fallen.

Die Transferpakete bilden die Übersetzungsschicht, die gemeinsame Sprache zwischen dem System des dLZAs und demjenigen des Produzenten, die intern über ihre eigenen Verwaltungs- und Repräsentationsmittel verfügen. Sie müssen deshalb sorgsam überlegt gebildet werden, damit die Informationsobjekte beim Produzenten so in Transferpakete umgewandelt werden, dass sie im dLZA wieder korrekt rekonstruiert werden können. So muss z.B. das Verhältnis zwischen Informationsobjekt und Transferpaketen kein 1:1-Verhältnis sein, sondern grosse Objekte müssen ggf. auf mehrere Transferpakete aufgeteilt werden oder Komponenten, die identisch bei einer Vielzahl von Objekten vorkommen (wie z.B. Formatvorlagen), werden nur einmal übertragen. Auch die Zuordnung der Metadaten zu einzelnen Dateien und die Identifikation der Transferpakete müssen eindeutig sein und festgehalten werden. Es gibt eine Reihe von

⁵⁰ siehe Wilson, Andrew: Significant Properties Report. 2007.

http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf.

sogenannten Paketformaten, die für diese Aufgaben entwickelt wurden.⁵¹ Üblicherweise werden dLZA bereits auf bestimmte Paketformate ausgelegt sein und dafür Vorgaben machen. Für den eigentlichen Transfer sind die Analyse der technischen und organisatorischen Möglichkeiten auf Seiten des Produzenten und des dLZA offensichtliche Aufgaben. Beachtet werden müssen insbesondere die gesetzlichen und vertraglichen Rahmenbedingungen, wie Sicherheitsmassnahmen bzgl. Vertraulichkeit, Verschlüsselung, digitale Signaturen etc. Unterschätzt wird dagegen häufig die Wichtigkeit der Festlegung der Reihenfolge der Transferarbeitsschritte und des Durchführens von Tests. Als Transferweg ist im Zeitalter der Netzwerktechnologie die Online-Übertragung sicherlich die nächstliegende Idee, aber die Offline-Übertragung durch Kurierversand der Datenträger kann eine einfachere, sicherere und schnellere Alternative sein. Es sind auch Fälle möglich, hauptsächlich wenn dLZA und Produzent derselben Institution angehören, in denen überhaupt keine räumliche Übertragung der Informationsobjekte stattfindet, sondern dem dLZA nur die Verantwortung für die Objekte und Infrastruktur übertragen wird.

Nach dem Transfer ist die Informationsübernahme aber noch nicht abgeschlossen. Es müssen Validierungen durchgeführt werden, um sicherzustellen, dass die Übertragung fehlerfrei und wie vorgesehen funktioniert hat. Zu den Tests, die der Produzent und das dLZA vereinbaren sollten, können Tests auf Vollständigkeit, Integrität und Validität gehören. Teil der Vereinbarung eines Tests ist die Festlegung des notwendigen Erfüllungsgrad und der Konsequenzen bei Nichterfüllung, der beteiligten Personen und Hilfsmittel und der zeitliche Abfolge.

Management

Das Management der Informationsübernahme umfasst eine Vielzahl von Themen, von denen die zwei grössten Komplexe Recht und Dokumentation sind. Weitere Teilbereiche des Managements sind Qualität, Sicherheit, Prozessmanagement, Kosten und Risiken.

Eine Reihe von rechtlichen und vertraglichen Fragen muss für die Informationsübernahme geklärt werden, um Planungs- und Rechtssicherheit für alle Akteure zu gewährleisten. Neben eher formalen Aspekten wie der Identifikation der juristischen Personen und Akteure und Fragen der Gewährleistung und Haftung gibt es auch Aspekte mit unmittelbaren Konsequenzen für die Langzeitarchivierung. Welche Verpflichtungen gehen das dLZA und der Produzent bezüglich des zu archivierenden Materials ein? Und welche Rechte existieren an dem zu archivierenden Material? Wenn das Material z.B. unter das Urheberrecht fällt, dann müssen dem dLZA die nötigen Rechte für die angestrebte Form der Archivierung eingeräumt werden, da die Archivierung eine Vervielfältigung ist, die vom Urheberrecht untersagt wird, und eventuell anstehende Migrationen das digitale Objekt verändern.

⁵¹ Die bekannteren Paketformate sind METS (Metadata Encoding and Transmission Standard, <http://www.loc.gov/standards/mets/>), OAI-ORE (Open Archives Initiative Object Reuse and Exchange, <http://www.openarchives.org/ore/>) und MPEG 21 – DIDL (Moving Picture Experts Group 21 - Digital Item Declaration Language, <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>).

Um die Nachvollziehbarkeit der Informationsübernahme zu ermöglichen, sind das Erstellen einer Übernahmevereinbarung, die das Ergebnis aller Planungen festhält, und die Dokumentation der Durchführung, insbesondere von Abweichungen vom normalen Verlauf, unverzichtbar. Sie tragen zur Authentizität und Integrität der Objekte und der Vertrauenswürdigkeit des dLZA bei. Übernahmevereinbarung und Protokoll müssen auch selbst erhalten werden und z.B. durch Provenienzmetadaten mit den übernommenen Objekten verknüpft sein.

Management insgesamt ist natürlich eine allumfassende Aufgabe, die einen Blick für das Ganze behalten muss, die Angemessenheit der Ziele und Mittel sicherstellt und die Umsetzung und Wirksamkeit der Massnahmen kontrolliert. Als letzter Themenkomplex des Leitfadens werden dazu einige Aspekte hervorgehoben wie der Abgleich mit den Qualität- und Sicherheitsstandards der Institutionen, das Management der Prozesse und die Integration in die Arbeitsabläufe (Wann und an welcher Stelle erfolgt LZA?) sowie die Kosten und Risiken. Zu den Kosten und Risiken sei noch angemerkt, dass es inzwischen eine ganze Reihe von generischen Modellen zur Kostenabschätzung gibt,⁵² die aber jeweils mit den konkreten Situationen und Zahlen abzugleichen sind. Viele Personen werden von den Kosten der Informationsübernahme abgeschreckt, die wie bereits erwähnt schon einen grossen Prozentsatz der Gesamtkosten der Langzeitarchivierung ausmachen. Um die Kosten und Konsequenzen der Durchführung in eine angemessene Perspektive zu rücken, ist es oft sinnvoll, sie mit den Konsequenzen (und Kosten) der Nichtdurchführung und den ursprünglichen Produktionskosten zu vergleichen.

Weiterentwicklungen

Der Leitfaden zur Informationsübernahme wurde bisher auf einigen Veranstaltungen vorgestellt und hat ein reges Interesse erfahren. Die bisherigen Rückmeldungen zeigen, dass der Umfang des Leitfadens gut gewählt und für die Praxis handhabbar scheint. Gewünschte Weiterentwicklungen und Ergänzungen bezogen sich u.a. darauf, die Perspektive anderer Anwendungsgruppen einzubringen (Museen, aber evtl. auch Industrie und Wissenschaft) und mehr Beispiele anzuführen. Ein Faktor für die erfolgreiche Informationsübernahme, der im Leitfaden nicht erwähnt, aber bei Übungen mit ihm deutlich wurde, ist, dass die beteiligten Personen, die Teams auf Seiten des dLZA und des Produzenten, wichtig sind. Sie müssen einerseits die inhaltlichen, andererseits aber auch die technischen und rechtlichen Fachkenntnisse besitzen, aber auch ein Bewusstsein für die gemeinsame Aufgabe und ihre Relevanz haben. Um diese und andere im Leitfaden noch nicht angesprochenen Themen besser auszuarbeiten, sollten idealerweise eine Reihe von Umsetzungsstudien durchgeführt werden, die als Sammlung den Leitfaden ergänzen. Ein weiteres Resultat dieser Arbeiten könnten Vorlagen für Übernahmevereinbarungen für idealtypische Szenarien sein.

⁵² Z.B. Ayris, P.; Davies, R.; McLeod, R.; Miao, R.; Shenton, H.; Wheatley, P.: The Life final project report. Research report. LIFE Project, London, UK 2008. <http://eprints.ucl.ac.uk/11758/1/11758.pdf> oder Beagrie, N.; Chruszcz, J.; Lavoie, B.: Keeping research data safe: a cost model and guidance for UK universities. Final Report, April 2008. HEFCE 2008. <http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf> oder Nationaal Archief, Costs of Digital Preservation, the Hague, 2005, <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf>.

Nestor ist als Projekt ausgelaufen und existiert nun als Kooperationsverbund von Gedächtnisinstitutionen. Die Fortsetzung der Arbeiten am Leitfaden für Informationsübernahme wird aber weiterhin im nestor-Umfeld und im Rahmen des DIN Normungsausschusses "Schriftgutverwaltung und Langzeitverfügbarkeit digitaler Informationsobjekte" erfolgen.

Handreichungen zur Webarchivierung des Arbeitskreises Dokumentation und Archivierung von Webpräsenzen

Rudolf Schmitz

Da die Handreichungen erst Ende des Jahres erscheinen werden, gestatten Sie mir zunächst einen Hinweis auf ein Buch, das den Vorteil hat, bereits erschienen zu sein, und den Titel trägt "The WWW as a challenge and as a chance for parliamentary and party archives". In diesem Buch wird ausführlich auf die Webarchivierung Bezug genommen. Daneben finden sich aber auch Beiträge zu anderen Aspekten der Nutzung des Internets durch Archive etwa als Präsentationsmedium für Sammlungen. Die Beiträge des Bandes sind zum Teil in englischer und zum Teil in deutscher Sprache verfasst. Im Anhang finden sich elf Thesen zur Webarchivierung, die auf der SPP-Tagung 2006 in Bonn verabschiedet wurden. Das Buch kann kostenlos über das AdsD bezogen werden und ist auch als PDF-Datei im Internet verfügbar⁵³.

Soweit ich sehe, sind zurzeit drei Archivverbünde bzw. -arbeitskreise mit der Arbeit an Handreichungen zur Webarchivierung befasst: zunächst

- die Bundeskonferenz der Kommunalarchive, in deren Arbeit Frau Dr. Becker auf der Tagung in Halle einen eindrucksvollen Einblick gegeben hat. Dann
- der Arbeitskreis "Elektronische Archivierung" in der Vereinigung deutscher Wirtschaftsarchivare, die sich besonders dem Problem einer Bewertungsmatrix für die unterschiedlichen Erfassungssoftwares widmet.⁵⁴ Sowie
- der Arbeitskreis "Dokumentation und Archivierung von Webpräsenzen" in der AWW, dessen Handreichungen den gesamten Workflow umfassen sollen.

Wir haben uns bei unserer Arbeit an den Richtlinien von zwei Grundsätzen leiten lassen.

Grundsatz 1: Die Handreichungen richten sich an Archivare und nicht vornehmlich an IT-Spezialisten. Das heisst einmal, dass sie – wo immer es geht – auf traditionelle archivische Verfahrenweisen rekurren, auch da, wo man glaubt, Verfahren modifizieren zu müssen (Bewertung). Das gilt auch für die verwendete Terminologie. Das heisst aber auch, dass man unterschiedliche Verfahren etwa der Erschliessung und der Bereitstellung nicht in einen unnötigen Gegensatz zu einander bringt, sondern Vorzüge und Nachteile gegeneinander abwägt.

Grundsatz 2: Die Handreichungen sollen dazu dienen, besonders kleine und mittlere Archive zu ermutigen mit der Webarchivierung zu beginnen – ich erinnere noch gut die ironischen Formen, mit denen auf der Stuttgarter Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" die kleineren Archive zur Berücksichtigung ihrer Belange aufgefordert

⁵³ URL: <http://www.fes.de/archiv/spiegelung/default.htm>,

http://www.fes.de/archiv/spiegelung/links/links_literatur.htm, <http://library.fes.de/pdf-files/adsd/05657.pdf>

⁵⁴ URL: <http://www.wirtschaftsarchiv.de/akea/webseiten.htm>

haben. Das heisst natürlich auch, Standards, die unverzichtbar sind, nicht so zu definieren, dass ihnen nur noch wenige Grossorganisationen Rechnung tragen können.

Bei der Erläuterung des Workflows möchte mich auf die sogenannten Grundsätze der Webarchivierung konzentrieren, die in der unteren linken Hälfte aufgeführt werden, sowie auf die Beschreibung der Arbeitsschritte der Erfassung.

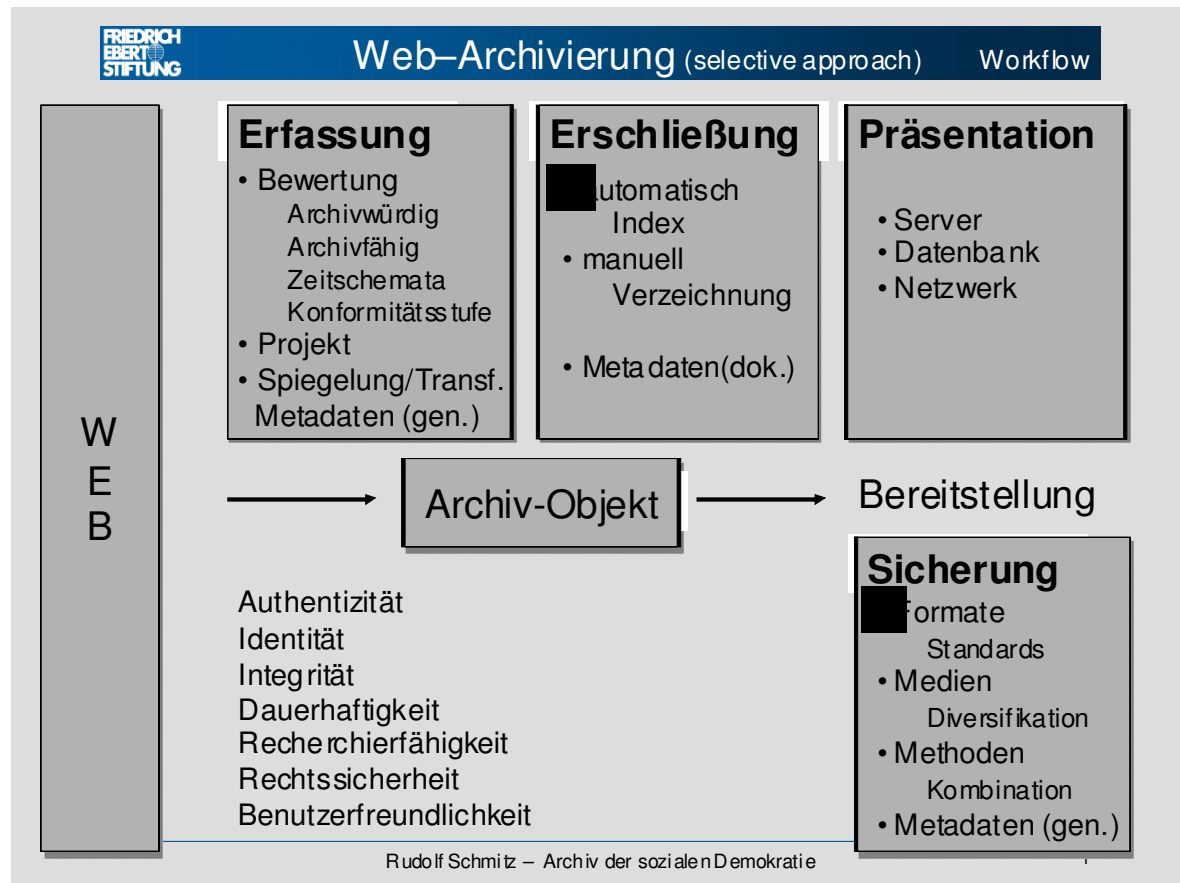


Abbildung 16: Übersicht Web-Archivierung

Zunächst jedoch einige Worte zum Web als einer neuen Quellengattung.

In immer stärkerem Masse räumen öffentliche und private Organisationen ihren Internetpräsenzen eine zentrale Stellung sowohl bei der Organisation ihrer Kommunikation als auch bei der Darstellung ihrer Inhalte und Personen ein. Planmässig werden die neuen Möglichkeiten der Informationstechnologie in Überlegungen zur Transparenz von Strukturen und zur Konzeption der Arbeit mit einbezogen. Dabei werden konventionelle Formen der Darstellung und Kommunikation immer stärker durch Internetangebote ergänzt oder gar ersetzt.

Damit sind die einschlägigen Archive vor die nicht übertragbare Aufgabe gestellt, auch diese neue Quellengattung als wichtigen Bestandteil des kulturellen Erbes langfristig zu sichern und einer interessierten Öffentlichkeit zur Verfügung zu stellen.

Das World Wide Web zeichnet sich durch die Flüchtigkeit seiner Inhalte und eine grosse Variabilität seiner Strukturen aus. Websites stellen sich nach einem Relaunch völlig neu dar, Dokumente werden verändert, ersetzt oder an eine andere Stelle verschoben. URLs können sich ebenso ändern wie die Darstellungsweise von Texten und Grafiken. Immer neue Strategien,

wie z.B. die Einbindung der Dienste von Video-Portalen, werden für die Präsentation eigener Inhalte in Anspruch genommen. Dies alles macht das Web zu einer komplexen Quellengattung, die eine ganz besondere Herausforderung für die Archivierung darstellt.

Da die Webarchivierung in einem vollständig von der Informationstechnologie geprägten Umfeld stattfindet, die ihrerseits mit einem auf den Aspekt der Sicherung verkürzten Archivierungsbegriff arbeitet, muss an dieser Stelle noch einmal ausdrücklich betont werden, dass der Aufbau eines Internet-Archivs neben der Sicherung auch die Erfassung, Erschliessung und Präsentation von Webseiten erfordert. Die Datenbestände müssen physisch im Archiv präsent sein, sie müssen recherchierfähig sein, langfristig gesichert und benutzerfreundlich aufbereitet. Die Archivierung von Websites erfolgt in der Regel nicht, um bestimmte Inhalte (content) des Internets zu sichern, sondern um definierte Websites (selektiver Approach) unter Wahrung ihrer Strukturen und Funktionalitäten in einer browserfähigen Form zu archivieren. Das bloße Herunterladen einzelner Dokumente (z.B. PDF-Dateien) stellt einen Grenzfall der Webarchivierung nicht deren Normalfall dar. Der Darstellungszusammenhang muss gewahrt werden, zumindest aber rekonstruierbar sein.

Ziel der Webarchivierung ist die Generierung und Bereitstellung aussagekräftiger Archivobjekte in einer angemessen dichten Überlieferung und in lesbarer Form.

Unsere Handreichungen orientieren sich am Normalfall der selektiven Spiegelung von Webpräsenzen Dritter durch einen Off Line Browser. Bei diesem Verfahren werden keine anderen Wege und Methoden benutzt als diejenigen, die auch dem normalen Internetnutzer zur Verfügung stehen. Das Spiegelungsverfahren ist gebunden an die Linkstruktur des Webs und findet seine Grenzen etwa bei Datenbanken, die mit diesem Verfahren nicht zu erfassen sind.⁵⁵

Die Spiegelung der Websites erfolgt in diskreten Schritten. Dabei werden servergestützte Formate umgewandelt und die absoluten Links in relative umgeschrieben. Die gespiegelten Dateien werden zunächst in einem einheitlichen Verzeichnis auf einer Festplatte gespeichert.

Der von uns verwendete "Off Line Explorer" liefert immer noch die besten Ergebnisse, zeichnet sich durch einen grossen Bedienungskomfort aus, der auch Eingriffe während des Spiegelungsprozesses zulässt. Er erlaubt die Verwendung von Macros und URL Substitutes, etwa um die Mehrfachspiegelungen von Dateien mit verschiedenen Session Ids zu vermeiden, sowie die Verwendung kleinerer Scripte, um Spiegelungen vorzuprogrammieren und zu einem späteren Zeitpunkt oder in festen Intervallen automatisch durchführen zu lassen. Da keine proprietären Formate erzeugt werden, lassen sich die Spiegelungsergebnisse auch unterschiedlicher Off Line Browser miteinander kombinieren. Das gilt auch für die unter Umständen notwendige Ergänzung von Spiegelungen durch Streaming Files, die nur mit einer speziellen Software, z.B. dem RM-Recorder, erfasst werden können. (Eine Beta-Version des Off Line Browsers, der eine vorbereitende Analyse der zu archivierenden Websites ermöglicht, wird zurzeit im AdsD getestet.)

⁵⁵ Andere Methoden und Aufgabenstellungen (comprehensive approach / direct transfer / Datentransfer über FTP) werden nur am Rande erwähnt.

Wie bereits erwähnt, sind die Ergebnisse eines eingehenden Vergleichs verschiedener Off Line Browser auf den Seiten des Arbeitskreises "Elektronische Archivierung" des VdW zu finden. Hier sei nur angemerkt, dass man wegen der unterschiedlichen, teilweise konventionswidrigen Programmierung von HTML-Seiten auch sehr unterschiedliche Erfolge mit verschiedenen Off Line Browsern erzielen kann. Es ist also unter Umständen nicht der optimale, sondern der für bestimmte Seiten geeignetste Browser zu suchen.

Selektion

Die Auswahl der zu archivierenden Seiten ergibt sich aus der Zuständigkeit bzw. dem Archivierungsauftrag des entsprechenden Archivs. Der ausgewählte Bereich kann mehrere Domains, Server, URLs umfassen. Es gibt also keine technische Grösse, auf die man sich bei der Definition von zu archivierenden Webbereichen positiv beziehen könnte. Auf die Einheit dieses Bereichs kann man nicht referieren, man muss sie zuallererst herstellen. Dabei kann es sich als durchaus sinnvoll erweisen, die Übernahme von Websites in unterschiedlichen Projekten zu organisieren, etwa um die Zeit, die der Datentransfer beansprucht, zu minimieren. Ein längerer Zeitraum etwa von mehreren Wochen birgt immer die Gefahr in sich, Dokumente als Teil eines Archivobjekts zu präsentieren, die nie gleichzeitig im Internet standen.

Wie sich eine Site darstellt (manifestiert), hängt unter Umständen auch vom Zeitpunkt – und gegebenenfalls vom Ort – des Aufrufs ab sowie von den Parametern des benutzten Browsers. Das gilt natürlich auch für die Erfassung von Websites durch einen Off Line Browser.

Grundsätze der Webarchivierung

In allen Bereichen der Erfassung, Erschliessung, Bereitstellung und Langzeitsicherung dienen die zu ergreifenden Massnahmen der Feststellung und Erhaltung der Authentizität, Identität, Integrität, Dauerhaftigkeit, Recherchierfähigkeit, Benutzerfreundlichkeit und Rechtssicherheit.

Authentizität

Authentizität bedeutet einmal die Feststellung der Evidenz und zusätzlich die Wahrung der Strukturen und Funktionalitäten der Websites, die möglichst in einer browserfähigen Form archiviert werden sollen (sog. "Look and Feel"). Da der Entstehungszusammenhang in der Regel nicht mit erfasst werden kann, wird die Erfassung des Darstellungszusammenhangs von Dokumenten in Websites umso wichtiger.

Die Überprüfung übernommenen Dateien ist abhängig von den Validierungsroutinen der jeweiligen Software. Sie sollte den HTTP-Status überprüfen, Authentifizierungsverfahren unterstützen und die Angaben im Header der jeweiligen Datei zur Überprüfung heranziehen. Da aber z.B. die Angaben zur Dateigrösse keine Pflichteinträge sind oder nur unvollständig vorgenommen werden, sollten ergänzende Verfahren für den Abgleich der Dateigrössen verfügbar sein. Das alles muss automatisiert von dem entsprechenden Off Line Browser während des Datentransfers vorgenommen werden. Eine nachträgliche Validierung der übernommenen Daten ist in der Regel nicht möglich.

Zur Wahrung der Authentizität gehört natürlich auch die Dokumentation der Daten, die zur Steuerung des Datentransfers eingegeben oder – etwa als Fehlermeldungen – generiert wurden.

Identität

Zur Wahrung der Identität sind Angaben zur Grösse, zur Anzahl der Dateien, zum Speicherort sowie eine Signatur erforderlich. Diese Angaben können durch Hashwerte und digitale Signaturen ergänzt werden.

Integrität

Die Integrität erfordert Massnahmen zum Ausschluss aller nichtdokumentierten Veränderungen. Alle Veränderungen des Archivobjekts, die das Archiv vornimmt, müssen dokumentiert werden. Alle willkürlichen Veränderungen durch den Benutzer müssen ebenso ausgeschlossen werden wie alle unwillkürlichen Veränderungen, die sich durch lange Speicherzeiträume ergeben können (Refreshing).

Dauerhaftigkeit

Im digitalen Bereich müssen besondere Massnahmen (Diversifikation der Speichermedien, Standardisierung der Formate, Migrationsstrategien) ergriffen werden, um eine langfristige Speicherung und Benutzbarkeit der Archivobjekte zu gewährleisten.

Bei der Archivierung von Websites steht neben dem Erhalt des Content die Bewahrung der Funktionalität im Vordergrund der Langzeitsicherung. Die Konversion in ein Archivformat kann also nur ergänzenden Charakter haben, da sie in der Regel mit dem Verlust der Funktionalität erkaufte würde. Ins Zentrum der Bemühungen muss also der Erhalt der Präsentationsformate gestellt werden. Zu den Besonderheiten der Archivierung von Webseiten Dritter gehört, dass nicht nur die grosse Zahl der unterschiedlichen Datei-Formate, die im Web Verwendung finden, von den Archiven nicht zu beeinflussen ist, sondern auch deren unterschiedliche, manchmal eigenwillige Programmierung. Dabei hat sich gezeigt, dass die Vielzahl der Anbieter mit den unterschiedlichsten Formen konventionswidriger Programmierungen der einzelnen Seiten eine automatische Konvertierung der gespiegelten Seiten in XHTML nicht zulässt.

Diversifikation der Speichermedien

- Sicherung auf einer Festplatte mit einem Raid-System
- Sicherung auf Bändern
- Externe Sicherung auf CD oder DVD

Anzustreben ist eine möglichst grosse Diversifikation der eingesetzten Medien zur Langzeitsicherung.

Standardisierung der Formate

- Da es keine verbindlichen Standards gibt, wird man sich im Fall der Ersetzung eines Formats an der Häufigkeit, mit der bestimmte Formate verwendet werden, orientieren müssen. Grundsätzlich ist darauf zu achten, dass bei Konversionen auch die entsprechenden Links umgeschrieben werden müssen.

Methoden

- Pflege eines Browserarchivs
- Dokumentation der Massnahmen zur Langzeitsicherung in den entsprechenden Metadaten
- Protokollierung der Umgebungsdaten
- Entwicklung von Erhaltungsstrategien als einer Kombination von unterschiedlichen Verfahren: Migration, Emulation, Ersetzungen mit Hilfe regulärer Ausdrücke.
- Zusätzliche Sicherung in komprimierter Form. (Bei der Komprimierung mit WinZip müssen bestimmte Parameter beachtet werden, um die Struktur der Dateien zu erhalten.)

Recherchierfähigkeit

Die Archivobjekte müssen durch geeignete Massnahmen erschlossen werden. Dazu gehören neben der Klassifikation die Verzeichnung, die Indexierung sowie die Dokumentation der Metadaten, die neben einer sichernden auch eine erschliessende Funktion haben können. Da die Erfassung in diskreten Schritten organisiert ist (Snapshots), sollte über die Indizes auch die Möglichkeit einer diachronen bzw. synchronen Suche geschaffen werden.

Indexierung:

Die automatische Volltextindexierung bietet neben der Verzeichnung und den Metadaten eine wertvolle Ergänzung zu den Rechercheoptionen der übrigen Erschliessungsformen. (Versuche, semantische Analyseverfahren in die Indexierung mit einzubeziehen, sind bisher leider gescheitert, weil die Handhabung der entsprechenden Software sich äusserst umständlich gestaltete und die Einbeziehung binärer Formate nicht gewährleistet werden konnte.) Die Indexierungssoftware DtSearch bietet die Möglichkeit unterschiedliche Indizes in eine Webform einzubinden, so dass einzelne Indizes bei der Recherche frei miteinander kombiniert werden können.

Eine Indexierungssoftware sollte folgenden Anforderungen entsprechen:

1. Freie Indexierung
2. Thesaurus (optional)
3. Verarbeitung einer Datenmenge von mindestens 10 GB
4. Sprachmodul der deutschen Sprache/Stammformensuche
5. Boolesche Operatoren/Trunkierungen
6. Webform
7. Gewichtete Anzeigen
8. Highlighting

9. Browserfähige Ergebnisseiten
10. Ermöglichung von diachroner und synchroner Suche

Dokumentation der Metadaten

Die Sicherung der Authentizität und Identität des archivierten Materials erfolgt über eine Dokumentation der im Prozess der Erfassung und Langzeitsicherung anfallenden Metadaten.

Ein Minimalset der Metadaten umfasst die Erfassungsdaten (Steuerungsdaten und Speicherdaten), die Erschliessungsdaten, die Evidenzdaten und die Sicherungsdaten. Diese Daten sollten in XML abgelegt werden.

Steuerungsdaten (Erfassungsdaten, Authentizität):

OffLine-Browser (Typ, Version), Programmeinstellungen⁵⁶, aufgenommene URLs, URL der Startseite, Fehler beim Spiegeln, Gebrochene Links, Datum der Spiegelung, (Nachbearbeitungen), (Abbruch der Spiegelung), Umgebungsdaten⁵⁷

Speicherdaten (Erfassungsdaten, Identität):

Umfang des Projekts, Anzahl der Dateien, Speicherverzeichnis, Projektname/Signatur

Erschliessungsdaten:

Seiteninformation (Metatags), Seiten-, Dateiinformationen des Servers

Evidenzdaten:

Anbieterdaten (Denic), Benutzerdaten

Sicherungsdaten:

Formate, Medien, Methoden

Benutzerfreundlichkeit

Die Benutzbarkeit schliesst die Zugänglichkeit, Handhabbarkeit und Lesbarkeit ein. Dem Benutzer muss über eine entsprechende Infrastruktur der Bereitstellung der Zugang zu den Archivobjekten ermöglicht werden, diese müssen mit normalen IT-Kenntnissen benutzbar sein und zusammen mit der einschlägigen Software bereitgestellt werden, die ihre Lesbarkeit garantiert. Daher sollte mit Beginn der Webarchivierung – wie bereits erwähnt – auch ein entsprechendes Browserarchiv angelegt und gepflegt werden.

Rechtssicherheit

⁵⁶ Abweichungen von definierten Standards.

⁵⁷ Abweichungen von definierten Standards.

Neben den vertraglichen Regelungen, die die jeweiligen Archive mit den Anbietern von Websites zu treffen haben, bleibt es eine Aufgabe im Rahmen einer weiteren Klärung des Urheberrechts darauf hinzuwirken, dass die Archive sowohl die Archivierung als auch die Bereitstellung digitaler Dokumente mit der notwendigen Rechtssicherheit betreiben können.

Arbeitsschritte

Die Erfassung von Webpräsenzen gliedert sich in folgende Arbeitsschritte:

Auswahl

Zusammenstellung der einschlägigen URLs (Seiten der satzungsgemässen Gliederungen, Gremien und Initiativen einer Organisation); Festlegung, in welchen Fällen auch darüber hinausgegangen werden kann.

Bewertung

Feststellung der archivwürdigen Teile des ausgewählten Bereichs; Festlegung von Intervallen und Anlässen der Spiegelung; Bestimmung der Konformitätsstufen (text-, content-, darstellungs-, struktur-, kontext- oder funktionsorientierte Spiegelungen).

Analyse

Wahl der Form des Datentransfers; Besonderheiten der Seiten (Weiterleitungen, Flash, dynamische Seiten).

Spiegelung

Organisation und Durchführung der Erfassung. Eingabe der URLs mit den entsprechenden Parametern. Festlegung des Speicherverzeichnisses.

Ergebniskontrolle

Überprüfung der Resultate; Kontrolle, ob das angestrebte Ergebnis erreicht wurde. Eventuelle Korrekturen.

Metadaten

Dokumentation der Parameter, mit denen der Off Line Browser den Spiegelungsprozess steuert.

Datenbank

Welche Möglichkeiten die Einbeziehung einer relationalen Datenbank in den Prozess der Webarchivierung bietet, soll abschliessend mit einigen Screenshots (neue und alte Version von OWA) verdeutlicht werden.

Eine wichtige Hilfe zur Entwicklung von Methoden der Langzeitarchivierung bietet die Möglichkeit, die vorhandenen Mime-Typs mit den entsprechenden Subtyps innerhalb einer Spiegelung anzuzeigen.

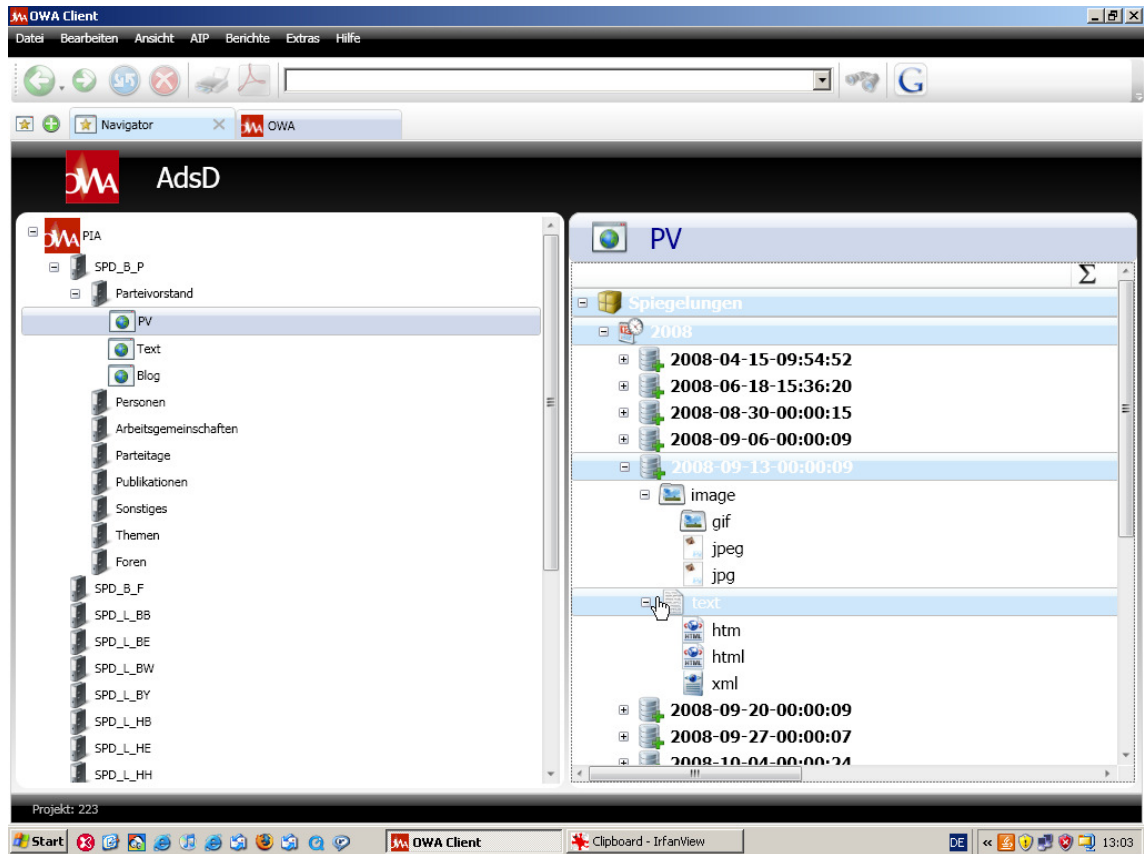


Abbildung 17: Spiegelung Mime-Typs und Subtyps

Man kann mit den entsprechenden URL-Angaben auch eine Detailansicht eines bestimmten Dateityps aufrufen.

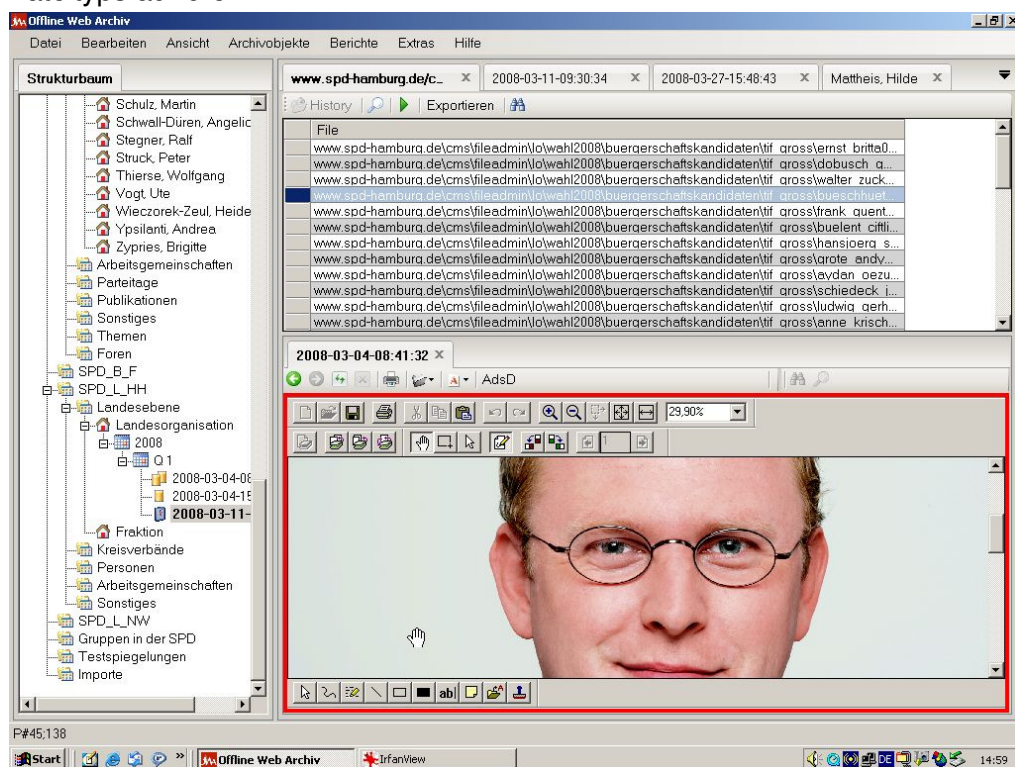


Abbildung 18: Detailansicht

Eine völlig neue Strategie der Webarchivierung eröffnet sich durch die Möglichkeit, die für die Darstellung notwendigen Umwandlungen nicht am Archiv-Objekt selbst vorzunehmen, sondern nur im und für den Prozess der jeweiligen Präsentation. Die Ersetzung mit Hilfe regulärer Ausdrücke und die Umwandlung von Dateien ohne Extension ausschliesslich für die Darstellung liesse sich natürlich auch für die Umwandlung absoluter Links anwenden, die dann im Quelltext des Originals nicht mehr umgeschrieben zu werden bräuchten.

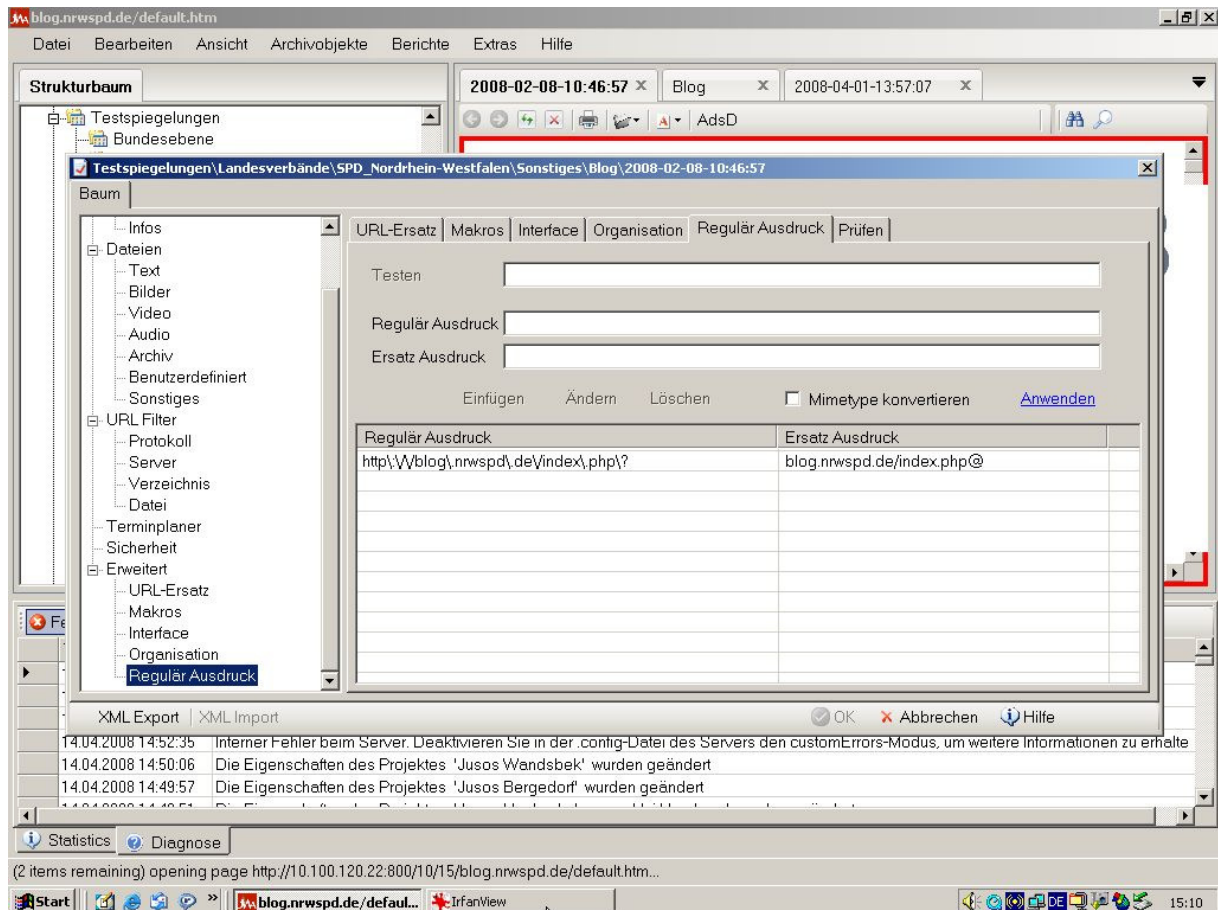


Abbildung 19: Ersetzung mit regulären Ausdrücken

Planung, Ausbau und Betrieb des brandenburgischen revisionssicheren digitalen Langzeitarchivs nach OAIS

Jörg Homberg

Rahmenbedingungen

Gesetzlicher Auftrag

Das Brandenburgische Landeshauptarchiv (BLHA) ist als zentrales Archiv des Landes Brandenburg zuständig für das Archivgut aller Stellen des Landes sowie ihrer Rechts- und Funktionsvorgänger.

Das BLHA wurde beauftragt eine Lösung zur Langzeitarchivierung zu entwickeln, die der Gemeinsamen Geschäftsordnung (GGO) respektive der Registraturrichtlinie folgt.

Ferner wurde gefordert, dass die elektronischen Akten gemäss §§ 239, 257 HGB, der Abgabenordnung und dem GoBS (Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen) revisionssicher archiviert werden können.

Infrastrukturelle Voraussetzungen im Land

Zum Zeitpunkt der Projektinitiierung (2003) stellte sich die Situation in Brandenburg wie folgt dar: Es gab verschiedene DMS/VBS in den Brandenburgischen Landesbehörden; es musste daher mit elektronischen Akten unterschiedlichen Formates und heterogener Qualität gerechnet werden. Ziel war es daher eine produktunabhängige Schnittstelle zu einem geeigneten Speichersystem zu entwickeln, die die Workflows der Anbietung, Bewertung, Aussonderung und der Übernahme abbilden sollte. Als Nutzen für das Land wurden Einsparungspotentiale bei ansonsten notwendigen Entwicklungskosten für spezifische Schnittstellen über das Setzen eines Landesreferenzmodells postuliert.

Parameter

Basieren sollte die Lösung auf XML als Beschreibungssprache, verlässlichen Format-Standards (TIF, PDF/A) sowie zuverlässigen und handhabbaren Speichertechnologien als Medien. Organisatorische Grundlage bildete das DOMEA-Konzept zur Aussonderung und Archivierung elektronischer Akten⁵⁸.

Projekthistorie

- **2003/2004** Projektidee wird Sieger im eGovernment-Wettbewerb des Landes
- **Ab 2004** Bemühungen um Finanzierung über Fördermittel, Partnerschaft mit H&T Greenline GmbH und Microsoft AG Deutschland
- **2004-2006** Erarbeitung einer GGO/Registraturrichtlinie für das Land Brandenburg
- **2005-2006** Mitwirkung am erfolgreichen Test zum XDOMEA-konformen eAktenaustausch

⁵⁸ www.kbst.bund.de/Anlage306096/Abschlussfassung-DOMEA-Organisationskonzept-2.0-pdf-1-25-MB.pdf.

- **Ab 2006** Beschäftigung mit Standard-Formaten zur Langzeitarchivierung, Marktanalyse zu langzeitarchivtauglichen Speichersystemen. Festlegung: TIFF, PDF/A und XML als Transfer- und Archiv-Formate. Entscheidung für ein "BLACKBOX"-Verfahren als Speichersystem. Testung und Validierung einer EMC²-CENTERA⁵⁹
- **Ab 2007** Mitwirkung bei der Ausschreibung eines "Landes-VBS/DMS" unter Berücksichtigung der Archivschnittstelle/Entwicklung des Moduls Aussonderung im Landesreferenzmodell, Prototypisierung der Archivschnittstelle
- **11/2007** "Bewährungstest": Havarie (ein Plaid ausgefallen, Rekonstruktion über Nacht, Datensicherheit jederzeit mind. 600%!)
- **1/2008** Inbetriebnahme einer 2. CENTERA (Spiegelfunktionalität), eines Indexservers und Aufrüstung der produktiven CENTERA auf 18 TB)
- **3/2009** Offizielle Inbetriebnahme der revisionssicheren dLZA (EL.ARCHIV) des Landes durch den Staatssekretär MWFK auf der CeBIT 2009
- **Ab 4/2009** Konzeptionelle Arbeiten an den Modulen "Digitales Zwischenarchiv des Landes Brandenburg" und Feinkonzeption der Ausbaustufe "Urkundsarchiv" (EL.NORM)
- **Ab 9/2009** Prototypisierung von EL.NORM

Ingest

Die "EL-Welt" des Landes Brandenburg

Sehr schematisch lässt sich die künftige Interaktion der verschiedenen Fachverfahren der Ressorts des Landes und des Langzeitarchivs im BLHA wie folgt darstellen:

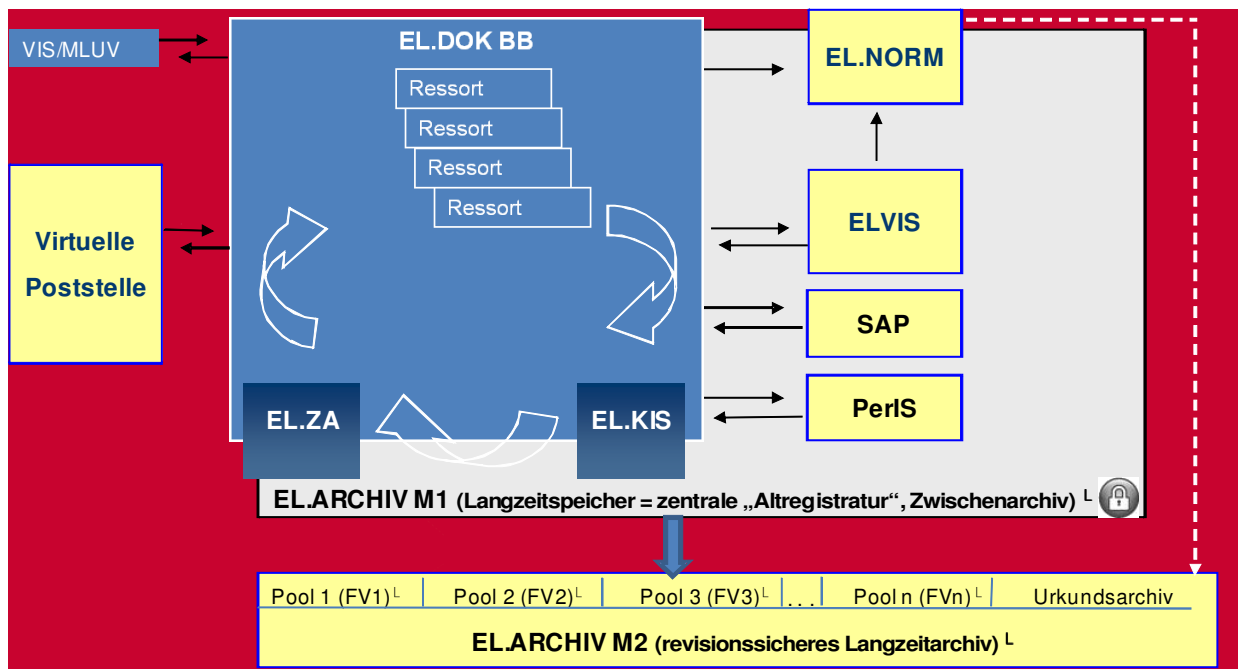


Abbildung 20: Schematische Darstellung EL-Welt im Land Brandenburg

⁵⁹ <http://germany.emc.com/products/family/emc-centera-family.htm>.

- **PerIS** = **P**ersonal**i**nformations**s**ystem
- **EL.ZA** = Ressortübergreifende Kollaborations (**Z**usammen**a**rbeits)-Plattform
- **EL.VIS** = **E**lektronisches Landtags **V**organgsbearbeitungs- und **I**nformations**s**ystem (alle parlamentarischen Initiativen werden digital eingebracht und gehandhabt; von der Einbringung von Anträgen, über die Protokollierung der Plenarsitzungen bis zur Verkündung der Gesetze)
- **EL.NORM** = **E**lektronische Verkündung von Rechtsvorschriften und **N**ormen
- **EL.KIS** = **K**abinet**i**nformations**s**ystem. Vor- und Nachbereitung der Kabinettsitzungen
- **EL.DOK** = Brandenburgisches **D**okumentenmanagement und VBS

Interaktion PRODEA-Client der abgebenden Behörde – Archivclient im BLHA

Die Kommunikation zwischen dem führenden DMS des Landes, EL.DOK und der digitalen Archivlösung im BLHA, EL.ARCHIV erfolgt webbasiert über eine Schnittstelle.

Momentan erlaubt die Applikation eine manuelle oder automatisch (über Fristen) getriggerte Anbietung, Bewertung, Übernahme und Speicherung der elektronischen Akten.

Die Benachrichtigung der beteiligten Personen über erfolgte Übernahme oder eine Kassationsentscheidung werden noch über E-Mail realisiert, in einer späteren Ausbaustufe sollen diese Funktionalitäten den entsprechenden Clients inkorporiert werden.

Bewertung Übernahme Archivierung Recherche Aktenplan Erschließung						
Aktualisieren						
Transferprozess			Titel		Archivieren	Archiv
MdI Anbietung			Testakten aus dem MdI			Endarchiv
Aktion	Nr	AKZ	Aktentitel	Inhalt	Band	Referat
	2	I/6-251-32/A2008#A0003	Einführung (DMS/VBS - landesweit)	Einführung (DMS/VBS - landesweit)		
	1	I/6-251-32/A2008#A0003#V2008#V0001	Teilprojekt Einführung im MI	Einführung (DMS/VBS - landesweit)		

Abbildung 21: Erfassungsmaste EL.ARCHIV – Ausschnitt Archivierungsfunktion nach Bewertung und Übernahme

Speicherung

Die Speicherung folgt der Content Addressed-Methode (CAS).

Der Zugriff erfolgt immer über eine autarke Applikation (über die CENTERA-API) auf den eigentlichen Speicher: D.h. bei Bedarf können neue zeitgemässe Technologien Verwendung finden; der Langzeitspeicher bleibt davon unberührt.

Der Zugriff erfolgt nicht über Filenamen bzw. lokationsabhängige Adresse, sondern über einen Schlüssel. Jedem Objekt wird beim Abspeichern eine weltweit eindeutige Identifikation (CA = Content Address) zugewiesen (Dieser sogenannte Hashwert der Grösse 256Bit wird aus dem binären Inhalt des Datenobjektes errechnet!). Dazu werden die das Objekt betreffenden Metadaten erzeugt (CDF = C-Clip Descriptor File).

Das CDF als XML-Datei kann mit beliebigen zusätzlichen Angaben ergänzt werden.

Das Objekt (BLOB = Binary Large Objekt), das CDF und redundante Spiegeldateien werden dann in der CENTERA gespeichert.

Das CDF erhält zusätzlich eine eigene CA, die dem Objekt übergeben wird, wenn BLOB und CDF zweifach gesichert wurden. Im Prinzip hat diese zusätzliche CA die Funktion eines "Gepäckscheines", über die das Objekt im Sinne der Wiederfindung angesprochen werden kann.

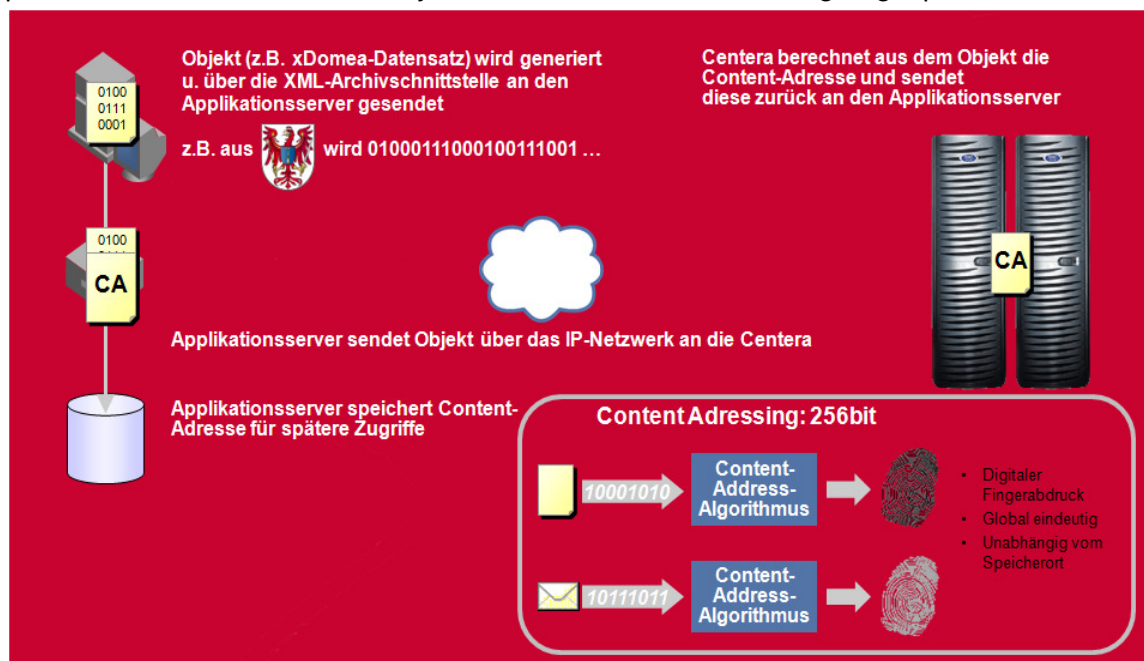


Abbildung 22: Funktionsmodell 1

Das zu archivierende Objekt wird mit zusätzlichen Metadaten angereichert. Diese enthalten einerseits Kenndaten zur Validitätsprüfung, andererseits werden zusätzliche Daten zum Prozess der Generierung des Transfer-Formats angelegt; dies sind technische Metadaten zur dLZA der Objekte, die es künftig erlauben sollen, Migrationsläufe im Batch-Verfahren durchzuführen. Ein dazu eingesetztes Werkzeug ist JHOVE⁶⁰, ein etabliertes Tool zur automatischen Generierung von technischen Metadaten:

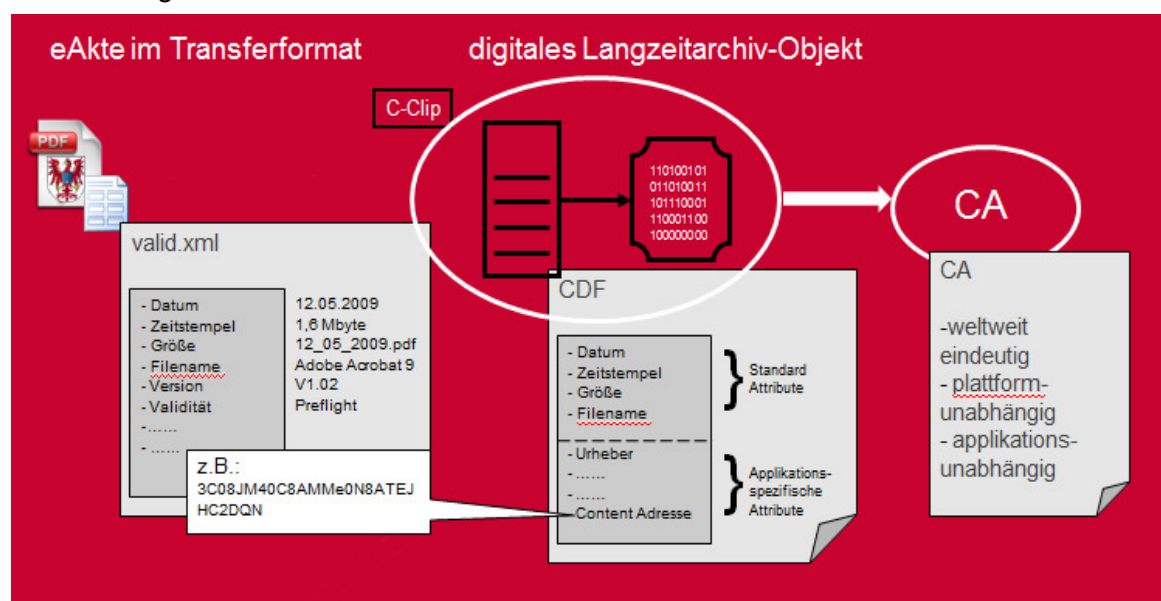


Abbildung 23: Funktionsmodell 2

⁶⁰ <http://hul.harvard.edu/jhove/distribution.html>.

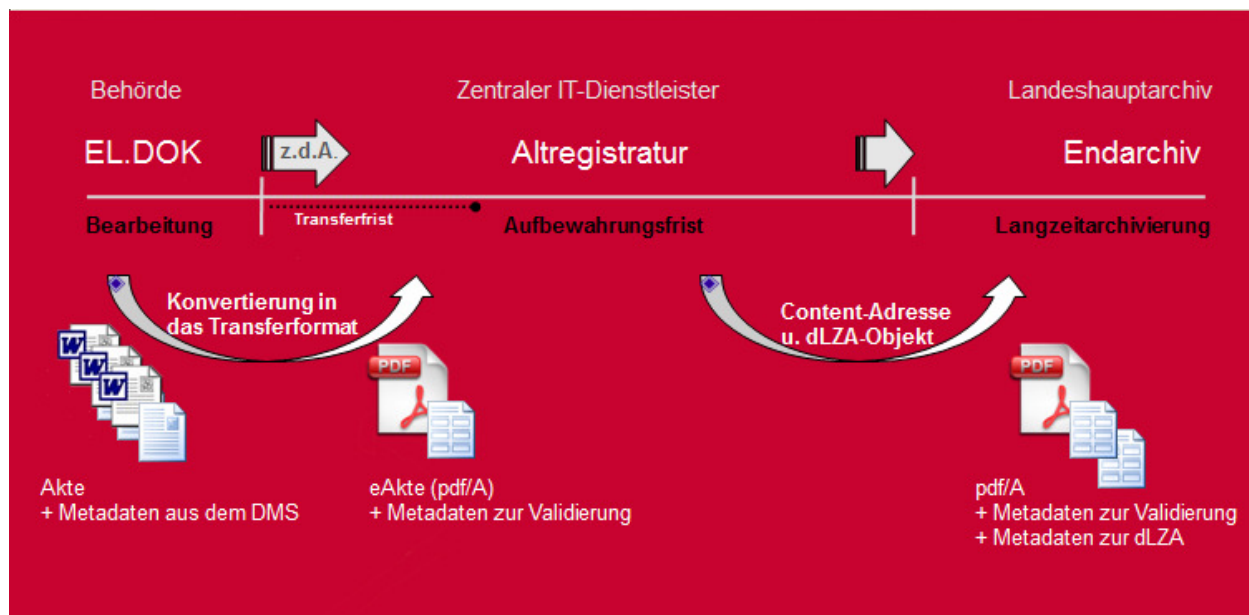


Abbildung 24: Funktionsmodell 3

Die digitalen Objekte (hier eAkten) sind danach:

- **revisionssicher** – bei jeder Änderung (Revision) wird ein neues Objekt angelegt. Das alte wird nicht überschrieben, es ist unantastbar,
- **ein(ein)deutig** – die Objekte werden als Binärdatenstrom (BLOB) gespeichert; begleitende Metadaten (CDF) werden mitgespeichert,
- und können über eine Hashwert-basierte Content-Adresse **sicher identifiziert werden**.

Datenmanagement

Datenretrieval (Recherche) ist sowohl feldspezifisch als auch im Volltext-Modus möglich.

Die aktuelle Version von EL.Archiv, Rev.1.4, erlaubt neben der Bewertung, Übernahme und Archivierung von elektronischen Akten (und den Metadaten von Papierakten!) auch eine Erschliessung von Archivdaten. Perspektivisch soll die Erschliessungsfunktion durch Schnittstellen zu etablierten Archiverschliessungsprogrammen (AUGIAS, SCOPE) ergänzt werden.

Die Metadaten aus der Erschliessung, Validierung des Transfer-Formats und die technischen Metadaten werden sowohl auf dem Langzeitspeicher als auch (aus Performance und Sicherheitsgründen) zusätzlich auf einem MICROSOFT SQL-Server 2005 vorgehalten.

OAIS-Kompatibilität

EL.ARCHIV verarbeitet definierte Formate. Die Daten werden nach dem DOMEA-Aussonderungskonzept kommuniziert. Als Transferformate (nach OAIS-Terminologie SIP) werden XML, TIF(F) G4, PDF/A akzeptiert.

Die SIPs werden zusammen mit den Metadaten als Datenblob gespeichert, d.h. es entstehen Langzeitarchivformate (AIP) als Bitstream, der nur über seinen eindeutigen Hashwert adressiert werden kann!

Die Visualisierung der archivischen Objekte (DIP) erfolgt rollenbasiert:

Das ganze Objekt, Teile des Objekts (wenn Restriktionen vorliegen) oder bei Bedarf auch nur Metadaten zum Objekt (Gastrecherche) können webbasiert repräsentiert werden.

Ausbaustufen

Zurzeit werden im Kontext der Enterprise Content Management-Aktivitäten des Landes Brandenburg Möglichkeiten einer zentralen Ausrichtung der dLZA diskutiert.

Es soll ein "Archivdienst" (mit zentraler Finanzierung) konzipiert und implementiert werden.

Zwei synergistische Faktoren scheinen zurzeit eine Entwicklung in diese Richtung zu forcieren:

1. eine derzeit durchgeführte Wirtschaftlichkeitsanalyse und 2. eine mittelfristig (bis 2010) geplante und kalkulierte Erweiterung von EL.DOK und EL.ARCHIV um die Komponente "Zentrale Altregistratur" des Landes. Aufgebaut auf EL.ARCHIV-Technologie und erweitert um Komponenten zur digitalen Signatur und zur Verschlüsselung werden nach derzeitiger Planung Akten schon nach zwei Jahren in ein zentrales Zwischenarchiv überführt. Dies soll das Produktiv-DMS-System entschlacken und den Hochverfügbarkeitsspeicher entlasten.

EL.ARCHIV ist so konzipiert, dass über elektronische Akten hinaus andere elektronische Objekte archiviert werden können.

Höchste Priorität hat hier die Entwicklung von Technologien zur Langzeitarchivierung von Datenbanken.

Eventuell ist über Kooperationen mit Hochschulen und anderen Institutionen (FH Potsdam, GFZ Potsdam, Bundesarchiv Schweiz u.a.) an entsprechenden Entwicklungen zu partizipieren. Die Archivierung von Audio und Video-Daten ist nachrangig. Hier kann unter Umständen auf Erfahrungen der Medienarchive bzw. von öffentlich-rechtlichen Rundfunkanstalten aufgebaut werden.

Ein weiterer Fokus in der Fortentwicklung der Anwendung liegt auf der perspektivischen Bedienung des XAM-Standards⁶¹.

Im Sinne des OAIS-Referenzmodells macht es Sinn über einen definierten, standardisierten Austausch von Daten über verschiedene Plattformen nachzudenken (Interoperabilität).

XAM kann ein gangbarer Weg bei Vernetzungen von Wissensspeicher-Storagesystemen sein und spielt in unseren Überlegungen hinsichtlich der Zukunftsträchtigkeit und Konsistenz der dLZA im BLHA eine Rolle.

Preservation Planning

Preservation Planning bedeutet im Sinne von OAIS organisatorische und technologische Massnahmen zu treffen, um Konsistenz und Nachhaltigkeit der dLZA über die Jahre hin zu gewährleisten.

⁶¹ <http://www.snia.org/forums/xam/>.

In diesem Zusammenhang seien nur punktuell ein paar Aktivitäten von Belang erwähnt:

- Aktives Verfolgen der Technologieentwicklung und Entwickeln von Empfehlungen in Bezug auf Archivierungsstandards und -politik.
- Überwachen der Archivierungsbemühungen.
- Ausarbeiten von Empfehlungen für die Erhaltung der Lesbarkeit der gespeicherten Information.
- Planen von Datenmigrationen und Kopiervorgängen.
- Haushalterische Vorsorge treffen für langfristige Finanzierung von IT-Infrastruktur, Telekommunikationswege, Wartungsverträge etc.
- Regelmässiger Dialog mit aktenabgebenden Instanzen, Schulungen und Informationsveranstaltungen zu Themen der dLZA.

Es wird davon ausgegangen, dass sich die Hardware-Halbwertszeitspannen künftig eher nicht vergrössern! Daher müssen Vorsorgemassnahmen in dem Sinne getroffen werden, als notwendige Migrationsszenarien unterstützt, ja schon vorangedacht werden. Flankierende Massnahmen in diesem Kontext sind:

- Ausschliessliche Bedienung allgemeingültiger, akzeptierter und wohldokumentierter Standards.
- Ablehnung proprietärer Ansätze und Pseudostandards.
- Redaktionelle (archivische) und technische Metadaten sind kontrolliert anzulegen (Dublin Core⁶²).
- Eine Zertifizierung hinsichtlich OAIS-Compliance⁶³ als "Vertrauenswürdige digitales Langzeitarchiv" ist anzustreben.
- Als Basistechnologien bzw. Richtlinien dienen: XML als Beschreibungssprache, DOMEA als Organisationskonzept.
- Wahrung der Unabhängigkeit der Datenspeicher.
- Applikation: Die Software der Lösung ist so zu konstruieren, dass eine Hardware-Migration jederzeit möglich ist. Die Software-Komponente muss vollkommen unabhängig vom Speichermedium CENTERA sein und auch mit anderen Systemen kommunizieren können.
- Umgekehrt ist auch die CENTERA unabhängig von der Applikation. Die Schnittstelle (API) der CENTERA ist öffentlich zugänglich.

⁶² <http://dublincore.org/>.

⁶³ http://nestor.cms.hu-berlin.de/moinwiki/AG_Vertrauensw%C3%BCrdige_Archive_-_Zertifizierung.

EL.NORM. Die geplante elektronische Normverkündung im Land Brandenburg (Stand: Mai 2009)

Simone Stumpe

Brandenburg plant die Einführung der elektronischen Verkündung von Rechtsvorschriften. Das Projekt trägt die Bezeichnung "Elektronische Normverkündung" (EL.NORM). Es betrifft die Verkündung aller Gesetze und Verordnungen des Landes sowie die Veröffentlichung der Verwaltungsvorschriften der Landesbehörden. Hierdurch soll die herkömmliche papiergebundene Verkündung dieser Rechtsvorschriften abgelöst werden.

Den unmittelbaren Anstoss zu dem Projekt hat ein Sonderausschuss des Landtages Brandenburg gegeben. Seine Aufgabe war es, Vorschläge für eine Modernisierung der Verwaltung zu entwickeln. Einer dieser Vorschläge betrifft die elektronische Normverkündung. Dazu hatte sich der Ausschuss zuvor in Österreich über das dort vom Bundeskanzleramt betriebene Rechtssystem RIS informiert. In Österreich werden die Bundesgesetzblätter seit 2004 ausschliesslich im RIS rechtsverbindlich bekannt gemacht.

Das Ministerium der Justiz wurde daraufhin um Vorlage einer Konzeption für die elektronische Normverkündung gebeten. Einen entsprechenden Auftrag hat das Ministerium der Justiz an ein externes IT-Fachunternehmen, die Firma IMTB, vergeben und die Erstellung der Konzeption von EL.NORM eng begleitet. Parallel dazu wurden die zur Einführung der elektronischen Verkündung erforderlichen Gesetzentwürfe erarbeitet.

Eckpunkte der Konzeption

Mit der Umsetzung von EL.NORM soll die webbasierte, amtliche Bereitstellung von Gesetzen und Verordnungen erfolgen. Dies schafft mehr Transparenz bei den am Rechtserzeugungsprozess Beteiligten und bei den Normadressaten und führt zu einer schnelleren Verfügbarkeit der Normen für die Öffentlichkeit.

Seit der Bereitstellung von nichtamtlichen Fassungen der Verkündungsblätter im Brandenburgischen Vorschriftensystem ist die Zahl der in Papierform abonnierten Fassungen stetig rückläufig. Daher ist zur Herstellung einer für die Verkündung von Rechtsvorschriften geeigneten Öffentlichkeit dringend eine rechtlich tragfähige Alternative zur Papierverkündung notwendig. Auf der elektronischen Verkündungsplattform sollen sich letztendlich Bürger, Wirtschaft und Verwaltung kostenfrei über erlassene Vorschriften informieren können. Um den ggf. weiter bestehenden geringen Bedarf an Papierexemplaren zu decken, wurden entsprechende Übergangsregelungen bzw. die Möglichkeit zum Bezug von Papierexemplaren in die Betrachtung einbezogen.

Die Ausfertigung und Verkündung einer Rechtsvorschrift gliedert sich in drei Prozessschritte: Ausfertigung der Norm, eigentlicher Arbeitsschritt der Verkündung der Norm, Herstellung der Öffentlichkeit. Nach der Konzeption ist für die Verkündung der Rechtsvorschrift die zentrale Verkündungsstelle zuständig. Sie besteht aus den Redaktionen der Verkündungsblätter des

Landtages (GVBl. I) und der Landesverwaltung (GVBl. II, ABl.), die künftig als Verkündungsstellen tätig werden sollen. Während die organisatorischen Rahmenbedingungen von den Verkündungsstellen dezentral durchgeführt werden, findet die Veröffentlichung der Verkündungsblätter auf einer zentralen Verkündungsplattform statt.

Künftig sollen Ausfertigung und Verkündung ausschliesslich elektronisch stattfinden. Die Anforderungen bezüglich der Authentizität und Integrität einer Norm lassen sich mittels einer angebrachten qualifizierten elektronischen Signatur sicherstellen. In der Konzeption zu EL.NORM ist vorgesehen, dass die ausfertigende Stelle (z.B. Präsident des Landtages) eine qualifizierte elektronische Signatur an die auszufertigende Originalnorm (PDF/A-Datei) anbringt. Bei der Signatur handelt es sich um eine qualifizierte elektronische Signatur entsprechend dem Signaturgesetz. Hierdurch wird eine Überprüfung für die zu erwartende lange Laufzeit der Normen sichergestellt. Weiterhin ist in der Konzeption zu EL.NORM eine ggf. notwendige Übersignierung einer mathematisch unsicher gewordenen Signatur (vgl. § 17 SigV) vorgesehen. Die ausfertigte Originalnorm ist nicht mehr veränderbar. Da für die Verkündung die Rechtsvorschrift im jeweiligen Layout des Verkündungsblattes anschliessend redaktionell aufzubereiten ist, erstellt die Verkündungsstelle aus der Rohfassung, die noch in einer .doc- bzw. .odf-Datei vorliegt, eine Ausgabenorm (PDF/A-Datei). Diese wird durch die Verkündungsstelle mit einer qualifizierten elektronischen Signatur signiert. Die Ausgabenorm ist zusammen mit einem Prüfbericht der Ausgabenorm auf der zentralen Verkündungsplattform des Landtages und der Landesverwaltung für die Öffentlichkeit kostenfrei recherchierbar.

Dadurch entstehen neue Anforderungen an eine unbegrenzte Haltbarkeit und Lesbarkeit von authentischen, elektronischen Normen. Hierfür ist es nicht nur notwendig, elektronische Normen zu speichern, sondern es ist auch sicherzustellen, dass die abgelegten Normen unter Wahrung der Authentizität, Integrität und Beweiskräftigkeit zeitlos wieder auffindbar (recherchierbar) und lesbar (benutzbar) sind. Normen sind ewig aufzubewahrende Quellen. In der Konzeption zu EL.NORM ist für die Aufbewahrung und Pflege dieser elektronischen Daten die mandantenfähige Infrastruktur von EL.ARCHIV und für den Recherchezugriff durch die Öffentlichkeit ein Auskunftssystem in einer Demilitarisierten Zone (DMZ) vorgesehen.

Um den Anforderungen an eine revisionssichere Aufbewahrung der in Kraft getretenen Rechtsvorschrift und eine historische Endarchivierung der ausser Kraft getretenen Rechtsvorschrift zu ermöglichen, sieht die Konzeption von EL.NORM eine Speicherung der Daten entsprechend im elektronischen Langzeitspeicher bzw. im Urkundsarchiv vor. Dies ermöglicht die Pflege der Haltbarkeit und Lesbarkeit aller elektronischen Normen.

Weiterhin werden in der Konzeption zu EL.NORM zwei Speichersysteme für eine redundante Sicherung definiert, damit auch bei Ausfall oder Verlust eines Systems ein vollständiger Datenbestand vorhanden ist. Aus diesem dann noch vorhandenen Datenbestand kann die Wiederherstellung der Daten für das zweite Speichersystem erfolgen. Beide Speichersysteme sind dabei an unterschiedlichen Standorten aufzustellen.

Einbettung in die E-Government-Architektur des Landes Brandenburg

Das Projekt EL.NORM wird unmittelbar in die übergreifende E-Government-Architektur des Landes Brandenburg eingebettet werden. Diese E-Government-Architektur wird derzeit in verschiedenen Einzelprojekten aufgebaut. Die Basiskomponenten sowie ein landesweites Dokumentenmanagement- und Vorgangsbearbeitungssystem (DMS/VBS) – EL.DOK genannt – stellt der zentrale IT-Dienstleister bereit. Mit EL.DOK, das ein elektronisches Kabinettinformationssystem (EL.KIS) einschliesst, ist die umfassende Abbildung elektronischer Geschäftsprozesse in der obersten Landesverwaltung geplant. Darüber hinaus sollen mit dem Elektronischen Landtagsvorgangsbearbeitungs- und -informationssystem (ELVIS) die parlamentarischen Prozesse im Landtag unterstützt werden. Ergänzt wird dies durch das Projekt EL.ARCHIV des Brandenburgischen Landeshauptarchivs (BLHA) mit der Konzipierung einer landesweiten Archivierungsstrategie.

Stand der Projektumsetzung

Ausgehend von der Konzeption EL.NORM hat das Ministerium der Justiz sodann die gesetzlichen Regelungen zur elektronischen Verkündung von Gesetzen und Verordnungen erarbeitet. Dies betrifft eine Ergänzung der Landesverfassung (Artikel 81), um eine verfassungsrechtliche Grundlage für die elektronische Ausfertigung und Verkündung zu schaffen, ferner den Regierungsentwurf für ein Brandenburgisches Ausfertigungs- und Verkündungsgesetz, durch das die Einzelheiten der elektronischen Verkündung von Gesetzen und Rechtsverordnungen im Gesetz- und Verordnungsblatt geregelt werden sollen. Die Gesetzentwürfe sind mittlerweile von der Landesregierung beschlossen und in den Landtag Brandenburg eingebracht worden. Die 1. Lesung der Gesetzentwürfe hat Anfang April 2009 stattgefunden.

Mit der technischen Umsetzung des Projekts wurde Anfang März 2009 begonnen. Unter der Voraussetzung, dass der Landtag die gesetzlichen Grundlagen für die elektronische Verkündung von Gesetzen und Rechtsverordnungen beschliesst, könnte das Gesetz- und Verordnungsblatt für das Land Brandenburg Anfang Oktober 2009 erstmals elektronisch veröffentlicht werden. Aufgrund der Komplexität des Gesamtprozesses von EL.NORM durch seine Vielzahl von Zulieferern und der damit verbundenen Masse an Informationen ist in der Konzeption eine stufenweise Einführung vorgesehen. EL.NORM als technisches System wird endgültig Anfang 2010 fertiggestellt sein. In seiner endgültigen Ausprägung soll es dann auch die Veröffentlichung der weiteren Verkündungsblätter ermöglichen.

Fazit

Mit EL.NORM könnten künftig der Landtags- und Landesverwaltung sowie der gesamten Öffentlichkeit Normen schneller, vor allem aber elektronisch, rechtssicher und nachhaltig zur Verfügung stehen. Durch die erstmalige durchgehende elektronische Ausfertigung und Verkündung würde der Zugang zu dem Gesetz- und Verordnungsblatt für das Land Brandenburg für

Nutzer erheblich vereinfacht und barrierefrei werden. Die Verbreitung des Gesetz- und Verordnungsblattes nähme zu. Der Aufwand für Recherchen würde deutlich sinken. Der Zugriff auf aktuelle amtliche Informationen würde beschleunigt. Die Umstellung könnte insofern auch zu einem gewissen wirtschaftlichen Nutzen für Wirtschaft und Verwaltung bei den Prozess- wie auch bei den Sachkosten führen. Das Verfahren der Gesetz- und Verordnungsgebung wird in seiner Schlussphase insgesamt effizienter gestaltet. Die Aufbewahrung der in Kraft getretenen Normen und die Endarchivierung der ausser Kraft getretenen Normen können durch den elektronischen Langzeitspeicher und das Urkundsarchiv sofort sichergestellt werden.

EL.NORM bedeutet daher auch für Archivarinnen und Archivare einen wichtigen Fortschritt. Die Einführung der elektronischen Verkündung von Rechtsvorschriften verändert auch die Übergabe von Normen an das Brandenburgische Landeshauptarchiv (BLHA). Während derzeit die papiergebundenen Originale am Jahresende ans BLHA ausgesondert werden, können die Normen als elektronisches Dokument sofort überführt werden. Dies ermöglicht dem Archiv eine jederzeit historisch vollständige und auf Grund der elektronischen Datenverarbeitung rationellere Archivierung aller Normen.

Eine wesentliche Voraussetzung bildet die feste Zusammenarbeit innerhalb der E-Government-Architektur des Landes Brandenburg.

Das Archivierungssystem AMphora der Aktion Mensch

Andreas Jüngling, Ulrich Kamp

Vorbemerkung

Als sich im Jahre 2004 der Geschäftsbereich Presse, Öffentlichkeitsarbeit und Aufklärung entschloss, nach vierzigjährigem Bestehen der Deutschen Behindertenhilfe – Aktion Mensch e.V. das Gedächtnis des Vereins in Form eines Archivs zu institutionalisieren, erlebte die Diskussion um adäquate Erschliessung und langfristige Aufbewahrung digitaler Archivalien gerade wieder eine Renaissance. Diese Diskussion war und ist für das Archiv der Aktion Mensch besonders wichtig, da von vornherein feststand, in diesem Archiv und dem dafür notwendigen Verzeichnungs- und Erfassungssystem alle Formen und Gattungen von Archivalien, von der bekannten Papierakte über Photographien und Plakate bis zu digitalen und digitalisierten Überlieferungen, aufzunehmen und für die Nutzer zugänglich zu machen.

Unter diesen Voraussetzungen wurde eine Marktanalyse durchgeführt, mit dem Ziel, die unterschiedlichen auf dem Markt befindlichen Systeme der Möglichkeit einer Eigenentwicklung gegenüberzustellen. Den Untersuchungen zufolge ergaben sich signifikante Vorteile für die der eigenen Problemstellung angepasste Beauftragung, ein neues und eigenständiges elektronisches Erschliessungssystem für Archivgüter entwickeln zu lassen. Ausschlaggebend für diese Entscheidung war, dass seitens der Aktion Mensch die Steuerung der Umsetzung unserer Vorgaben vollständig ermöglicht wurde. Zu diesen Vorgaben zählte auch, AMphora als Open-Source-System zu gestalten. Begründet wird diese prospektive Erwartung mit der Möglichkeit, mit einem Open-Source-System nicht nur neue Impulse im Entwicklungsbereich der digitalen Langzeitarchivierung zu setzen, sondern auch die Diskussion und den theoretischen wie technischen Austausch in den Fachkreisen entlang eines konkreten Projektes weiterzutragen.

Im Vorfeld wurden grundsätzliche Anforderungen an die Neuentwicklung gemäss den mittlerweile gängigen Standards des Preservation Planning nach OAIS für die digitale Langzeitarchivierung und gemäss den Prinzipien der Stufenverzeichnung nach ISAD/G formuliert. Ausserdem musste eine den aktuellen IT-Architekturen folgende strukturierte Suche und Volltextsuche bereitgestellt werden. Für die Aufgabe zur Archivierung digitaler Archivalien waren die Sicherstellung der Datenauthentizität durch den HASH-Wert und die Bereitstellung einer digitalen Signatur (PURL) wichtig. Die Steuerung einer langfristigen Erhaltung der digitalen Archivalien war eine weitere wichtige Anforderung. Ein wesentlicher Vorteil einer modernen IT-Architektur findet sich in der fein austarierbaren Steuerbarkeit von Verzeichnungs- und Nutzungsrechten. Das zugrunde gelegte Modell verschiedener zueinander in Beziehung zu setzender Regel-, Attribut- und Metadatenschemata bedingt daher nicht bloss die Umsetzung jener Anforderungen, sondern eine einfache und individuelle Anpassbarkeit an die Erschliessungs- und Verwaltungsordnungen jedes einzelnen Archivs.

Einleitung

Vor dem Hintergrund der Vielfalt der Überlieferungen und ihres häufig nicht aktenplanmässigen Zustandekommens sowie einer möglichst schonenden Behandlung bestimmen daher zwei Aspekte die grundsätzlichen Funktionen von AMphora: Erstens muss eine möglichst hohe Flexibilität bei der Bildung der Archivtekonik ebenso gewährleistet sein wie eine nach dem ISAD/G-Standard und eigenen Grundsätzen festgelegte präzise Verzeichnungsfähigkeit, die eine objektive und logisch-nachvollziehbare Ordnung für die Überlieferungen gestattet. Zweitens müssen neben der Einspielung (Ingest) und Integration von digitalen Archivalien auch Digitalisate (z.B. von papiernen Akten) separat von den rein elektronisch erzeugten Archivalien verzeichnet und nutzbar gemacht werden können. Hinzu kommen Anforderungen an das Archivsystem der Aktion Mensch, die durch die Eignung als aktuelles Pressearchiv bzw. Dokumentationsdatenbank weitere Funktionen für digitale Objekte erfordern. In diesem Rahmen sollten auch digitale Datenbestände aus Produktionssystemen übernommen werden können.

Prinzipiell sollte also das geplante Archivierungssystem gegenüber allen denkbaren Archivalien, Archivierungsmethoden und Verzeichnungsweisen offen bzw. bei Bedarf unkompliziert umzugestalten sein, um künftige Entwicklungen in den Archivwissenschaften auf der einen Seite und den Erfordernissen einer sich verändernden Archivierungspraxis auf der anderen Seite adäquat aufgreifen und mitgestalten zu können.

Die Bestände im Archiv der Aktion Mensch

Aus fünf hauseigenen Überlieferungsprovenienzen (Geschäftsführung/Vorstand; Lotterie/Buchhaltung; Marketing; Förderung; Presse/Öffentlichkeitsarbeit/Aufklärung) fügen sich die hauptsächlichen Inhalte des Archivs mit den institutionellen Vorläufern zusammen. Nachlässe aus dem Personenkreis des Gründungsumfeldes und aus thematisch oder institutionell mit der Aktion Sorgenkind/Aktion Mensch verbundenen Organisationen ergänzen den Gesamtbestand und erschliessen die historischen Tiefenschichten. Weiterhin wurden und werden Materialien von Mitgliederinstitutionen wie dem ZDF (z.B. Referenzüberlieferungen der Aktion Sorgenkind im ZDF-Archiv), Kooperationspartnern und Fremdfirmen (z.B. Agenturen als Dienstleister) übernommen. Der Aktenanteil im Bestand Presse, Öffentlichkeitsarbeit, Aufklärung ist daher gegenüber den verschiedenen Filmformaten, Originalfotographien oder Werbematerialien geringer als in den anderen Bestandsüberlieferungen. Da für ein Unternehmen die Aussenwahrnehmung und deren Steuerung besonders wichtig ist, sammelt das Archiv zugleich Zeitungsausschnitte, lässt Medien-Clippings anfertigen, übernimmt unabhängige wissenschaftliche Arbeiten und ergänzt diese Analysehilfsmittel durch Zeitzeugeninterviews und Gutachten sowie Erhebungen der Markt- und Meinungsforschung. Als Archiv mit dem um einen Dokumentationsauftrag erweiterten Aufgabenbereich erwirbt und katalogisiert es ebenso Fachliteratur, Zeitschriften, graue Literatur und die Sammlung der Zeitungs- wie Zeitschriftenausschnitte zu den Themen, die die Arbeit der Aktion Mensch unmittelbar und mittelbar betreffen. Dieser Bibliotheksbestand ist integraler Bestandteil des Archivs und wird über das Archivsystem gleichbe-

rechtigt als eigener Bestand mit allen weiteren Archivalien verwaltet und den Mitarbeitern wie Interessierten zugänglich gemacht.

Aufgaben und Funktionsweise von AMphora

Bereits in der Vorbemerkung wurde auf die grundgelegten Strukturen von AMphora nach OAIS (Preservation Planning) sowie ISAD/G hingewiesen. Die Vorteile dieses OAIS-Standards liegen auf der Hand: Erstens die präzise Zuweisung der deskriptiven Metadaten beim Ingestieren, zweitens die Nutzbarkeit der AIP als Erfassungs- und Verzeichnungssystem, drittens die klare Sicherung der digitalen Metadaten. Zudem bieten sich im Bereich des Nutzers metadatensichernde Überleitungsmethoden an, die die authentifizierende Abrufbarkeit der Archivalien garantieren. Neben der Authentifikation steht die eindeutige identifikatorische Abrufbarkeit der digitalen Archivalien aus den Metadaten- und DMS-Datenbanken⁶⁴ auf der DIP-Seite im Vordergrund.

Die Umstellungen sämtlicher Bearbeitungs- und Verwaltungsvorgänge auf elektronische Verarbeitung, in die komplementär papierene Verschriftlichungsstufen integriert sind, macht daher eine gleichzeitige und doch systematisch trennbare IT-Lösung sinnvoll. Klassische Archivalien werden gleichberechtigt und gleichwertig mit digitalen und digitalisierten Archivalien auch in Bezug und wo notwendig in Abhängigkeit zueinander aufgenommen und verzeichnet. Auf diese Weise sind alle Archivalien in einem System katalogisiert und abrufbar.

Auf der anderen Seite erfolgt der primäre Zugriff auf die vorhandenen und erfassten Archivalien materialschonend und einfach über das IT-Verzeichnungssystem für alle internen wie externen Nutzer. Zugriff heisst Suche, Sichtung, Ansicht, Bestellung, Bereitstellung, Verwaltung der Ausleihe und Ausgabe sowie Hilfe bei der Erarbeitung von Chroniken, Festschriften und anderem.

Strukturelle Prinzipien von AMphora

Funktionen und Besonderheiten des Archivsystems

Grundsätzlich werden sechs Funktionseinheiten unterschieden. Erstens nach den Verzeichnungs- bzw. Ordnungskriterien, zweitens nach der Erfassung und Gliederung der Verzeichnungsmetadaten, drittens nach der Recherchefunktion entsprechend den Verzeichnungsdaten und Inhalten, viertens nach den Archivverwaltungsprozessen sowie fünftens und sechsten nach Benutzerrechten und Sichtberechtigungen.

Ordnungskriterien der Verzeichnung:

Innerhalb der nach ISAD/G definierten Stufenverzeichnung im tektonischen Aufbau lassen sich die Archivalien nach ihrer Herkunft sowie nach ihren formalen Eigenschaften gliedern. So als "klassische", d.h. in den archivisch bekannten Papierformaten, als "logische" im Sinne eines strukturierenden Formats sowie als "digitale" Formate eingeordnete originäre digitale Objekte. Dabei werden Digitalisate von reinen digital entstandenen Materialien unterschieden. Unab-

⁶⁴ Die Abkürzung DMS steht für Document-Management-System.

hängig von ihren Formaten und Eigenschaften gestattet AMphora die Verzeichnung der Archivgüter je als Einzelarchivalien gegenüber der logischen Sammlungseinheit. Die jeweiligen Archivaliengattungen (siehe unten) können je nach archivspezifischer Anforderung einander zu-, über- und/oder untergeordnet werden. Diese Gattungsgliederung entspricht ausdrücklich dem Prinzip der Stufenverzeichnung nach ISAD/G. Die Gattungen sind dabei ebenso wie die Archivtektonik vollständig selbst konfigurierbar.

Deskriptive Metadaten:

Die gattungsspezifischen Attribute, die die klassischen, logischen und digitalen Formate formal und inhaltlich erschliessen lassen, können sowohl nominell als deskriptiv individuell definiert werden. Gemäss des ISAD/G-Standards empfiehlt sich die Verwendung von "Titel" der Archivalie, "Signatur" zur Verortung und Wiederauffindung, "Datierung" oder "Laufzeit", Provenienzstelle und Umfang.⁶⁵ Zusätze und Ergänzungen, die über die Minimalbeschreibung hinausgehen, können sehr einfach je nach Metadatenmaske hinzugefügt werden. Das Archiv der Aktion Mensch beispielsweise erfasst "Enthält"-Vermerke, Bearbeitungsvermerke und weitere formale oder physikalische Merkmale der Archivalie. Ausserdem werden Hinweise zum Lagerort, Schlagworte sowie die Angaben zu den natürlichen und juristischen Personen (siehe unten), die mit dem Gut in Zusammenhang stehen, aufgeführt. Jeder physikalischen oder logischen Archivalie kann ein entsprechendes Digitalisat zugeordnet werden, was eine schonendere Nutzung gewährleistet. Darüber hinaus gibt AMphora die Möglichkeit, Schlagwörtern zu vergeben und in gesonderte Textfelder einzutragen.

Recherche auf Verzeichnungsdaten und auf Inhalte digitaler Dokumente:

Neben der Volltextsuche bietet AMphora verschiedene detaillierte Suchfunktionen an. Dazu zählen u.a. die Auswahl aller deskriptiven Metadaten aller Archivaliengattungen, wobei diese Auswahl auch gattungsspezifisch, d.h. für und innerhalb der digitalen Archivalien vorgenommen bzw. mit der Volltextsuche oder der Schlagwortsuche kombiniert werden kann.

Archivverwaltungsprozesse:

Alle Verwaltungsfunktionen lassen sich in AMphora vordefinieren und umsetzen. Dies betrifft zum Beispiel die Erstellung von Chroniken, Findbüchern oder Ausleihvorgängen. Die dabei erfassten Daten erlauben zudem Auswertungen über die Art und den Umfang der Nutzung einzelner Archivalien oder Bestände.

⁶⁵ Vgl. ISAD(G) – Internationale Grundsätze für die archivische Verzeichnung, übersetzt und neu bearbeitet v. Rainer Brüning, Werner Heegewaldt, Nils Brübach, (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaften, 23), 2., überarb. Ausgabe, [2002], S. 20.

Benutzerrechte:

Für alle Anwender von AMphora können Bearbeitungsrechte wie Ändern oder Löschen eingeschränkt werden. Administrative Funktionen sind von vornherein nur für einen ausgewählten Nutzerkreis verfügbar.

Grundsätzlich sind nur jene berechtigt, Neuanlagen und Änderungen in AMphora einzugeben, die Mitarbeiter des Archivs sind. Darüber hinaus kann jedem Mitarbeiter des Hauses und auch Externen die lesende Nutzung des Archivdatenbestandes ermöglicht werden.

Sichtberechtigungen auf Bestände:

Unabhängig von Benutzerrechten können auf allen tektonischen Ebenen (von den Beständen bis zur einzelnen Archivalie) auch die Sichtrechte eingeschränkt werden. Zusätzlich sind Sperrfristen möglich, die ebenfalls die Sichtbarkeit von Archivalien für bestimmte Nutzergruppen einschränken. Diese Regelsätze werden im System als "Domänen" bezeichnet.

Die Sichteinschränkungen wirken sich sowohl auf die Archivtektonik als auf die Ergebnisse von Recherchen aus.

Archivierung digitaler Archivalien

Ein besonderes Kennzeichen von AMphora bei der Archivierung digitaler Archivalien ist die Verwendung einer elektronischen Signatur. Verwendet wird hierfür die sogenannte PURL, eine persistente URL, die auf der technischen und logischen Grundlage des World Wide Web (HTTP) basiert. Diese PURL ist gleichsam eine festgelegte eindeutige und nicht revidierbare Adresse, die mit dem Ingest der digitalen Archivalie an deren Metadaten (Dateiname, Dateiattribute, Dateinhalt etc.) angekoppelt wird. Ihrer Funktion nach ist die kostenlose PURL mit dem Persistent-Identifier-Standard der Deutschen Nationalbibliothek vergleichbar, aber folgt wie AMphora dem Open-Source-Gedanken.

Der Prozessablauf ist dabei recht einfach. Ausgehend von einer für Archivare und Archivnutzer gemeinsamen Benutzeroberfläche von AMphora wird einerseits die digitale Datei (gilt für alle Dateiformate, bspw. DOC, PNG, TIF, TXT, PDF) auf dem DMS und andererseits die deskriptiven Metadaten sowie die Nutzungs- und Systemrechte der Archivalie in einer für diese Metadaten bestimmten SQL-Datenbank (z.B. Oracle) gespeichert. Sowohl beim DMS als auch bei der Metadatenbank handelt es sich um logisch voneinander getrennte Systeme. Die PURL garantiert die Zuordnung der archivalischen Metadaten zur gespeicherten Datei. Beim Zugriff auf eine digitale Archivalie wird die PURL mittels des HTTP-Protokolls an das DMS gesandt und die Datei geladen.

Zusammengefasst heisst das Folgendes: Wird eine digitale Archivalie vom Nutzer geladen (begrifflich analog dem "Heben" im Archiv), entspricht dieser Abruf dem bekannten Gebrauch von Archiv-Signaturen. Die PURL wird an das AMphora-DMS gesandt und die entsprechende digitale Archivalie identifiziert und aus der Datenbank abgerufen. Man erhält damit keine Spiegelung oder ein Duplikat der digitalen Archivalie, sondern die Originaldatei. Als "Ansichtsexemplar" ist ein verkleinertes Bild (Thumbnail) verfügbar. Einspielung, Sicherung, Lagerung

und Gebrauch einer digitalen Archivalie unterscheiden sich nicht von den Regeln, wie sie für klassische Archivgüter gelten. Die Trennung der IT-technischen Seite von den fachlichen Anforderungen der Archivierung umgeht eine Vielzahl von Problemen, die mit Fragen der Modifizierung oder Manipulation der Metadaten von digitalen Archivalien bei der Langzeitarchivierung in anderen Verfahren entstehen. Das heisst konkret, dass der Zugriff auf die digitale Datei nicht von deren Benennung (nachträgliche Titeländerung möglich) abhängig ist, Namensdubletten (bspw. bei seriellen Akten) somit auch unterscheidbar bleiben. Durch die Trennung von IT- und Fachanforderung können verschiedene Speichertechnologien eingesetzt werden. Dies bietet auch im Falle einer "digitalen Umbettung" zur Langfristerhaltung den Vorteil, dass diese unabhängig von den Archivmetadaten, also unabhängig von der eingesetzten Software, der Art der Formatmigration und der verwendeten Speichertechnik erfolgen kann.

Begrifflichkeiten in AMphora

Tektonische Prinzipien: Provenienz und Pertinenz

Grundsätzlich orientiert sich der Aufbau, d.h. die Tektonik und die archivalische Beschreibungsweise, am klassischen Prinzip der Provenienz, also an der Herkunft des Archivgutes. So können die chronologischen und strukturellen Bezüge und Zusammenhänge in der Gänze der Organisation (z.B. Unternehmen, Behörde) abgebildet werden.

Grundsätzlich heisst allerdings auch, dass in Anbetracht der Besonderheit von Unternehmens-, Themen- oder Wirtschaftsarchiven nicht völlig auf einen Aufbau verzichtet werden kann, der auf der sachlich und thematisch begründeten Zusammenstellung von Archivalien beruht, also dem Pertinenzprinzip folgt. Auch solche gemischten Gliederungen können von AMphora unterstützt werden, z.B. durch spezifische Gattungen (Akte/Bestand vs. Sammlung).

Archivaliengattung

Grundlegendes Kriterium bei der Erfassung von Archivalien in AMphora ist neben der Tektonik die Archivaliengattung. Die Gattung legt fest, um welchen Typ von Archivalie es sich jeweils handelt, wie diese archivisch erfasst und wo sie innerhalb der Tektonik eingeordnet werden kann.

Mit der Festlegung der Gattung der einzelnen Archivalien werden gleichzeitig Festlegungen für die konservatorische Behandlung und für spätere Nutzungsmöglichkeiten getroffen. Die Gattung bildet so die Grundlage des in AMphora festgelegten Archivierungsregelwerks.

Klassische, physisch vorhandene Archivalien werden von digitalen Archivalien unterschieden, für gliedernde Einheiten werden logische Archivalien verwendet.

Aufbauend auf dieser Unterscheidung bestimmt die jeweilige Gattung, welche Eigenschaften (Eingabefelder, Attribute) bei der Erfassung der einzelnen Archivalien dokumentiert werden können oder welche Personenrollen bei der Erfassung zur Verfügung stehen.

Darüber hinaus bestimmt die Gattung, an welchen Stellen innerhalb der Tektonik die Archivalien aus erfassungssystematischer Sicht eingestellt werden können. So wird die Einhaltung der

definierten Erfassungsregeln sichergestellt und die Qualität der Erfassung innerhalb der Archivtektonik nimmt deutlich zu.

Bei der Definition der einzelnen Gattungen wird festgelegt, wie die Eingabemasken aussehen und welche Symbole (Icons) in den Masken und Tabellen für die jeweilige Gattung verwendet werden.

Dieses gattungsbezogene Regelwerk kann somit an geänderte oder neue Archivalientypen angepasst werden. So kann über das Gattungsregelwerk AMphora flexibel auf Änderungen in Archivabläufen reagieren.

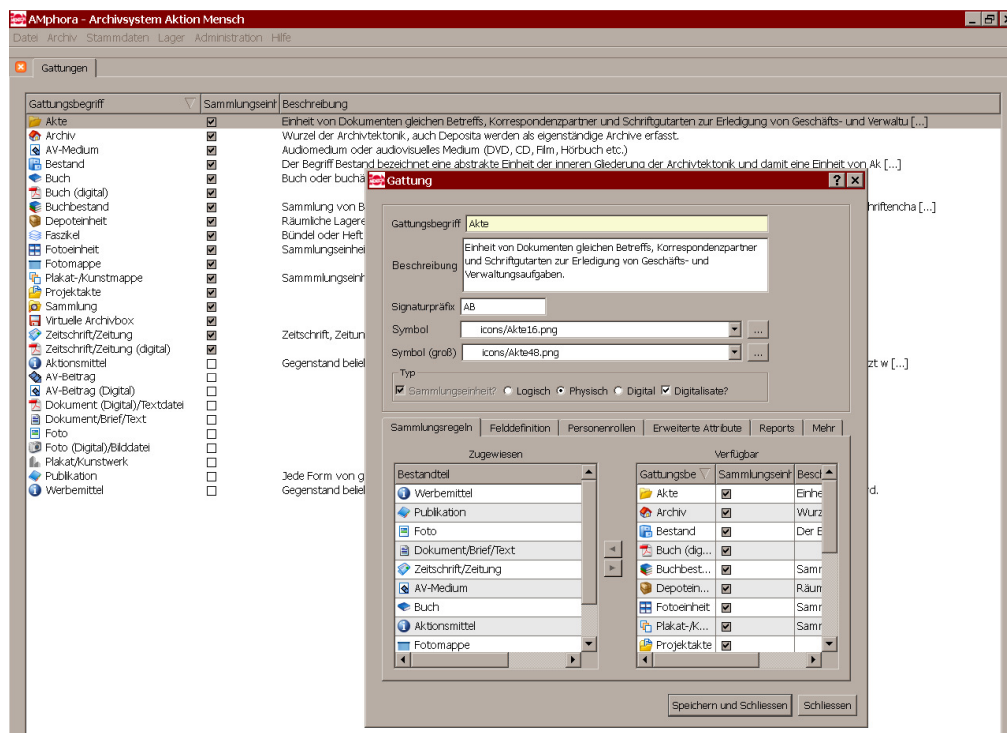


Abbildung 25: Ansicht einer Gattungsmaske

Personenverwaltung und -zuweisung, Personenrollen

Jede zu erfassende Person, die in Beziehung zu einer Archivalie steht, wird zunächst in die Personenliste aufgenommen. So ist sichergestellt, dass für Personen definierte Ansetzungsformen bestehen oder bei der Erfassung gebildet werden. Daher ist vor jeder Neuanlage einer Person zu prüfen, ob diese bereits eingetragen ist.

Bei der Personenerfassung werden natürliche und juristische Personen unterschieden. Juristische Personen können Ansprechpartner enthalten, die wiederum natürliche Personen sind. Zu den Personen sollten Adress- und Kontaktinformationen erfasst werden, soweit diese zum Zeitpunkt der Erfassung bekannt sind. Zu Personen können Notizen und Bilder, zu natürlichen Personen darüber hinaus biografische Angaben erfasst werden.

Wenn neue archivrelevante Informationen zu Personen bekannt werden, sollten diese während der Erfassung von Archivalien nachgetragen werden. Auf diese Weise wächst das Wissen über den Personenbestand im Archiv beständig an und erleichtert in Zukunft die Erfassung erheblich.

Bei der Erfassung der Archivalien wird die Person aus der Personenliste ausgewählt und es

wird mit Hilfe der Personenrolle bestimmt, in welcher Beziehung die Person zur Archivalie steht. Welche Rollen bei der Erfassung zur Verfügung stehen, wird von der verwendeten Archivaliengattung bestimmt.

Funktionen zur Verwaltung digitaler Objekte in AMphora

Digitale Archivalien (Virtuelle Archivboxen)

Digitale Archivalien (digitale Fotos, PDF-Dokumente u.a.) lassen sich in AMphora als Virtuelle Archivbox (analog der Gattung "Akte") speichern und erfassen. Um diese anzulegen, müssen die einzuspielenden Daten entweder auf einem tragbaren elektronischen Medium oder auf einem Laufwerk bereitstehen.

Dabei muss vor dem Einspielen darauf geachtet werden, dass für die LZA geeignete Dateitypen verwendet werden. Gegebenenfalls sind vorher notwendige Dateikonvertierungen durchzuführen.

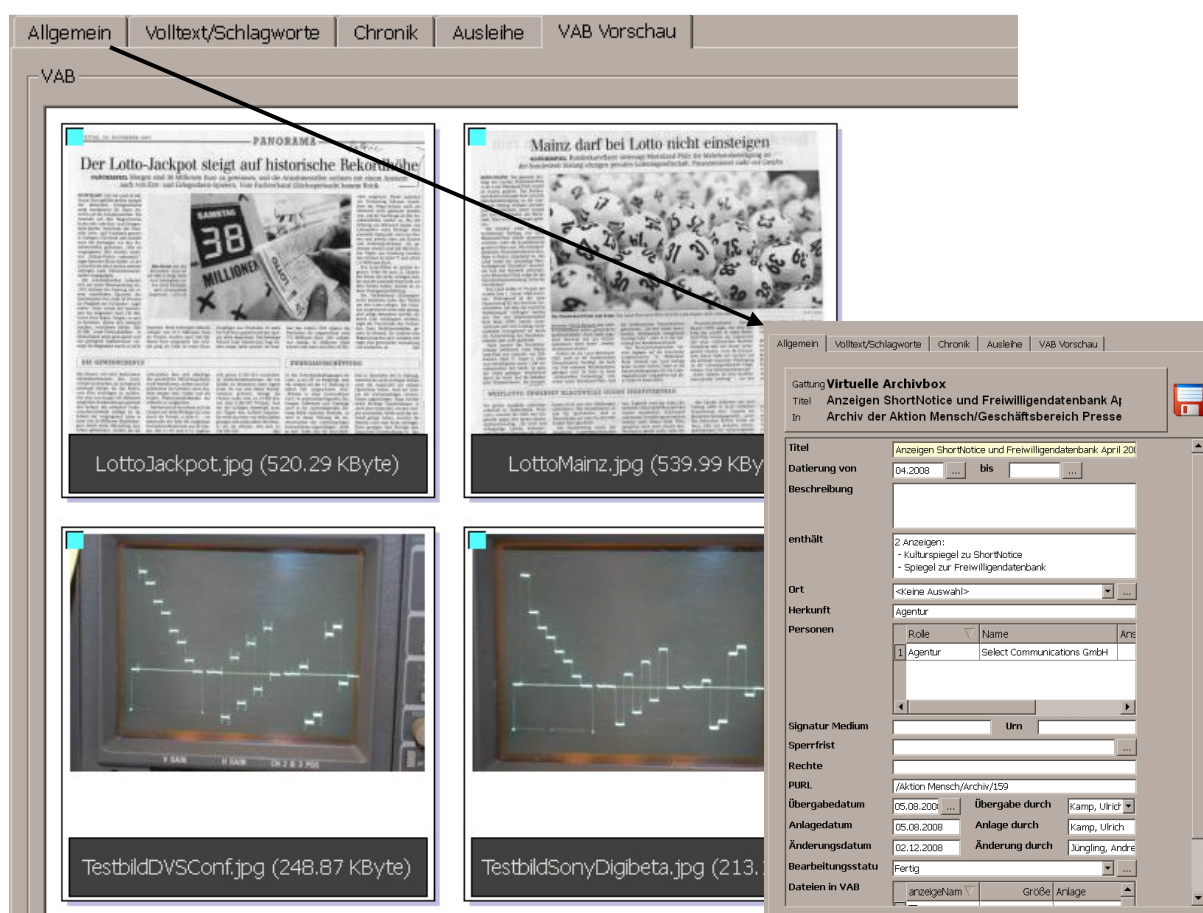


Abbildung 26: Beispiel für Virtuelle Archivbox mit Vorabsicht (Thumbnail)

LZA-Definition für digitale Dateiformate:

Um eine effiziente und langfristig orientierte Verwaltung der digitalen Archivalien zu gewährleisten, gibt es in AMphora besondere Funktionsmerkmale. Vor dem Archivieren digitaler Objekte können verwendete Dateiformate nach einem zuvor definierten Status bezüglich LZA-Eignung geordnet werden ("LZA-Status"). Abhängig von der firmeneigenen IT-Kapazität und den vorhandenen Ressourcen, von der Erwartung der Entwicklung bestimmter Dateiformate und auch von persönlichen Einschätzungen können die LZA-Status feingliedrig in verschiedene Stufen

aufgeteilt werden. Zur besseren und schnellen Orientierung wird dem LZA-Status ein frei bestimmbares Farbfeld zugewiesen, das in der Voransicht des Bildes als kleines Rechteck mit erscheint. Dieser Status wird der digitalen Archivalie mit dem Ingest zugeordnet. Die Liste der Dateiformate kann dazu erweitert werden, auch kann sich die Einschätzung einzelner Dateiformate im Laufe der Zeit ändern. Dann ändern sich gleichzeitig auch die LZA-Status der verwendeten Dateien und es kann eine Dateikonvertierung der archivierten Dateien angestoßen werden. Durch die Verbindung zwischen LZA-Status und Dateiformat wird so in der Praxis deutlich, welche Dateiformate aktuell noch beibehalten werden können; ggf. wird der Archivar implizit zu einem Formatwechsel aufgefordert.

Wesentlich ist vor allem, dass auf diese Weise eine Automatisierung der Konvertierung von Dateien ermöglicht wird.

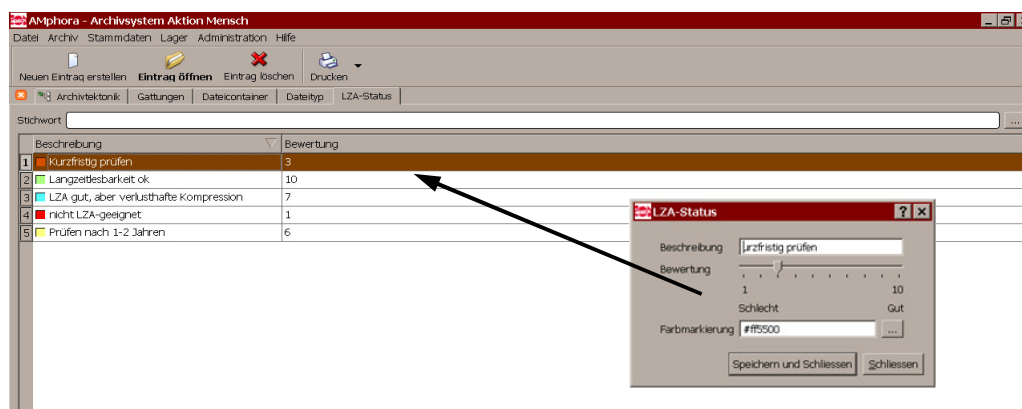


Abbildung 27: Definition der LZA-Status

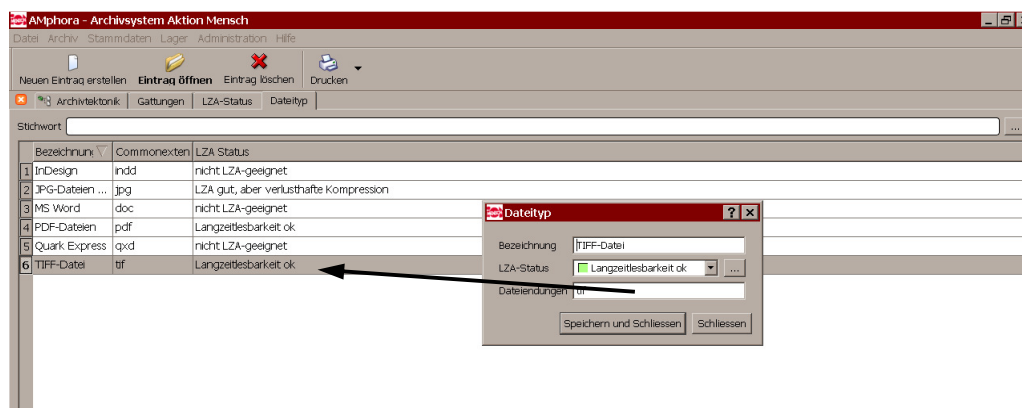


Abbildung 28: Definition der LZA-Status für Dateiformate

Beispiel:

Dem JPG-Format wird eine gute LZA-Erwartung zugeschrieben. Diese Zuschreibung wird erweitert mit dem Hinweis auf die formattypische verlusthafte Kompression. So lässt sich auch für den technischen Laien erkennen, welche massgeblichen Qualitätsunterscheidungen vorliegen, ob eine Migration in ein anderes Dateiformat oder auf einen anderen Datenträger angebracht ist.

Digitalisate

Im Gegensatz zu digitalen Archivalien stellen Digitalisate Ansichts- und Vorschaubilder zu Archivalien dar, die bei der Recherche der besseren Auffindbarkeit dienen. Für Digitalisate eignen sich in der Regel fotografierte oder gescannte Abbildungen der Archivobjekte, deren bildliche Auflösungsqualität aufgrund ihrer Ansichtsfunktion nicht hoch sein muss. Da es sich nicht um qualitativ hochwertige, zur weiteren Benutzung bestimmte und daher auch nicht um langzeitarchivierte Abbildungen handelt, werden sie folglich nicht im DMS gespeichert. Hierfür reicht eine Verknüpfung der Bilddatei mit den deskriptiven Metadaten der Archivalie für die Wiedergabe auf der Benutzeroberfläche von AMphora.

Die Möglichkeit, den jeweiligen Gattungen Digitalisate zuordnen zu können, wird über die Gattungsregeln bestimmt.

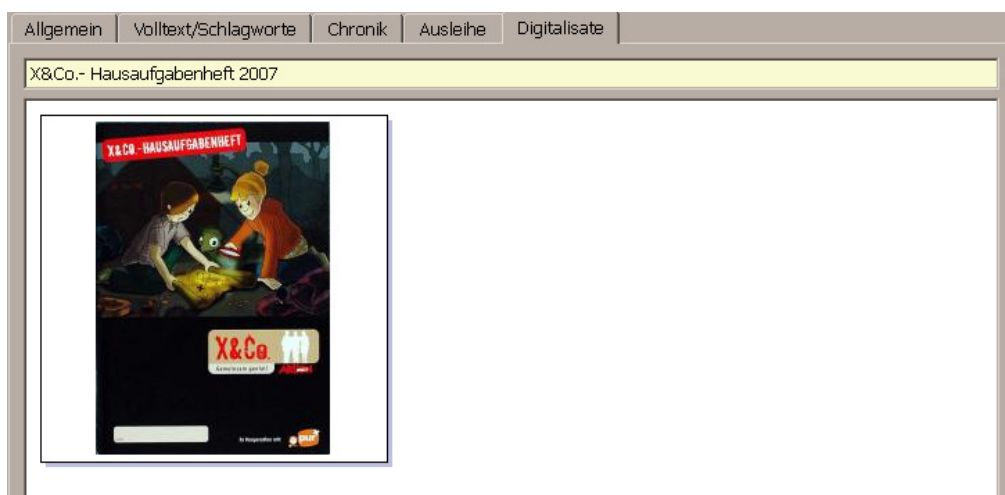


Abbildung 29: Beispiel für eine Digitalisatvoransicht

Nachwort

Mit dem Archivierungsprogramm AMphora steht allen Institutionen in Wirtschaft und Verwaltung ein in der Praxis zuverlässig erprobtes und bewährtes System zur Verfügung. Es gestattet auf der Basis eingeführter und allgemein verbindlicher Archivierungs- und Verzeichnungsstandards die freie, den jeweiligen Anforderungen und langfristigen Erwartungen wie Möglichkeiten in der Entwicklung der Archivwissenschaften angepasste Verwendung. Die Erfassung, Verzeichnung, Lagerung bzw. Speicherung digitaler Archivalien ist dabei eine ebenso natürliche Aufgabe wie die Verwaltung und Nutzung sogenannter "klassischer" Archivgüter. Beide Arten von Archivalien werden in AMphora prinzipiell gleich und gleichberechtigt behandelt. Vom allgemeinen ISAD/G-Standard der archivischen Stufenverzeichnung ausgehend lassen sich sämtliche Arten des Archiveinsatzes (z.B. Pressearchiv, Langzeitarchiv, Bildarchiv) in gleicher Weise berücksichtigen. Die nach dem Prinzip der Provenienz reglementierten Ordnungsvorbehalte, die durch die Stufenverzeichnung abgebildet werden, eignen sich daher für eine genaue historische wie nach sachlichen Gesichtspunkten orientierte Rekonstruktion der inneren Abhängigkeiten der Archivinhalte.

Die strikte Trennung, die in AMphora zwischen der technischen Bewältigung der gestellten Aufgaben und den archivischen Anforderungen besteht, gestattet es, die Integrität der grundle-

genden Verzeichnungsprinzipien zu wahren. Insbesondere die sich vielfach noch im Stadium des Experiments und der Erforschung wie Erprobung befindlichen Entwicklungen zur digitalen Langzeitarchivierung werden in AMphora auf struktureller Ebene mitgedacht. Die Lösungen, wie sie unter anderem mit der PURL-Signatur gefunden wurden, sind so einfach wie möglich gehalten. So orientiert sich diese PURL-Signatur am modernen Signaturbegriff der Archive und Bibliotheken, gibt allerdings anders als dieser keine sachlichen bzw. tektonischen Strukturvorgaben wieder. Im Sinne der eindeutigen und unverwechselbaren Einordnung in das Sammlungs Ganze und der Wiederauffindbarkeit erfüllt die PURL ihre Funktion. Wie sich im OAIS-Modell zeigt, führt die elektronische Datenverarbeitung auch zu erheblichen Funktions- und Prozessänderungen in Archiven. Das System AMphora zeigt jedoch auch, dass dabei grundlegende Archivprinzipien nicht über Bord geworfen werden müssen.

Das für AMphora grundlegende Systemmodell gestattet somit einerseits auf der IT- und andererseits auf der Verzeichnungsseite jeweils adäquate Lösungen, die die Weiterführung von Bewährtem oder die getrennte Anpassung an moderne Entwicklungen ermöglichen. Darüber hinaus ist es als modifizierbare ("agile") Lösung für Anpassungen an jede Art von Archiven offen. Der klassische und der künftig vorherrschende digitale Weg der Archivierung werden in AMphora zusammengedacht. Als Open-Source-Programm fordert es daher eine möglichst breite und vielfältige Anwendung und Mitarbeit zur Fortentwicklung heraus.

PADUA: Produktivsystem für die Archivierung digitaler Unterlagen im (Stadt-)Archiv Stuttgart⁶⁶

Katharina Ernst, Heike Maier

Der Beitrag bietet zunächst einen Überblick über die Funktionalitäten von PADUA. Die Strukturierung der Unterlagen und die Metadaten, die in PADUA erfasst, extrahiert und verwaltet werden, werden im zweiten Abschnitt ausführlicher beschrieben.

Funktionalitäten von Padua

Das Stadtarchiv Stuttgart stellt im Bereich der dLZA einen kleinen Player dar. Mit 17.5 festen Planstellen kann es sich in der Grösse mit den Bundes- und Landeseinrichtungen nicht messen. Obwohl es ein kleiner Player ist, gelten aber die gleichen Spielregeln und Anforderungen: Seine Bestände sind genauso komplex, die zu übernehmenden digitalen Daten sind genauso vielfältig, die Mengen sind genauso gross und seine AIPs müssen ebenso gut sein wie die der grossen Player.

Wir versuchen daher, modular einzelne Elemente des OAIS-Modells umzusetzen. Bei der Tagung des Arbeitskreises vor zwei Jahren in Stuttgart haben wir über unseren digitalen Speicher berichtet.⁶⁷ Den Bereich Speicherung haben wir also als Erstes abgedeckt. Wir wurden dabei von der Notwendigkeit getrieben, für die grossen digitalen Bestände, die uns angeboten worden waren und die wir als archivwürdig bewertet hatten, einen geeigneten Langzeitspeicher zu schaffen, damit wir unser digitales Archivgut überhaupt sicher ablegen konnten. Wir haben also mit dem virtuellen Magazin begonnen. Das Magazin war auch gut und verfügte über Klimatisierung und Regale, aber das Archivgut, das wir darin ablegten, war – um im Bild zu bleiben – weder entmetallisiert noch verpackt, weder signiert noch beschriftet. Eine langfristige Erhaltung des Archivguts war in diesem Zustand nicht möglich, ebenso wenig die perspektivische Benutzung der Bestände. Die Menge des Archivguts verbot eine Bearbeitung von Hand. Um nur die Extreme zu nennen: Der Bestand mit der grössten Zahl an Dateien besteht aus knapp 260.000 Dateien, der mengenmässig grösste Bestand umfasst knapp 770 GB. Solche Mengen lassen sich nur kontrolliert bearbeiten, wenn die Bearbeitung weitgehend automatisiert erfolgt. Eine Bearbeitung von Hand wäre auch zeitlich gar nicht zu leisten.

Mit diesem Problem sind wir in unserer Verwaltung vorstellig geworden und konnten 2008 eine Ingest-Anwendung mit integriertem Archiv-Repository ausschreiben. Wir haben den Auftrag an die Firma HP als Generalunternehmer mit der Firma SER als Subunternehmer vergeben.

Die Firmen HP und SER sind auch die Dienstleister, die das Digitale Archiv des Bundesarchivs aufgebaut haben. Die Ausschreibung des Stadtarchivs Stuttgart und die Auswahl des

⁶⁶ Der Vortragsstil wurde für die Publikation beibehalten.

⁶⁷ Vgl. Heike Maier, Der Langzeitspeicher des Stadtarchivs Stuttgart – ein strukturintegrativer Lösungsweg, in: Katharina Ernst (Hrsg.), Erfahrungen mit der Übernahme digitaler Daten – Bewertung, Übernahme, Aufbereitung, Speicherung, Datenmanagement, Stuttgart 2007 S. 52-57.

Dienstleisters erfolgten jedoch völlig unabhängig vom Bundesarchiv. PADUA deckt auch einen anderen Bereich ab als das Digitale Archiv des Bundesarchivs, das vor allem für das Übernehmen von Daten aus Dokumentenmanagementsystemen konzipiert ist. PADUA hingegen ist für alles ausser für elektronische Akten da: für Daten aus Fachverfahren, aus Datenbanken, Bilddateien, für digitale Nachlässe – also für alles das, was in unserer Stadt und im Verantwortungsbereich der meisten Archive schon seit vielen Jahren entsteht und was von vielen Archiven immer noch etwas stiefmütterlich behandelt wird.

Eine Ausbaufähigkeit in Richtung Dokumentenmanagement ist aber gegeben, und die Einführung von elektronischem Dokumentenmanagement befindet sich in der Stadt Stuttgart gerade in der Pilotierung. Die Einführung von DMS erfolgt hier ebenfalls durch die Firmen HP und SER.

PADUA wird die folgenden Funktionalitäten bieten, die teils vollautomatisch ausgeführt werden, teils halbautomatisch durch die Mitarbeiter und Mitarbeiterinnen des Stadtarchivs.

Strukturierung: SIPs, die beim Stadtarchiv anlangen, haben keine fest vordefinierte Struktur. Sie können ganz unterschiedlich aussehen. Viele sind völlig ungeordnet. Zunächst müssen die angelieferten Daten also strukturiert und mit administrativen Metadaten für die gesamte Lieferung versehen werden: Die Dateien müssen AIPs zugeordnet werden, die AIPs Taxonomieebenen, zwischen inhaltlichen Einheiten müssen Beziehungen hergestellt werden (vgl. hierzu die Ausführungen weiter unten). Dies findet im Strukturierungseditor statt. Er stellt eine eigenständige Anwendung dar, die von HP entwickelt wurde.

Aus dem Strukturierungseditor werden dann IIPs an die eigentliche Ingest-Anwendung übergeben. Die IIPs haben wir neu eingeführt, weil man nicht davon sprechen kann, dass SIPs weitergegeben werden, denn diese Daten haben ja bereits eine wesentliche archivische Bearbeitung erfahren. Die Ingest-Anwendung beruht auf dem Produkt PRODEA der Firma SER.

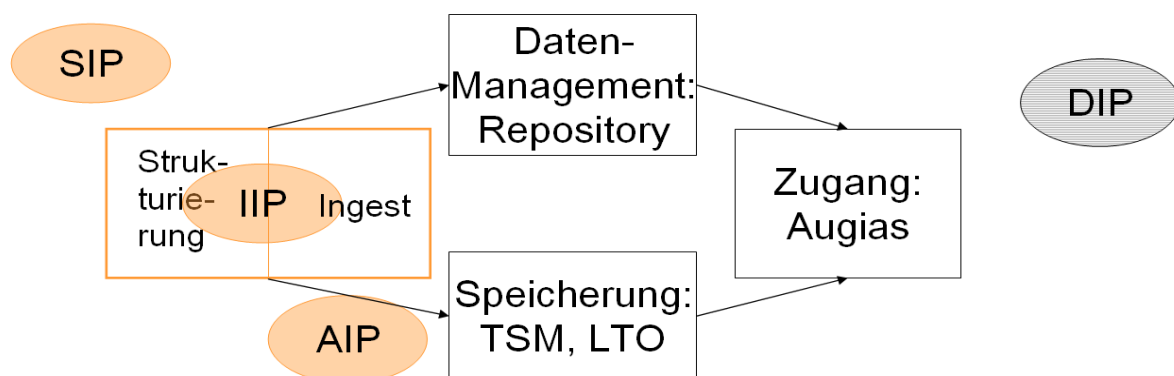


Abbildung 30: Ingest Information Packages (IIP)

Akzessionierung: Hier werden folgende Aktionen automatisch ausgeführt: eine Eingangsprüfung des IIP auf Datenkonsistenz und Datenintegrität, eine Formaterkennung mit DROID und eine Formatprüfung mit JHOVE. Das System trägt eine Reihe von Metadaten in die XML-Datei ein, die zum jeweiligen AIP gehört, z.B. die Ergebnisse der Formaterkennung und Formatprüfung, den Hashwert, den Original-Dateinamen, die Dateigrösse. Metadaten, die bereits im

Strukturierungseditor vom Stadtarchiv für die ganze Lieferung vergeben wurden, werden falls vom System automatisch zu jedem AIP in die XML-Datei geschrieben. Die Archivsignaturen werden erzeugt. Die ganze Lieferung geht dann weiter zum Prozessschritt Bearbeitung.

Bearbeitung: In dieser Phase werden von Hand weitere Metadaten ergänzt. Eine detaillierte Verzeichnung soll nicht in PADUA erfolgen, sondern perspektivisch in Augias. Wir wollen aber den AIPs zumindest ein Minimum an archivisch-deskriptiven Metadaten mitgeben, so dass jedes AIP so autonom wie möglich ist und alle notwendigen Informationen zu seinem Verständnis beinhaltet. PADUA unterstützt dabei die Bearbeiter auf verschiedene Weise, so ist beispielsweise das automatisierte Vererben von Informationen möglich.

An die Bearbeitung schliesst sich die **Transfer-** und **Archivierungsphase** an, in der alle Daten dort hingeschoben werden, wo sie dauerhaft bleiben sollen. Das ist zunächst einmal das Archivgut: Jedes AIP, bestehend aus einer oder mehreren Primärdateien und einer XML-Datei mit allen Metadaten, wird in ein TAR-File gepackt und über den Tivoli Storage Manager auf unseren Langzeitspeicher geschoben. Dann sind da die Metadaten, die man ja nicht nur gut verpackt in der XML-Datei des jeweiligen AIP haben möchte, sondern auch recherchierbar in einer Datenbank. Sie werden also in das Repository geschrieben. Die Metadaten werden ausserdem in ein Verzeichnis ausgespielt, wo sie zum Import in andere Anwendungen (Augias) bereitstehen. Der erfolgreiche Transfer aller dieser Daten wird vom System überprüft. Dann erst werden die Daten im System gelöscht und die Ingest-Anwendung wird von den Daten bereinigt.

Was die erforderliche technische Ausstattung betrifft, so wird die zentrale IT-Abteilung der Stadt Stuttgart drei Server nur für PADUA betreiben. Auf Anwenderseite benötigen wir Rechner mit 4 GB Hauptspeicher und einem 64bit-Betriebssystem. Das Stadtarchiv gehört damit zu den technisch anspruchsvollsten Anwendern in der Stadtverwaltung.

Dieser Überblick stellt nicht den Endzustand von PADUA dar. Das Stadtarchiv wird weiterhin modular vorgehen und nach seinen Bedürfnissen und Möglichkeiten weitere Module hinzufügen. Insbesondere betrifft dies zum einen die automatisierte Konvertierung, denn eine Konvertierung von Hand ist bei unseren Beständen allein von der Grösse her nicht möglich. Zum anderen betrifft es den Zugang zum Archivgut, das DIP.

Strukturelle Organisation und Metadaten digitaler Unterlagen im Stadtarchiv

Stuttgart

Strukturrahmen

Das Stadtarchiv übernimmt zurzeit digitale Unterlagen, die in einer vom Lieferanten bestimmten Struktur vorliegen (SIPs)⁶⁸. Die Ingest-Anwendung PADUA benötigt aber strukturierte Ein-

⁶⁸ Es werden keine Akten übernommen, d.h. es gibt keine Aktenstruktur (Akte – Vorgang – Dokument) und kein einheitliches Datenaustauschformat zwischen Lieferant und Archiv, wie dies für Akten mit XDOMEA existiert.

gangslieferungen. Daher ist der Ingest-Anwendung PADUA ein Strukturierungseeditor vorge-schaltet. Hier wird jede Lieferung von Objekten durch das Archiv so strukturiert (und auch die AIPs bereits gepackt), dass die Ingest-Anwendung damit umgehen kann (IIPs).

Strukturierung heisst hier: "Objekte in eine logische hierarchische Ordnung zu bringen und Beziehungen zwischen den Entitäten (logische Einheiten) und Objekten (physischen Einheiten = Datei) zu erstellen, die notwendig für das Verständnis und zur Bestandserhaltung des Inhaltes sind."

Hierarchisch besteht die Möglichkeit, eine Lieferung in fünf Taxonomiestufen (Bestand, Serie, Subserie, Kollektion, Subkollektion) einzuordnen⁶⁹. Mindestens zwei Taxonomiestufen sind zwingend (Bestand und Serie)⁷⁰. Auf der untersten Taxonomiestufe einer Lieferung befinden sich die Primärobjekte (Dateien, die die archivischen Inhalte beinhalten).

Im Strukturierungseeditor können in der hierarchisch geordneten Lieferung zwei Arten von Beziehungen definiert werden:

Kontextbeziehung

Mit der Kontextbeziehung wird ausgedrückt, dass eine Entität zum Verständnis eine andere Entität benötigt. Kontextbeziehungen werden im Stadtarchiv Stuttgart nur zwischen logischen Einheiten vergeben.

Repräsentationsbeziehung

Mit der Repräsentationsbeziehung wird ausgedrückt, dass ein Objekt denselben Inhalt in einer anderen Daseinsform repräsentiert. Repräsentationsbeziehungen entstehen immer bei Migration von Objekten.

Konkretisierung des Strukturrahmens am Beispiel der Daten zu Zeitzeugeninterviews im Rahmen des Projekts Zeitgeschichte

Beispiel:

In den 80er-Jahren wurde in Stuttgart beschlossen, eine stadtgeschichtliche Sammlung einzurichten. Im Rahmen dieses Projekts wurden verschiedene zeitgeschichtliche Ausstellungsprojekte realisiert, die differenzierte Blicke auf Stuttgart in den 30er- und 40er-Jahren eröffneten. Exemplarisch soll hier auf einen Teil der Tondokumente, auf denen Zeitzeugeninterviews festgehalten sind, eingegangen werden. Diese Zeitzeugeninterviews wurden im Original auf analogen Tonkassetten aufgenommen.⁷¹ Als bestandserhaltende Massnahme wurde diese durch den städtischen Sitzungsdienst auf CD digitalisiert. Diese CDA-Dateien wurden im Stadtarchiv in WAV- und MPEG-Dateien gerippt. Zu den Primärdateien – den Interviews – liegen zwei Dokumentationsdateien in PDF/A vor: Das erste Dokument ist ein Akzessionierungsdokument,

⁶⁹ Es müssen nicht alle Taxonomiestufen belegt werden, d.h. es gibt Lieferungen, die sich in drei Taxonomiestufen einordnen, andere in fünf.

⁷⁰ D.h. ein AIP kann nie direkt unter einem Bestand hängen. Jedes AIP benötigt eine Entität über sich, die den Inhalt des AIPs beschreibt.

⁷¹ Tonkassetten sind ca. 10–25 Jahre lang haltbar.

das den formalen Rahmen der Übernahme und den Prozess der Digitalisierung beschreibt. Das zweite Dokument besteht aus einer Liste der Interviewten und weiteren formalen inhaltlichen Angaben, die für alle Interviews zuträfen. Der Bestand selbst setzt sich aus umfangreichen analogen Unterlagen zusammen (Foto, Filme, Ausstellungsdokumentation etc.), die im Rahmen des Projekts entstanden sind bzw. gesammelt wurden.

Strukturierung

Im Stadtarchiv wird das gesamte digitale Interviewmaterial aus WAV/MPEG-Dateien und den beiden PDF/A-Dokumentationsdateien als eine Lieferung betrachtet, die dem Bestand Zeitgeschichte zugeordnet wird.

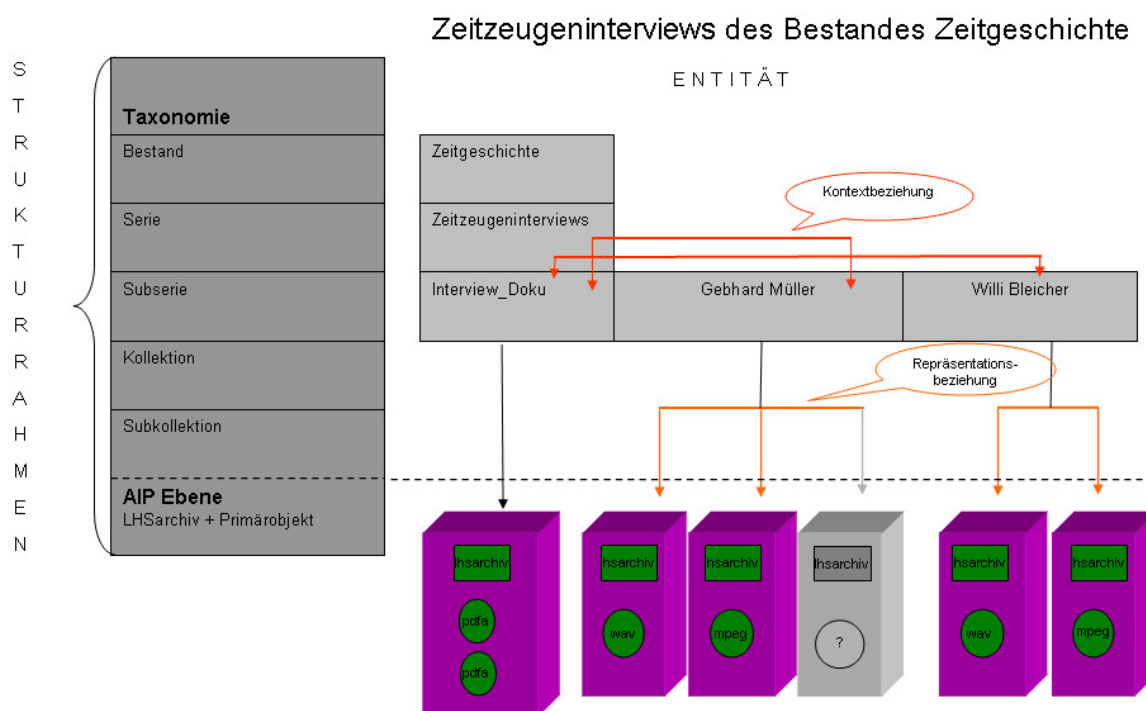


Abbildung 31: Zeitzeugeninterviews – Entität und Strukturrahmen

Die Zeitzeugeninterviews werden hierarchisch in den Bestand eingegliedert und bilden darin eine Serie wie auch die Fotos, Filme etc.⁷² Die Serie Zeitzeugeninterviews wird weiter hierarchisch unterteilt in Subserien. Dies ist zum einen eine Subserie, die das Dokumentationsmaterial repräsentiert, und zum anderen eine Subserie, die aus den einzelnen Interviews besteht.⁷³

⁷² Dieser Bestand wird stark nach Medientypen strukturiert, dies ist aber nur eine Möglichkeit der Strukturierung.

Eine stärker inhaltliche Strukturierung wäre theoretisch denkbar gewesen, z.B. hätte man einem Zeitzeugen auch Fotos zuordnen können. Diese Zuordnung wäre aber bei dem gesamten Bestand nicht sinnvoll gewesen, d.h. eine Strukturierung orientiert sich immer an verschiedenen Gesichtspunkten und ist pro Lieferung individuell zu betrachten.

⁷³ In diesem Fall liegen drei Taxonomiestufen vor. Das Stadtarchiv hat noch die Möglichkeit, mit der Kollektion und der Subkollektion zwei weitere Hierarchiestufen zu nutzen.

Den einzelnen Entitäten – logischen inhaltlichen Einheiten – sind die Stücke zugeordnet, die in AIPs verpackt sind. Konkret sind in diesem Fall der Entität "Interview mit Gebhard Müller" zwei Stücke zugeordnet, nämlich eine WAV- und eine MPEG-Datei, die jeweils getrennt in AIPs gepackt worden sind. Diese beiden Stücke sind Repräsentationen der Entität "Interview mit Gebhard Müller". Sie repräsentieren beide den gleichen Inhalt, nur in unterschiedlichen Daseinsformen. Weitere Repräsentationen, die durch Migrationen entstehen, sind denkbar (im Schaubild grau dargestellt). Die Repräsentationsbeziehung wird über das PREMIS Metadatenfeld `relationshipSubType` mit dem Ausdruck "hasSibling" ausgedrückt.⁷⁴

Neben der Repräsentationsbeziehung, die das Stadtarchiv Stuttgart als technische Beziehung definiert, gibt es eine weitere für den Inhalt der Entitäten notwendige Beziehung. Diese sogenannte Kontextbeziehung drückt aus, dass zum Verständnis einer Entität eine andere Entität benötigt wird. Die Entität "Interview mit Gebhard Müller" benötigt, um vollständig verstanden zu werden, die Entität, die das Dokumentationsmaterial enthält.⁷⁵ Diese Beziehung wird im Bereich der deskriptiven Metadaten durch den Ausdruck "requires" im Element `relationTypeEntity` dargestellt.

Die Kontextbeziehung wird nur auf der Ebene der Entität, der logischen Einheit, ausgedrückt⁷⁶. Vorteil dieser Vorgehensweise ist, dass erstens die Zahl der Referenzierungen nicht zu umfangreich wird und zweitens diese Beziehung – die eine inhaltliche ist – stabil ist, d.h. egal welche Migrationsprozesse im Laufe der Zeit mit dem Material vorgenommen werden, immer wird die Entität "Interview mit Gebhard Müller" zum vollständigen Verstehen die Entität "Interview Doku" benötigen.

Bestellbares Archivgut ist die Entität – die logische Einheit –, die oberhalb des AIPs liegt.⁷⁷ Ein AIP besteht in der Regel aus mindestens einer Primärdatei und einer Metadatendatei im XML-Format. Das Metadatenschema wird nachfolgend dargestellt.

⁷⁴ Bislang kann im Stadtarchiv Stuttgart das Repräsentationsmodell konzeptionell nur im Bereich der digitalen Unterlagen umgesetzt werden. Die Tonkassette und auch die Transkription des Interviews – beides analoge Objekte – sind logisch ebenso Repräsentationen der Entität "Interview mit Gebhard Müller". Eine Anwendung, in der digitale und analoge Objekte gemeinsam nach dem Repräsentationsmodell verwaltet werden können, gibt es in Stuttgart bislang nicht.

⁷⁵ Eine noch deutlichere Kontextbeziehung ist bei den Wahldaten sichtbar. Die eigentliche Wahldatei kann nur interpretiert werden, wenn man die Datei Strassenverzeichnis, Wahlbezirkseinteilung, Urnen-Briefwahlbezirke verfügbar hat, denn in der Wahldatei sind nur Schlüssel, aber keine inhaltlich verstehbaren Angaben hinterlegt.

⁷⁶ Theoretisch möglich ist es auch, die Kontextbeziehung auf der Ebene des konkreten Stückes anzusiedeln, z.B. "das Stück Gebhard Müller.wav benötigt die Stücke Interview_Doku1.pdf und Interview_Doku2.pdf". Hier muss dann aber pro Daseinsform (Repräsentation) die Beziehung dargestellt werden. Dies vervielfacht die Referenzierung.

⁷⁷ Es wird dabei davon ausgegangen, dass der Nutzer primär einen Inhalt sucht und nicht eine bestimmte Daseinsform.

LHSArchiv

LHSArchiv ist das Metadatenschema des Stadtarchivs Stuttgart. Die Metadaten dieses Schemas dienen der Verwaltung der digitalen Bestände wie auch den bestandserhaltenden Massnahmen. Diese Metadaten, die in Form einer XML-Datei vorliegen, und das Primärobjekt bzw. die Primärobjekte werden zusammen in ein AIP gepackt; damit ist das AIP weitgehend autonom und selbsterklärend. Die Metadaten des Schemas werden natürlich auch in das Indexmanagement des Repositorys übernommen und ermöglichen so eine differenzierte Recherche nach digitalen Unterlagen.

Für das LHSArchiv wird kein standardisiertes logisches Packaging-Format wie METS oder XFDU benutzt. Das Stadtarchiv hat sich entschieden, keinen der beiden Standards einzusetzen⁷⁸, sondern in einer ersten Version des LHSArchivs eine einfache Lösung umzusetzen.⁷⁹ Für die digitale Langzeitarchivierung ist ein logisches Standard Packaging-Format keine zwingende Voraussetzung.

Das LHSArchiv besteht aus ca. 250 Elementen, davon sind 50% den deskriptiven⁸⁰ Metadaten zuzuordnen, 38% den bestandserhaltenden, 12% sind administrativ bzw. AIP-bezogen. Das LHSArchiv besteht aus folgenden 4 Hauptcontainern:

1. AIP Metadata
2. Admin Metadata
3. Technical Metadata
4. Descriptive Metadata

1. AIP Metadata:

AIP Metadaten haben die Funktion eines Paketaufklebers auf dem digitalen Archivpaket. Ähnlich dem Magaziner, der in einem Karteikasten den Lagerort des bestellten Archivgutes ermittelt und dann im entsprechenden Regal den Karton mit der passenden Kartonnummer aushebt, wird das digitale Archivpaket aufgrund seiner Metadaten von seinem Speicherort gesucht und gefunden.

2. Admin Metadata:

Administrative Metadaten sind alle Daten rund um die Akzessionierung, vorwiegend werden hier Angaben über die abgebenden Stellen erfasst.

⁷⁸ Keller-Marxer, Peter: Ingest und Digitale Langzeitarchivierung im Stadtarchiv Stuttgart, Bern 2007, S. 41f. (unveröffentlicht).

⁷⁹ LHSArchiv_1.0 wird im Stadtarchiv gemäss seinen Bedürfnissen weiterentwickelt. Prinzipiell ist der Einsatz von METS oder XFDU weiterhin möglich.

⁸⁰ Die Zahl der deskriptiven Metadaten ist so gross, weil auf jeder Taxonomieebene alle deskriptiven Metadaten verfügbar sind.

3. Technical Metadata:

Technische Metadaten beschreiben die Objekte in ihrer technischen Beschaffenheit. Bei den technischen Metadaten verwendet das Stadtarchiv den Standard PREMIS 1.0⁸¹. Nicht verwendet werden die Container "rights" und "signature"⁸² PREMIS Metadaten werden sowohl durch Anwendungswerkzeuge extrahiert wie auch manuell vergeben.

4. Deskriptive Metadata:

Das Stadtarchiv Stuttgart verwendet für seine analogen Unterlagen keinen festgelegten Metadatenstandard zur Erschliessung. Es wurde anhand der bisher verzeichneten analogen Unterlagen geprüft, welche Metadaten real im Stadtarchiv vergeben werden. Hierbei wurde festgestellt, dass diese weitgehend dem DC-Standard entsprechen. Obwohl dieser Standard kein archivischer ist und auch nicht für Unterlagen dieser Art gemacht wurde, hat sich das Stadtarchiv entschlossen, sich für seine digitalen Unterlagen im ersten Schritt an diesem Standard zu orientieren. Im Laufe der Entwicklung wurde dabei deutlich, dass Dublin Core-Felder nicht ausreichend sind, daher wurden weitere Felder ergänzt, wobei das Stadtarchiv sich in der Syntax an PREMIS und nicht an den DC-Standard anlehnt.

Ausblick

Nach einem Jahr konzeptioneller und administrativer Arbeit im Rahmen der digitalen Langzeitarchivierung sind die Ziele für 2009/2010 sehr praktisch angelegt. Die Ingest-Anwendung PADUA und der Strukturierungseitor sollen produktiv genutzt werden.

Zum einen müssen hier die bereits im Stadtarchiv vorhandenen digitalen Bestände, die momentan nur in einer Ordnerstruktur auf dem Langzeitspeicher (LTO3 WORM) abgelegt sind, vom Speicher abgerufen und mit dem Strukturierungseitor und PADUA archivisch bearbeitet werden, um dann wieder auf LTO3 archiviert zu werden⁸³. Zum anderen muss der Übernahmestau im Bereich der digitalen Objekte bei den Fachämtern abgearbeitet werden. Organisatorisch ist es notwendig, die Ingest-Anwendung PADUA in die Arbeitsprozesse des Stadtarchivs zu integrieren. Diese Phase der Konsolidierung ist notwendig, um mit den hier gewonnenen Erfahrungen – dies hat sich im bisher modularen Ansatz gezeigt – das nächste Modul "Accession" zu realisieren.

Beispiele für Metadaten im LHSarchiv

AIP Metadata:

archivalIdentifierOfAIP – AIP Signatur

titleOfLevelOverAIP – AIP Titel

⁸¹ <http://www.loc.gov/standards/premis/>.

⁸² Rechte werden in den archivischen Metadaten dargestellt. "Signature" erfährt keine Verwendung, da das Stadtarchiv prinzipiell keine signierten Unterlagen übernimmt.

⁸³ Damit werden erste Migrationserfahrungen im digitalen Bereich gemacht.

Admin Metadata:

dateSubmitted – Eingangsdatum der Lieferung

provenanceName – Name der abgebenden Stelle

Technical Metadata:

Extrahiert:

formatName – Formatnaame

size – Grösse

messageDigest – Wert des Algorithmus

manuell

significantProperties – Spezifische Eigenschaften, die über Migrationen hinweg erhalten werden sollen

relationshipSubType – Beziehungstyp

swName – Softwarename

arcun. Überlegungen zur digitalen Langzeitarchivierung als Dienstleistungsangebot

Martin Kaiser

Abstract

Die KOST beabsichtigt mit dem Projekt *arcun* (rätoromanisch für "Speicher"), ihren Mitgliedarchiven (kommunale, kantonale und nationale Archive) eine preisgünstige Gesamtlösung für die dLZA zur Verfügung zu stellen. Die KOST übernimmt dabei Spezifikation, Auditing und Inkasso für die gewählte Lösung. Der Dienstanbieter stellt ein mandantenfähiges digitales Langzeitarchiv zur Verfügung.

In einem ersten Teilprojekt soll nur die Speicherung im Sinne von Bitstream Preservation als Dienstleistung angeboten werden. Dieses Teilprojekt befindet sich in der Realisierungsphase und soll bis Ende 2009 abgeschlossen sein.

In einem zweiten Teil *arcun+* soll der gesamte Archivierungsprozess als Dienstleistung angeboten werden. Es geht hier um eine Art "Archivsteckdose", über welche die Archive Ablieferungen in ein zentrales dLZA einstellen können und über die auf die einzelnen Unterlagen in diesen Ablieferungen jederzeit wieder zugegriffen werden kann. Der Dienst umfasst dabei die Bitstream Preservation, die Sicherstellung der Lesbarkeit der Datenobjekte sowie die Kontrolle über die notwendige Anreicherung mit Metadaten.

Dieses Fortsetzungsprojekt befindet sich noch in der Planungsphase und soll als Ergebnis Spezifikationen für Prozesse und Schnittstellen einer umfassenden Archivierungslösung liefern.

Einleitung

In einer Kolloquiumsreihe "Archivtaugliche Speicherinfrastruktur" haben sich die in der KOST zusammengeschlossenen Archive im Herbst 2007 mit der Frage beschäftigt, wie ein archivtaugliches Speicherkonzept auszusehen hätte. Im Vordergrund stand dabei als Erstes einzig die Speicherung der archivierten Unterlagen, also die sogenannte *Bitstream Preservation*⁸⁴. Aus diesen Überlegungen ist das Projekt *arcun* entstanden. Ziel und Zweck des Projektes ist es, den Mitgliedarchiven preisgünstigen Speicherplatz für die digitale Langzeitarchivierung zur Verfügung zu stellen. Der für diese Lösung notwendige Speicherplatz soll als Mietspeicherplatz für vorläufig fünf Archive bis Ende 2009 realisiert werden. Im ersten Teil dieses Artikels werden die grundlegenden Überlegungen, die zu dieser Lösung geführt haben, erläutert.

In der Folge hat sich schnell auch die Frage gestellt, ob, wenn die Speicherung gemeinsam betrieben werden kann, nicht auch der gesamte archivische Prozess inklusive Bewirtschaftung der Unterlagen (Metadatenverwaltung, Migrationen etc.) der digitalen Archivierung in eine gemeinsame Lösung ausgelagert werden könnte. Dieses weiterführende Projekt *arcun+(plus)*

⁸⁴ Die *Bitstream Preservation* ist die Sicherung der Daten auf der physischen Ebene (Datenträgermigration: Kopieren auf neue Datenträger und Anpassen an neue Speicher-Technologien); sie ist die Grundlage aller weiteren Archivierungsaktivitäten.

befindet sich noch in der Planungsphase. Im zweiten Teil dieses Artikels wird auf die Details dieses Projekts eingegangen

***arcun* – Analysekriterien für eine archivtaugliche Speicherinfrastruktur**

Archivische Anforderungen

Die Rahmenbedingungen für eine archivtaugliche Speicherinfrastruktur werden definiert durch die archivischen Anforderungen. Diese Anforderungen unterscheiden sich zum Teil erheblich von den Lösungen, welche die Informatikindustrie für operative Datenhaltung und Backup zur Verfügung stellt:

- Datenintegrität und Datensicherheit sind absolut zentral.
- Eine tiefe Verfügbarkeit ist tolerierbar: Nachts und am Wochenende können die Daten problemlos *offline* sein; es können auch Ausfalltage verkraftet werden.
- Zugriffszeiten im Minutenbereich, wie sie *Nearline*-Speicherung⁸⁵ bietet, sind im Archivbetrieb akzeptabel.
- Es gibt keine Transaktionen auf den archivierten Daten, deshalb sind die *Backup*-Anforderungen einfach zu erfüllen.
- Der Zugriff auf die Daten erfolgt arbiträr; ein hierarchisches Speichermodell, d.h. verschiedene Arten von Speicherplatz mit unterschiedlichen Zugriffszeiten, macht deshalb wenig Sinn.

Es existieren kaum Speicherkonzepte und -angebote, die auf dieses Anforderungsprofil zugeschnitten sind, und es besteht die Gefahr, die falsche Leistung einzukaufen. Das archivische Anforderungsprofil ist also das Werkzeug bei der Analyse und Beschaffung von Speicherlösungen.

Datenvolumen

Zurzeit sind die digitalen Bestände in den Mitgliedsarchiven der KOST noch klein. Verschiedene Erhebungen der KOST haben ergeben, dass pro Archiv maximal mit digitalen Archivalien im Umfang von 2 TB und einem jährlichen Zuwachs im dreistelligen GB-Bereich gerechnet werden muss. Eine Gemeinschaftslösung für 5-10 Archive wird folglich zu Beginn nicht grösser als 10 TB dimensioniert sein müssen.

⁸⁵ *Nearline*-Speicherung ist zwischen *Online* (Sofortzugriff, hohe Kosten) und *Offline* (ausgelagert, günstig) angesiedelt; in der Regel handelt es sich dabei um Systeme, die auf *Tape-Libraries* basieren.

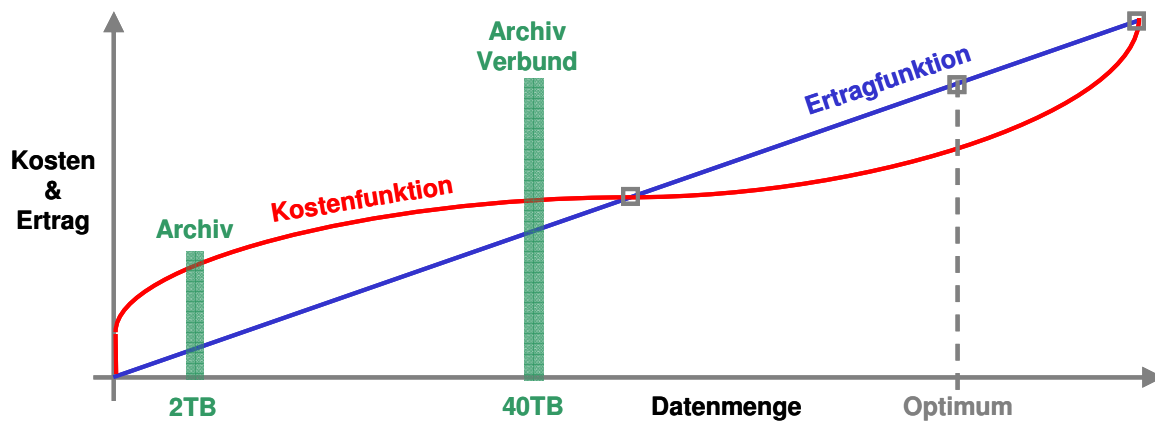


Abbildung 32: Datenspeicherung unterliegt den Gesetzmässigkeiten der Economies of Scale

Organisatorisch-rechtliche Aspekte

Die erwähnten Datenmengen liegen unter der kritischen Masse, die bei verschiedenen Lösungen für einen wirtschaftlichen Betrieb nötig wäre. Gerade die Datenspeicherung unterliegt einem Skaleneffekt; eine interkantonale oder auch eine innerkantonale Zusammenarbeit ist deshalb unbedingt zu prüfen. Zwar könnte eine interkantonale Lösung im Konflikt mit heutigen rechtlichen Vorgaben stehen, diese Hindernisse müssen aber in einem politischen Prozess überwunden werden. Zu beachten sind daneben mögliche "psychologische" Probleme, die mit der Datenspeicherung ausserhalb der physischen Kontrolle des Archivs oder der kantonalen Verwaltung verbunden sind.

Kosten

Die Kosten für die digitale Langzeitarchivierung befinden sich im Fluss und variieren sehr stark zwischen den Anbietern. Es ist deshalb nicht einfach, längerfristig verbindliche Prognosen aufzustellen. Ein Kostenvergleich zwischen *Bitstream-Preservation*-Anbietern zeigt erstaunliche Unterschiede, vom verbindlichen Angebot eines kantonalen Informatikdienstleisters für 30'000 Fr./TB/Jahr bis zum S3-Speicherdienst von Amazon für 2'400 Fr./TB/Jahr (Zahlen von Anfang 2008).

arcun – Unterschiedliche Lösungen

Im Rahmen des KOST-Kolloquiums "Archivtaugliche Speicherinfrastruktur" haben wir fünf unterschiedliche Lösungen für die reine *Bitstream Preservation* betrachtet. Im Detail folgen hier die einzelnen Lösungsvarianten und die Bewertung aus Sicht der oben dargelegten Kriterien.

Speicherplatz mieten

Das Archiv mietet den benötigten Speicherplatz bei einem (kommerziellen oder öffentlichen) Anbieter. Der dabei ins Auge gefasste Zeithorizont umfasst etwa fünf Jahre. Der Anbieter ist für die Datenpersistenz verantwortlich; das Archiv hat damit nichts zu tun. Die Anforderungen des

Archivs werden in einem *Service Level Agreement* (SLA)⁸⁶ festgehalten. Die kommerziellen Angebote sind sehr unterschiedlich gestaltet und müssen deshalb auf Archive massgeschneidert werden. Dies verteuert den Speicherplatz bei kleinem Volumen.

Speicherzentrum

Ein Speicherzentrum ist ein Rechenzentrum, das primär Dienstleistungen im Bereich der Datenspeicherung und Langzeitarchivierung erbringt. Dabei ist Folgendes zu beachten.

- Der Aufbau eines Speicherzentrums setzt ein minimales Datenvolumen voraus und darf in seiner Komplexität nicht unterschätzt werden. Grundsätzlich ist vom Neuaufbau eines Rechenzentrums abzuraten. Das Mieten bestehender Rechenzentrumsinfrastruktur scheint sinnvoller.
- Das Beherrschen der Hardwaremigrationszyklen ist eine der grössten Herausforderungen für ein Rechenzentrum im Bereich der Langzeitarchivierung.
- Eine klare Trennung von Speichermanagement im Speicherzentrum und Archivsoftware unter Aufsicht des Archives ist unerlässlich.

Blackbox

Unter einer Blackbox verstehen wir eine Speicherbox, d.h. ein physisch greifbares Gerät, das im Archiv selber oder bei einem Raumanbieter stehen kann. Die interne Funktionsweise dieser Box ist unbekannt und irrelevant. Wichtig ist, dass der Benutzer mit der Box über ein API (Application Programming Interface [Programmierschnittstelle]⁸⁷, meist proprietär) kommunizieren kann, um Daten abzuspeichern und wieder anzufordern. Die Datenpersistenz wird vom Hersteller garantiert; vom Anwender muss alle paar Jahre nur die Migration auf eine neue Box gelöst werden. Ein Hauptvorteil der Blackbox ist der relativ geringe Administrationsaufwand im laufenden Betrieb. Ein weiterer, vielleicht eher psychologischer Vorteil kann die physische Kontrolle über die Speicherinfrastruktur sein.

Datenauslagerung

Die digitalen Daten werden auf ein langzeitstabiles Speichermedium umkopiert und in dieser Form aus der operativen Informatikinfrastruktur ausgelagert. Während der Phase der Auslagerung fallen nur die Kosten für die Lagerung des Trägermediums an. Diese Lösung wird schon länger diskutiert und eingesetzt, z.B. bei der COLD-Archivierung (*Computer Output to Laser Disk*), erlangt aber mit dem Einsatz von Mikrofilm als Trägermedium eine neue Dimension bezüglich Auslagerungszeitraum.

Mit den heutigen Methoden können auf einen 35-mm-Film von 600 m Länge (entspricht etwa einer Rolle von 40 cm Durchmesser) ungefähr 250 GB Nutzdaten geschrieben werden. Eine

⁸⁶ Ein Service-Level-Agreement (SLA) oder eine Dienstgütevereinbarung (DGV) legt fest, mit welchen Leistungseigenschaften (Ausfallzeit, Reaktionszeit, Geschwindigkeit etc.) eine wiederkehrende Leistung angeboten wird.

⁸⁷ Siehe <http://de.wikipedia.org/wiki/Programmierschnittstelle>.

Standardisierung sowohl des Informationslayouts auf dem Mikrofilm als auch der notwendigen Prozesse hat bis jetzt noch nicht stattgefunden. Die aktuellen Auslagerungsprojekte sind kundenspezifisch und befinden sich eigentlich noch im Experimentierstadium. Als preislicher Richtwert für die Ausbelichtung von 1 TB auf Film können 25'000 € angenommen werden.

Speichern im Netzwerk:

Diesen Begriff verwenden wir für die verteilte Speicherung in *Peer-to-Peer*-Netzwerken. Das Prinzip ist allgemein bekannt und bewährt (beispielsweise bei Tauschbörsen im Internet)⁸⁸, im archivischen Umfeld jedoch nicht in Gebrauch. Im Bibliotheksbereich wird es seit einigen Jahren erfolgreich für die Archivierung von E-Journals angewendet. Das Konzept beruht auf einem Verbund einzelner Computer (Server), die über ein Netzwerk miteinander verbunden sind.

Hauptcharakteristikum ist die sehr grosse Ausfallsicherheit, die mit dem Fehlen eines zentralen Verwaltungspunktes einhergeht.

Aus archivischer Sicht sind zwei Varianten dieses Konzepts denkbar, nämlich ein Verbund mehrerer Archive oder ein internes Netzwerk. Beide sind interessant und würden eine nähere Betrachtung rechtfertigen. Interessant sind vor allem die geografische Redundanz und die intellektuelle Kontrolle über die vom Träger getrennte Speicherlogik. Allerdings ist dieses Konzept für Staatsarchive im Moment nicht anwendungsreif: Es existiert keine Lösung, die einfach implementiert werden könnte.

***arcun* – Fazit**

Innovation

Betrachten wir die fünf Lösungen aus Sicht des Innovationspotentials und der Entwicklungsmöglichkeit.

- *Speichern im Netzwerk* ist wahrscheinlich die innovativste Lösung, es besteht aber noch keine im eigentlichen Sinne archivtaugliche Variante. Weitere Forschung scheint nötig, welche auch herkömmliche, in der Praxis bewährte P2P-Netzwerke einbeziehen sollte.
- *Datenauslagerung* besitzt ebenfalls ein hohes Innovationspotential und könnte in standardisierter Form langfristig die Archive finanziell entlasten, indem die Bewirtschaftungskosten für Teile der Bestände gesenkt werden können.

Best Practice

Speicherplatz mieten und *Blackbox*-Lösungen sind die aktuellen Angebote der Informatikindustrie.

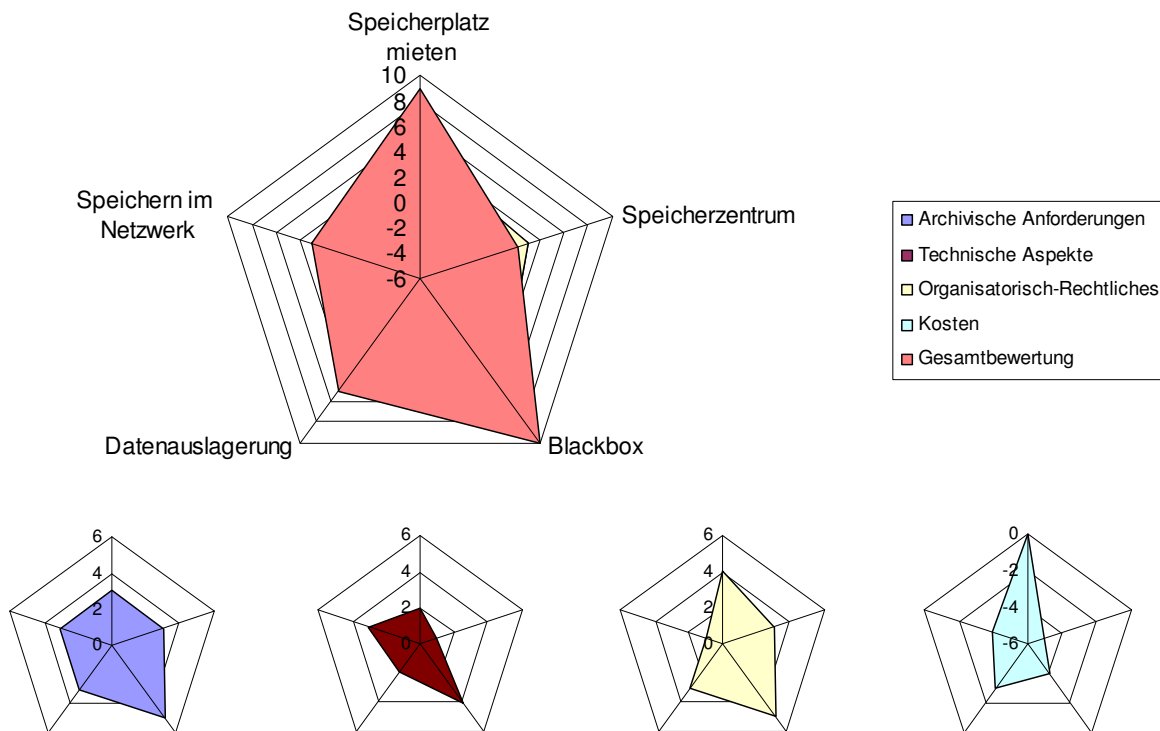
- *Blackbox*-Lösungen können heute als ausgereift betrachtet werden und lassen dem Archiv viel Gestaltungsfreiraum. Im Einsatz sind solche Systeme vornehmlich im KMU-Bereich, bei Firmen und Institutionen, die sich auf Grund ihrer Grösse kein eigenes Rechenzentrum leisten können, dennoch aber digital archivieren müssen.

⁸⁸ Siehe dazu auch <http://de.wikipedia.org/wiki/Peer-to-Peer>.

- *Speicherplatz mieten* verursacht wenig Aufwand im Archiv, schafft aber externe Abhängigkeiten. Diese Lösung ist erst seit kurzer Zeit wieder am Markt, nachdem die ersten Versuche in diese Richtung Ende der 1990er Jahre gescheitert waren.
- Das *Speicherzentrum* (Rechenzentrum) ist die klassische Lösung für die digitale Archivierung, ohne besonderes Entwicklungspotential, aber mit viel Erfahrungshintergrund. In der Regel ist es für kleinere Datenvolumen zu teuer.

Entscheidungsmatrix

Mit einer Entscheidungsmatrix vergleichen wir die fünf Speicherkonzepte aufgrund der oben besprochenen Kriterien.



Abbildungung 33: Entscheidungsmatrix mit folgenden Kriterien: Archivische Anforderungen, Technische Aspekte, Organisatorisch-Rechtliches, Kosten

Die Matrix zeigt nicht ganz unerwartet, dass die Speicherkonzepte *Blackbox* und *Speicherplatz mieten* die besten Resultate erhalten. Im Detail sehen wir, dass *Speicherplatz mieten* vor allem bei den Kosten vorne liegt. Im Projekt *arcun* wird dieses Konzept weiterverfolgt.

arcun – Speicherlösung

Auftrag

Aufgrund dieser Abwägungen hat die KOST-Aufsichtskommission Anfang 2008 beschlossen, das gemeinsame Beschaffen von Speicherplatz weiter zu verfolgen und die Geschäftsstelle beauftragt, bei kommerziellen oder öffentlich-rechtlichen Anbietern entsprechende Offerten einzuholen. Dazu sollte sie einen *Service Level* für den Dienst definieren (siehe unten) und eine Speicherschnittstelle spezifizieren. Als Zielgröße für die Speicherkosten werden 4'500

Fr./TB/Jahr festgelegt (basierend auf der SDSC-Studie⁸⁹, die von 1'500 Fr./TB/Jahr für eine Kopie ausgeht). Die KOST übernimmt Auditing und Inkasso für die gewählte Lösung.

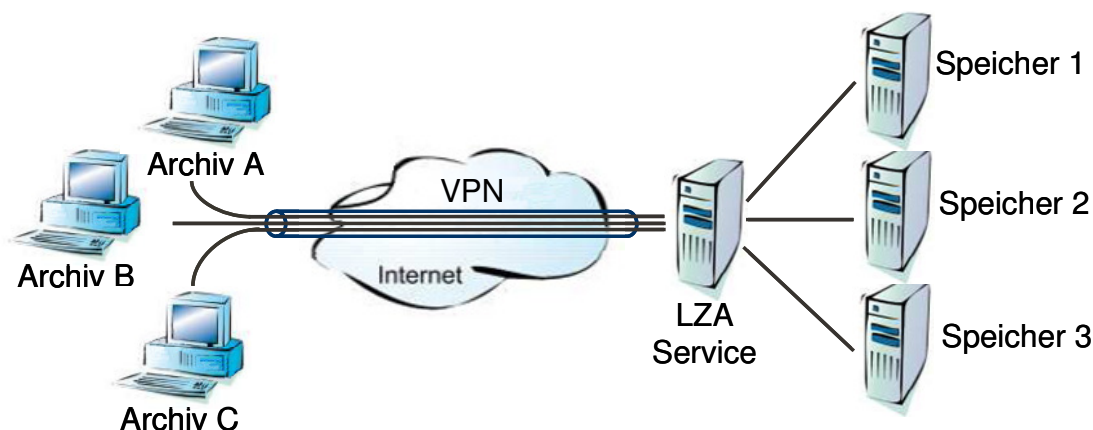


Abbildung 34: *arcun*: Mehrere Archive benutzen das gleiche digitale Speicherzentrum

Realisierung

Eine erste Offertrunde im August 2008 hat ergeben, dass sich das Projekt mit diesen technischen, organisatorischen und preislichen Vorstellungen realisieren lässt. Es haben sich aber zu diesem Zeitpunkt nicht genügend interessierte Archive gefunden. Nach erneuter grundsätzlicher Analyse der Problemstellung und der Bedürfnisse beteiligen sich im Frühjahr 2009 nun fünf Archive mit einem anfänglichen Speicherplatzbedarf von insgesamt 5 TB an der Realisierung von *arcun*. Die wichtigsten Eckpunkte der Ausschreibung sind (Auszug aus dem SLA):

- Die Daten sollen über eine verschlüsselte Verbindung und eine Standard-Speicherschnittstelle vom jeweiligen Datenbesitzer im dLZA gespeichert und wieder gelesen werden können. Einmal gespeicherte Daten werden nicht mehr verändert oder gelöscht. Der Anbieter wählt eine oder mehrere Methoden zur sicheren Datenübertragung Archiv – Speicherinfrastruktur und offeriert den Aufwand für die gewählte Lösung für ein einzelnes Archiv.
- Die Daten sollen im LZA-Speicher dreifach und an zwei geographisch voneinander getrennten Standorten gespeichert sein.
- Ein periodischer Integritätscheck muss die Integrität der Daten bzw. die Lesbarkeit der verwendeten Medien gewährleisten und soll mindestens alle drei Monate den Gesamtdatenbestand einmal überprüfen.
- Der Zugriff (speichern/lesen) soll sieben Tage pro Woche rund um die Uhr möglich sein. Geplante Ausfallzeiten (fünf Tage vor dem Beginn der Ausfallzeit bekannt gegeben) sind während den Wochenenden und ausserhalb der Bürozeiten (8-18 Uhr) sowie max. während fünf Arbeitstagen/Jahr vorzusehen.

⁸⁹ Richard L. Moore et al., *Disk and Tape Storage Cost Models*, IS&T Archiving Conference 2007: <http://www.imaging.org/store/phypub.cfm?seriesid=28&pubid=765>.

- Die *upstreaming*-Rate (fortlaufendes Datenspeichern) muss mindestens 2 MB/s insgesamt betragen. Zum Abrufen einer Datei reicht eine Datenrate von 0.6 MB/s. Die mittlere Suchzeit bei wahlfreiem Zugriff darf eine Minute nicht übersteigen.
- Es muss möglich sein, alle Daten eines Mandanten zusammen mit dem Schlüssel des jeweiligen Objektes sequentiell innerhalb angemessener Zeit (14 Tage) auszulesen, z.B. im Falle des Vertragsendes oder der Vertragsauflösung.
- Zu den zu verwendenden Speichermedien werden keine Vorgaben gemacht.
- Die Lösung soll im Betrieb transparent, d.h. für den Benutzer nicht sichtbar sein. Die gewählte Hardware- und Softwarelösung soll in der Offerte aber beschrieben und bei Vertragsvereinbarung offengelegt werden, soweit das für ein *Auditing* im Auftrag der KOST notwendig ist.

arcun+ – Der Archivierungsprozess als Dienstleistung

In Erweiterung des Projekts *arcun* wurde im Projekt *arcun+* angedacht, den gesamten Archivierungsprozess als Dienstleistungsangebot zu realisieren. Es würde hier also gewissermaßen um eine Archivsteckdose gehen, über welche die Archive Ablieferungen in ein zentrales digitales Langzeitarchiv einstellen können und über die auf die einzelnen Unterlagen in diesen Ablieferungen jederzeit wieder zugegriffen werden kann. Der Dienst umfasst dabei die *Bitstream Preservation*, die Sicherstellung der Lesbarkeit der Datenobjekte sowie die Kontrolle über die notwendige Anreicherung mit Metadaten. Es ist hier also eine vollumfängliche digitale Archivierungslösung im Sinne von OAIS geplant, die von einem Dienstanbieter nach Vorgabe der Spezifikation (des SLA) angeboten wird.

arcun+ – Abgrenzung zu arcun

Aus Sicht des OAIS-Modells umfasst die *arcun*-Lösung (*Bitstream Preservation*) den grün schraffierten Bereich. *Ingest*, *Access*, *Preservation Planning* und *Data Management* bleiben die Aufgabe des Archivs, bei dem die Datenhoheit liegt.

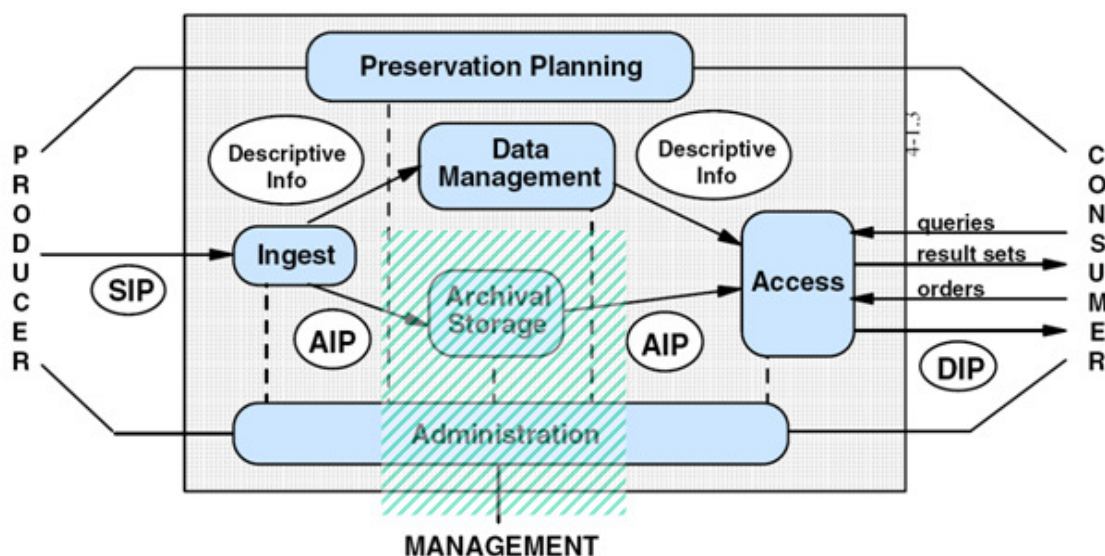


Abbildung 35: *arcun* im OAIS-Modell: Archival Storage, Administration

Übertragen auf Soft- und Hardware-Layer bedeutet dies, dass nur der unterste Software-Layer und die Hardware Teil der Lösung sind:

Layer	Funktion und Beschreibung
Archival Management Software	Findmittel bzw. Katalog erlaubt den (strukturierten) Zugriff auf die archivierten Daten
Digital Archiving Software	Entgegennehmen, Speichern, Verwalten und Ausliefern von digitalen Objekten
Storage Management Software	Verwalten des Speicherplatzes auf dem Speichersystem
Storage Hardware	Physisches Speichermedium

In der *arcun+*-Lösung (*digitale Archivierung*) hingegen werden folgende zusätzliche Dienstleistungen erbracht:

- An der Ingest-Schnittstelle zum digitalen Archiv wird kontrolliert, dass die Datenspeicherung in Form eines definierten SIPs, angereichert mit den archivisch notwendigen Metadaten, stattfindet.
- Die Unterlagen werden im digitalen Archiv sodann nicht nur gespeichert, sondern es wird bei Bedarf auch durch Formatmigrationen ihre weitere Lesbarkeit gewährleistet.
- Dadurch, dass die Datenobjekte im digitalen Archiv als AIPs mit Metadaten gespeichert sind, kann nach Bedarf vom Dienstanbieter auch ein Findmittel aufgebaut und gepflegt werden.

Im OAIS-Modell sieht diese *arcun+*-Lösung nun folgendermassen aus (wiederum grün schraffiert):

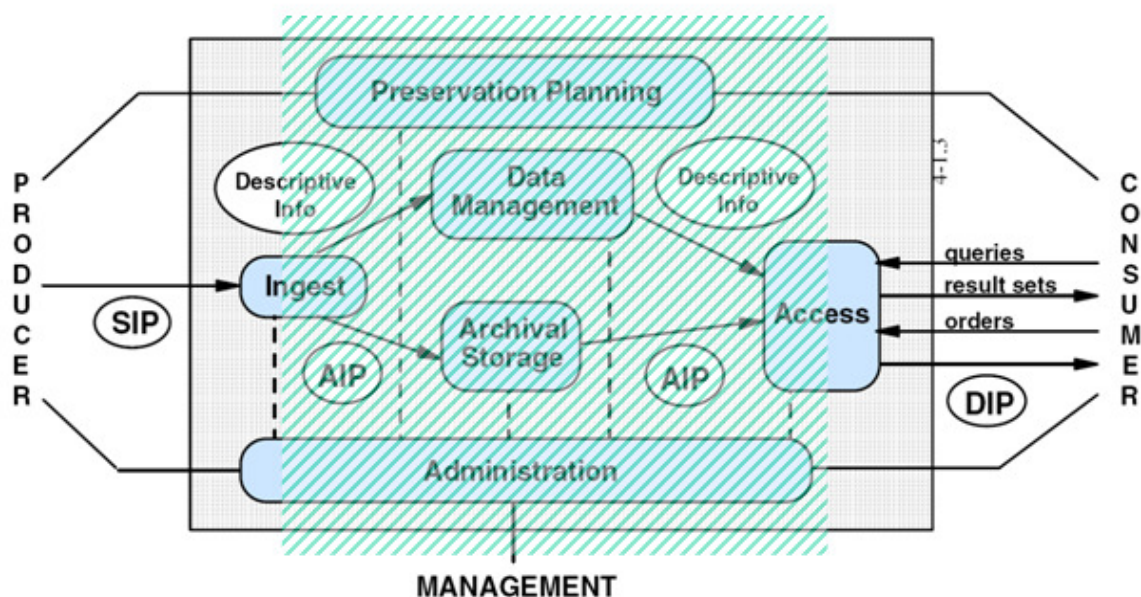


Abbildung 36: *arcun+* im OAIS-Modell: *Preservation Planning*, *Data Management*, *Archival Storage*, *Administration*

Hier ist neben der Hardware nicht nur die *Storage Management Software*, sondern auch die *Digital Archiving Software* Teil der ausgelagerten Dienstleistung:

Layer	Funktion und Beschreibung
Archival Management Software	Findmittel bzw. Katalog erlaubt den (strukturierten) Zugriff auf die archivierten Daten
Digital Archiving Software	Entgegennehmen, Speichern, Verwalten und Ausliefern von digitalen Objekten
Storage Management Software	Verwalten des Speicherplatzes auf dem Speichersystem
Storage Hardware	Physisches Speichermedium

Weiterhin in der Obhut der Archive bleiben:

- Der *Pre-Ingest*-Prozess: Entgegennahme einer digitalen Ablieferung zusammen mit den notwendigen Metadaten und Aufbereitung zu einem SIP.
- Die *Vermittlung*: Suchen, Finden und Anschauen von Archivobjekten mit einem geeigneten Findmittel.
- Die *Access-Control*: Kontrolle der Zugriffsberechtigungen.

Diese Aufgaben werden mit Hilfe der *Archival Management Software* erbracht.

arcun+ – Ingest und Access im Archiv

Zum reibungslosen und automatisierbaren Funktionieren der Datenübergabe Archiv ⇔ *arcun+* bei Ingest und Access müssen die entsprechenden Schnittstellen (Prozesse und Datenstrukturen) zwischen Archiv und Anbieter der Archivierungsdienstleistung *arcun+* genau spezifiziert werden.

***Pre-Ingest*⁹⁰**

Mit einem *Pre-Ingest*-Tool kann das Archiv die übernommenen digitalen Daten in ein SIP umwandeln. Das heisst, die Ablieferung wird mit diesem Tool im Archiv in eine vorgegebene Struktur gebracht, mit bestimmten, definierten Metadaten versehen und anschliessend automatisch an das digitale Archiv *arcun+* übermittelt. Das SIP enthält alle für die Archivierung notwendigen Informationen. Sinnvollerweise werden in diesem Tool die Metadaten für den Import in das lokale Findmittel ebenfalls bereitgestellt.

Access Control

Im Vermittlungsfall, das heisst beim Zugriff auf die archivierten Unterlagen, übernimmt das Archiv die Zugriffskontrolle bezüglich Berechtigung und Einhaltung der Schutzfrist. Der Dienstleister überprüft nur die Zugriffsberechtigung des entsprechenden Archivs.

⁹⁰ Die Bezeichnung ist dem kopal-Projekt entlehnt: <http://kopal.langzeitarchivierung.de>.

***arcun+* – Archivierung beim Dienstanbieter**

Eine erste Marktbeobachtung und die weiteren Diskussionen unter den beteiligten Archiven haben schnell gezeigt, dass eine Lösung in der Art von *arcun+* für eine vollumfängliche Archivierung im Augenblick noch nicht zu realisieren ist. Es gibt keine Anbieter mit Erfahrung, es fehlen die entsprechenden Standards, um die Dienstleistung entsprechend zu spezifizieren, und es sind keine fertigen Softwareprodukte auf dem Markt.

***arcun+* – Fazit**

Aufgrund dieser Analyse hat die KOST-Aufsichtskommission im Frühjahr 2009 beschlossen, den Projektauftrag zu ändern. Statt den gesamten Archivierungsprozess als Dienstleistung zu planen und dafür Anbieter zu suchen, soll neu daraus ein Standardisierungsprojekt werden. Der neue Auftrag lautet, die Prozesse und Schnittstellen, wie sie in *arcun+* bereits bezeichnet sind, soweit zu beschreiben und zu spezifizieren, dass Anbieter und Dienstleister, aber allenfalls gewisse Archive auch selbst Lösungen für den digitalen Archivierungsprozess entwickeln können. Die Standards sollen dabei so gesetzt werden, dass minimale archivische Anforderungen an die digitale Langzeitarchivierung erfüllt werden und dass ein Austausch bzw. eine Interpretation von AIPs zwischen den Produkten verschiedener Anbieter von Lösungen oder verschiedenen Archiven möglich ist.

Im Zentrum steht dabei die Spezifikation eines AIPs, d.h. einer gebündelten abstrakten Archivseinheit, die sowohl die originalen Primärdaten als auch zusätzliche Informationen (Metadaten) zu ihrer Darstellung und Einordnung in den ursprünglichen Entstehungszusammenhang umfasst. In zweiter Ordnung bedeutet das die Spezifizierung der Ingest-Schnittstelle (Prozesse und Daten), über welche die digitalen Unterlagen vom Ablieferer ins Digitale Archiv gelangen. Nachgeordnet im Projekt sind dann die weiteren Prozesse des OAIS-Modells.

Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (GEVER): Projekt AUGev und Pilotprojekt LARIS I

Martin Lüthi

Im Herbst 2006 machte die KOST⁹¹ eine Umfrage zu einem Modellprojekt, welches als Fortsetzung und Erweiterung zum Projekt KOSTPROBE⁹² gedacht war. Anhand exemplarischer Unterlagen aus der Verwaltung sollten geeignete Archivierungsformen untersucht werden. Es sollten insbesondere die technischen und organisatorischen Voraussetzungen untersucht werden, die gegeben sein müssen, damit elektronisches Archivgut in für den restlichen Lebenszyklus geeigneter Form aufbewahrt werden kann. Der Bau eines Prototyps sollte einer Überprüfung der gewonnenen theoretischen Erkenntnisse dienen und die Basis für ein produktives System bilden.

Da das Staatsarchiv St.Gallen ohnehin plante, im Jahr 2007 ein konkretes Projekt im Rahmen der Langzeitarchivierung durchzuführen, beteiligte es sich an der Ausschreibung. Nach Sichtung der eingegangenen Vorschläge entschied sich die KOST, die Projekte der Staatsarchive Basel-Stadt, Zug und St.Gallen, welche sich allesamt die Übernahme von Daten aus einem GEVER-System zum Ziel setzten, als Projekt Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (AUGev) zu verwirklichen. Gegenstand des st.gallischen Projekts bildete das sogenannte "Ratsinformationssystem" (RIS), das im Bereich von Kantonsrat (Parlament) und Regierung seit 2006 in Betrieb stehende elektronische Geschäftsverwaltungssystem, das dem hier zur Diskussion stehenden St.Galler Projekt denn auch den Namen gab (LARIS = Langzeitarchivierung Ratsinformationssystem).

Das unter dem Dach der KOST abgewickelte Gesamtprojekt (= Projekt AUGev) gliederte sich in folgende zwei Phasen:

- eine Konzept-Phase (Phase I), die der gemeinsamen Erarbeitung von Prozessmodellierung und Anforderungsanalyse diene;
- eine Pilotprojekt-Phase (Phase II), bei der die beteiligten Archive in jeweils individueller Arbeit versuchten, die in der ersten Phase erarbeiteten Konzepte in die Praxis umzusetzen.

Die Erarbeitung des Konzeptes (Phase I) erfolgte in zehn Workshops zwischen Februar und September 2007 sowie bei diversen Arbeiten zwischen den Workshops. Am Ende der ersten Phase wurde ein Schlussbericht erstellt. Dieser Bericht diene als Grundlage für die Phase II, in der von den beteiligten Archiven und der KOST parallel drei verschiedene Vorgehensweisen gewählt und schliesslich drei unterschiedliche Prototypen entwickelt wurden:

⁹¹ Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen.

⁹² Fragen der Langzeitarchivierung und die damit verbundenen Herausforderungen.

- Teilprojekt der KOST-Geschäftsstelle: Referenzimplementierung – Schnittstellen Prototyp in XML⁹³/XSLT⁹⁴
- Teilprojekt des Staatsarchivs Zug: Pilot-Implementierung – funktionaler Prototyp in J2EE⁹⁵
- Teilprojekt des Staatsarchivs St.Gallen: DIAS Pilot – auf der Basis einer kommerziellen Lösung

Das gemeinsame Ziel der Phase II des Projektes AUGev war es, eine Liste von Empfehlungen zu erstellen. Diese Empfehlungen sollten den Aufbau eines AIP, Bewertungskriterien für eine SIP-Schnittstelle, die Granularität von AIP und SIP, das Mapping von Metadaten, das Containerformat und einen Vorschlag für AIP und SIP betreffen. Was die Durchführung der Pilotprojekt-Phase (Phase II) im Staatsarchiv St.Gallen anbelangt, so wurde bereits während der Konzeptphase eine interne Evaluation durchgeführt, um einen geeigneten externen Partner zu finden, welcher Erfahrung im Bereich Archivierung mitbringt. Das von diesem verwendete Produkt sollte möglichst Standardsoftware nutzen und Standards einsetzen. Die Wahl fiel schliesslich auf das Produkt DIAS (Digital Information Archiving System) der Firma IBM.⁹⁶ Gemeinsam mit den Partnern der Firma IBM wurden hierauf folgende Ziele für den Piloten festgelegt:

- Der Pilot muss in einer Form Daten anzeigen können, um das "Look and Feel" der Lösung zu vermitteln.
- Die Datenübernahme aus RIS (Ratsinformationssystem) muss durchgeführt werden können.
- Die Einbettung von scopeArchiv (Archivsoftware des Staatsarchivs) soll gezeigt werden, bzw. der Weg soll aufgezeigt werden.
- Der Zugriffsprozess des Archivars von seinem Arbeitsplatz aus muss aufgezeigt werden können.

Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (AUGev)⁹⁷

Das Projekt AUGev diente einerseits als Konzeptphase andererseits als Grundlage für die Erstellung der Prototypen. Das Projekt AUGev umfasste die Modellierung der Archivierungsprozesse und Grundlagen der Datenmodellierung, ergänzt um Überlegungen zum digitalen Magazin, ein Glossar sowie eine Bibliografie.

⁹³ XML: Die Extensible Markup Language (engl. für "erweiterbare Auszeichnungssprache"), abgekürzt XML, ist eine Auszeichnungssprache zur Darstellung hierarchisch strukturierter Daten in Form von Textdateien.

⁹⁴ XSLT: XSL-Transformation, kurz XSLT, ist eine Programmiersprache zur Transformation von XML-Dokumenten.

⁹⁵ Java Platform, Enterprise Edition, abgekürzt Java EE oder früher J2EE.

⁹⁶ <http://www-05.ibm.com/nl/dias/> [Stand: 29.07.2009].

⁹⁷ Teile des Berichtes wurden aus den Schlussberichten der KOST übernommen und gekürzt. Die vollständigen Unterlagen sind unter www.kost-ceco.ch zu finden oder über das Staatsarchiv St.Gallen.

Abgrenzung der Aufgabenstellung mittels OAIS

Für die Modellierung diente das im Bereich der digitalen Archivierung weit verbreitete OAIS-Referenzmodell als Leitfaden. Das OAIS beschreibt einerseits die Prozesse, die ein digitales Archiv umfasst, andererseits die Schnittstellen zur Übergabe von Informationen zwischen Prozessen. Das OAIS beschreibt ein digitales Archiv als Interaktion von Prozessen und Spezifikation von Schnittstellen, die dazu dienen, digitale Akten zu speichern und wieder verfügbar zu machen. Im Projekt wurde die Archivierung von Unterlagen aus GEVER-Systemen behandelt. Der Projektfokus war eingeschränkt auf ein digitales Archiv inklusive Bestandserhaltung mit Schnittstellen zu einem GEVER-System, zum archivinternen Findmittel und zur Benutzung. Daraus ergibt sich, dass der vorarchivische Bereich, die Ausprägung des Findmittels und die Benutzung nicht Teil des Projekts waren. Diese Bereiche wurden jedoch mitgedacht und flossen in die Spezifikation der Schnittstellen ein. Auf das OAIS-Modell abgebildet bedeutete die Projektausrichtung eine Konzentration auf die vier Kernprozesse Ingest (Übernahme), Archival Storage (Speicherung), Data Management (Verwaltung), Access (Zugang). Daneben mussten die vier Schnittstellen SIP (Schnittstelle zum GEVER-SYSTEM), AIP (Datenkapsel), DIP (Zugangsformat) und Descriptive Info (Schnittstelle zum Findmittel) spezifiziert werden.

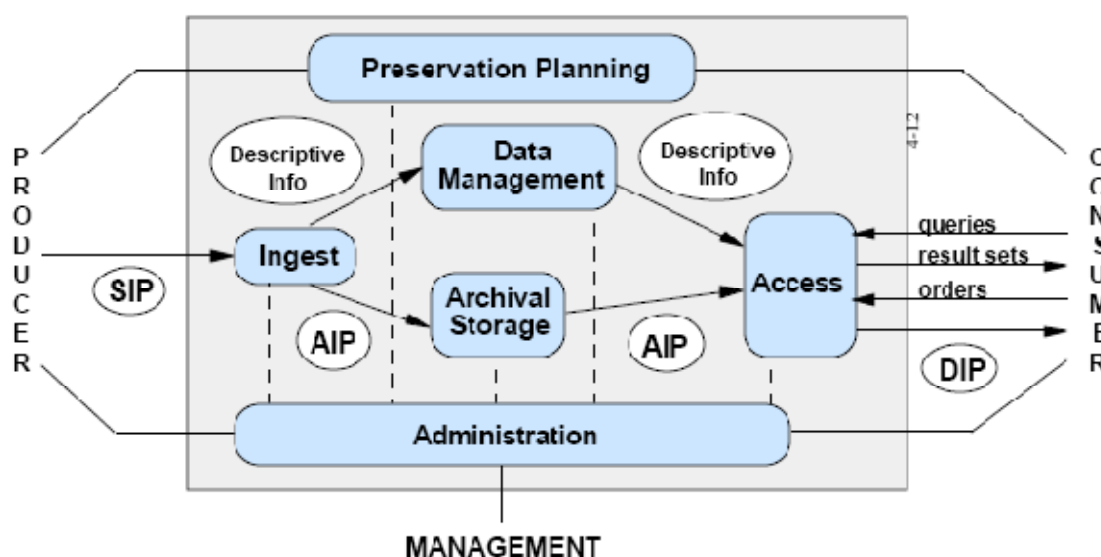


Abbildung 37: OAIS-Referenzmodell⁹⁸

Prozessmodellierung

Zur Strukturierung und Darstellung wurde die Modellierungssprache UML (Unified Modelling Language) verwendet. Diese erlaubt ein iteratives Vorgehen zur Prozessmodellierung: In einem ersten Durchgang wurde die Interaktion eines Anwenders mit dem modellierten System mittels Anwendungsfällen beschrieben. Diese wurden danach in weitere Anwendungsfälle unterteilt, in eine Abfolge von Aktivitäten aufgegliedert und detailliert beschrieben. So entstand eine abstrakte Abbildung des gesamten Archivierungsprozesses auf einzelne Schritte und ihre Resultate. Ausgehend vom Lebenszyklus der Unterlagen identifizierte und untersuchte das

⁹⁸ Übersicht Referenz-Modell: Consultative Committee on Space Data Systems, Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, BLUE BOOK, January 2002.

Projekt die Archivprozesse "Unterlagen übernehmen", "Unterlagen magazinieren" und "Unterlagen erhalten". Die Prozessmodellierung hält zudem fest, welcher Input und welcher Output für jeden Schritt erforderlich ist. Diese Objekte wurden im Rahmen der Datenmodellierung genauer spezifiziert, ebenfalls mit Hilfe der UML. Es handelt sich dabei einerseits um die Schnittstellenobjekte, die vom OAIS gefordert werden, andererseits um weitere Objekte, die aus der Prozessmodellierung hervorgehen. Im Verlauf der Arbeit hat sich gezeigt, dass die Spezifikation der Objekte sinnvollerweise direkt an Prototypen erprobt und verbessert wird. Dies ist ein Hauptbestandteil der Pilotprojekt-Phase (Phase II).

Lebenszyklus digitaler Unterlagen:

Das Use-Case-Diagramm Lebenszyklus bildet den Lifecycle digitaler Unterlagen aus Sicht des Archivs ab. Die Definition des Systems ist hier etwas künstlich, besteht der Lebenszyklus doch aus den Provenienz-, Archivierungs- und Benutzungsprozessen.

Das Projekt AUGev fokussiert auf die Anwendungsfälle der Archivierung im engeren Sinne (unten grau hinterlegt): "Unterlagen übernehmen", "Unterlagen magazinieren" und "Unterlagen erhalten".

"Unterlagen bewerten" beinhaltet Bewertung und Beratung in der vorarchivischen Phase, wenn sich die Unterlagen noch beim Amt befinden. Bei "Unterlagen erschliessen" handelt es sich um archivische Nacherschließung, insbesondere mit Angaben auf höheren Erschließungsniveaus (Archiv, Bestände, Ablieferungen). Im Gegenzug dazu umfasst "Unterlagen ordnen und verzeichnen" im Use Case "Unterlagen übernehmen" die automatisierte Einordnung in die Archivtektonik und die Übernahme der Metadaten.

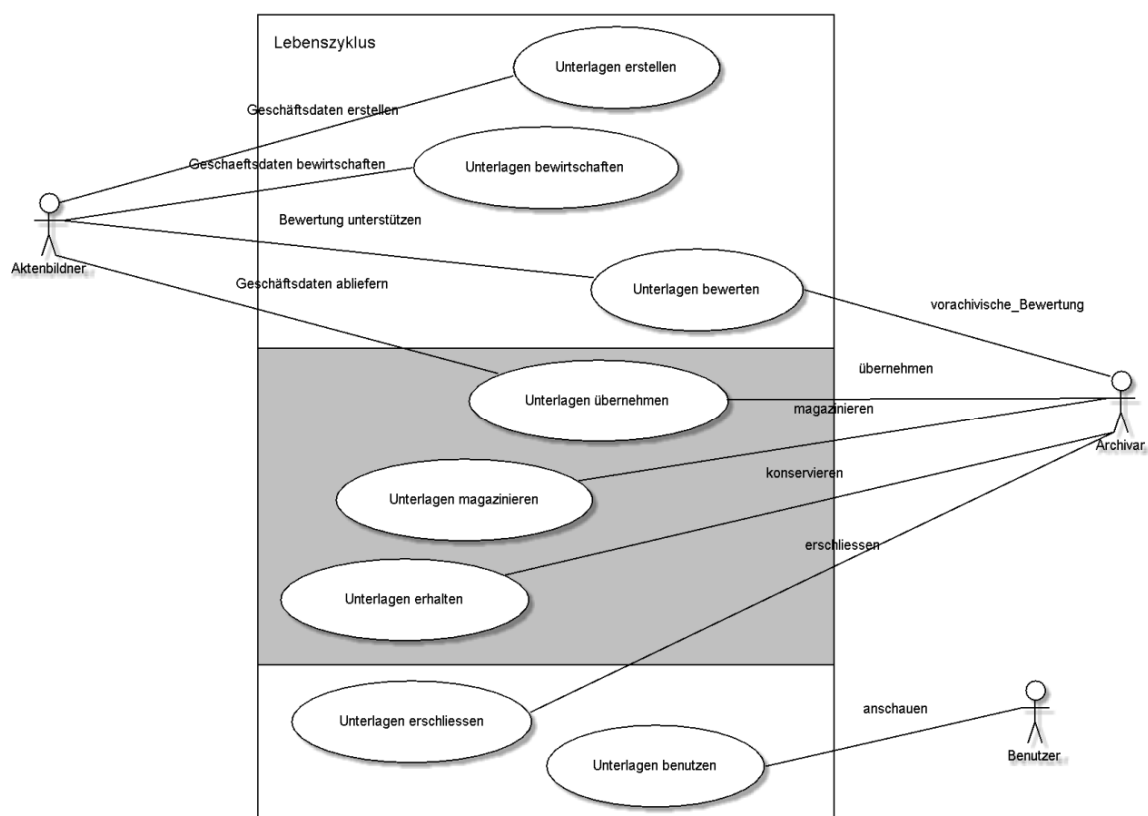


Abbildung 38: Use Case-Lebenszyklus

Aktivitätsdiagramm:

Das Aktivitätsdiagramm detailliert die einzelnen Schritte in den Prozessen und die dazu benötigten Objekte. Es funktioniert analog einem Ablaufdiagramm. Es werden die einzelnen Schritte von einem Startpunkt bis zum Ziel aufgezeigt und definiert.

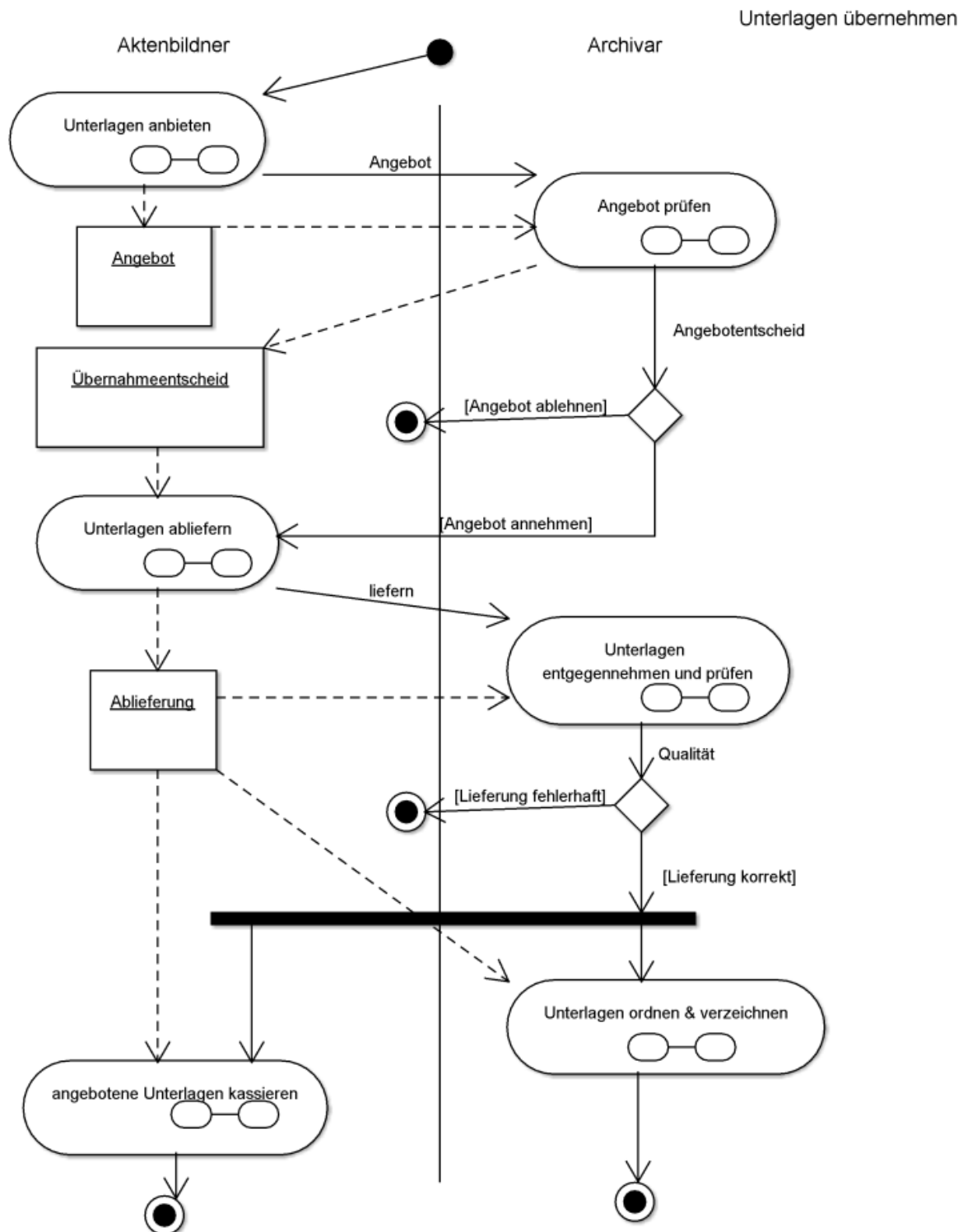


Abbildung 39: Aktivitätsdiagramm "Unterlagen übernehmen"

Die detaillierten Beschreibungen der einzelnen Use Cases können dem Schlussbericht der KOST entnommen werden⁹⁹. Es wird aus Platzgründen auf die Publikation in diesem Rahmen verzichtet.

Datenmodellierung und Schnittstellenspezifikation

Die zweite Projektphase diente der Verfeinerung der Datenmodellierung mit Spezifizierung der Systemschnittstellen. Auf dieser Grundlage erarbeiteten anschliessend die beteiligten Archive drei unterschiedliche Prototypen. Bei der Erarbeitung und Diskussion der Prozess- und Datenmodellierung, aber auch in Auseinandersetzung mit den Prototypprojekten hat die Arbeitsgruppe verschiedene Erkenntnisse gewonnen, die als Quintessenz aus ihrer Arbeit gelten können. Es wurde eng an der Nomenklatur von OAIS gehalten und beschreibt die Datenstrukturen SIP, AIP und DIP, wobei je nach Sichtweise darunter Schnittstellen oder Datenpakete verstanden werden können.

Archivtektonik:

Ein Objekt im Archiv ist durch seine Signatur identifiziert und zitierbar. Bei den meisten Archiven widerspiegelt die Signatur in sprechender Syntax die Ablagestruktur im Archiv, mithin die Archivtektonik. Wir können bei der Archivierung aus GEVER-Systemen folgendermassen für die Signatur verallgemeinern:

<i>Ebene</i>	<i>Sicht Amt</i>	<i>Sicht Archiv</i>
Archivtektonik	Provenienz	Archiv
		[Bestand]
		[Ablieferung]
Aktenplan Geschäftsverwaltung	[OS-Position]	[Serie]
	Dossier-Nr.	Dossier-Nr.
	Dokument-Nr.	Dokument-Nr.

Abbildung 40: Signaturstruktur¹⁰⁰

Modellierung AIP:

Wir bilden das AIP als Komposition von Verzeichniseinheiten (Komposition: Ganze-Teile-Hierarchie, bei der die Teile vom Ganzen existenzabhängig sind). Die Menge aller AIPs in einem Archiv bildet so die Baumstruktur der Archivtektonik ab.

⁹⁹ URL: http://kost-ceco.ch/cms/index.php?augev_public_de [Stand: 17.8.2009].

¹⁰⁰ [...] sind fakultative Verzeichnungsebenen bei der Signaturbildung. Provenienz/Bestand und Ordnungssystemposition/Serie können rekursiv definiert werden.

Bewertungskriterien für ein AIP:

Das AIP nach OAIS hat im Falle der Archivierung von Unterlagen aus GEVER-Systemen die Aufgabe, den Inhalt eines Geschäftes – das sind die Dokumente und die zusätzlichen Informationen zu ihrem Entstehungszusammenhang – zusammenzufassen und mit den für die Langzeitarchivierung notwendigen technischen Informationen zu versehen. Dieses Ziel wird erreicht, indem inhaltliche und technische Metadaten strukturiert nach einem Metadatenschema zusammen mit den Primärdaten in einem Container archiviert werden. Mit Vorteil wird bei den Schemata auf Standards zurückgegriffen.

Für die Beurteilung unterschiedlicher AIP-Lösungen legen wir einen einfachen, aber hinreichenden Bewertungskatalog vor. Dieser ist nach den Regeln der Softwarespezifikation unterteilt in vier funktionale und drei nicht-funktionale Kriterien¹⁰¹.

Funktionale Kriterien für die Bewertung eines AIP:

- Aus dem AIP können die notwendigen Metadaten für ein ISAD(G)-konformes Findmittel extrahiert werden.
- Notwendige ISAD(G)-Metadaten sind schemakontrolliert und nicht als Freitext oder Key-Value-Paare abgelegt.
- Aus dem AIP kann wieder eine universelle GEVER-Sicht (DIP) generiert werden.
- Formatmigrationen bei Übernahme oder Formatobsolenzen können im AIP schemakonform und mit vernünftigem Aufwand abgebildet werden.

Nicht-funktionale Kriterien für die Bewertung eines AIP:

- Ordnungssystem, Metadaten und digitale Objekte sind eng miteinander verbunden.
- Unnötige Redundanz ist vermieden, d.h. die Vererbung von Informationen erfolgt entlang der Ordnungsstruktur oder durch Referenzierung.
- Es werden möglichst wenige Standards verwendet.

Bewertungskriterien für ein SIP:

Das SIP nach OAIS ist einerseits das Gefäß (Container), in dem die Daten einer Ablieferung zur Archivierung abgeliefert werden, andererseits definiert es eine Schnittstelle zwischen dem GEVER-System und dem Ingest-Prozess auf Archivseite (siehe dazu die Prozessmodellierung "Unterlagen übernehmen" in AUGev Phase I).

Für die Bewertung von SIP-Schnittstellen legen wir einen einfachen Bewertungskatalog vor. Wie der Kriterienkatalog für die AIP-Lösungen ist er in funktionale und nicht-funktionale Kriterien unterteilt.

Funktionale Kriterien für die Bewertung eines SIP:

- Die Schnittstelle kann sämtliche GEVER-Metadaten aufnehmen.

¹⁰¹ Siehe zu "functional and non-functional requirements" ISO/IEC 9126-1:2001 und http://en.wikipedia.org/wiki/Non-functional_requirements.

- Die notwendigen ISAD(G)-Felder werden schemakontrolliert abgebildet.

Nicht-funktionale Kriterien für die Bewertung eines SIP:

- Die Schnittstelle implementiert anerkannte Standards.
- Die Schnittstelle verwendet ein Containerformat.
- Die Schnittstelle enthält als zusätzliche Ablieferungsmetadaten eine innerhalb der abliefernden Stelle eindeutige ID, das Datum und die Kontaktperson.

Granularität von SIP und AIP:

Die Frage, welche Granularität für die Informationspakete SIP und AIP ideal ist, ist sowohl eine archivische als auch eine technische Frage. Zu beachten ist deshalb, welche Sicht auf die Objekte der Archivierung bei der Übernahme und bei der Speicherung Sinn macht und realisierbar ist. Die AUGev-Projektgruppe empfiehlt als ideale Granularität für das SIP die Ablieferung. Unter der Ablieferung verstehen wir die Menge der gleichzeitig übernommenen Dossiers und Dokumente, zusammen mit dem Ordnungssystem der Provenienz im Zeitpunkt der Übernahme. Die Ablieferung ist die Informationseinheit, die im Übernahmeprozess von Unterlagen ins Archiv im Mittelpunkt steht. Mit dem Fokus auf die Ablieferung als SIP können Redundanzen bei der Übernahme des Ordnungssystems vermieden werden. Für das AIP sieht die Projektgruppe als ideale Granularität das Dossier vor. Ein Dossier ist die Menge der Dokumente, Vorgänge und Beteiligten zu einem Geschäft. (Dabei bilden Vorgänge und Beteiligte die Dossierprozessdaten, sind also Metadaten zum Dossier.) Traditionell ist das Dossier die unterste Stufe bei der archivischen Erschliessung. Deshalb empfiehlt sich, das SIP für die archivische Speicherung auf AIPs der Granularität Dossier herunterzubrechen. Ein solches AIP umfasst folglich die Primär- und Metadaten einer Verzeichnungseinheit, die zugehörigen Ablageinformationen (eine Menge von Verzeichnungseinheiten in einer "besteht aus"-Beziehung mit einem Dossier als Blatt) und die technischen Metadaten, und zwar sowohl die mitgelieferten als auch die beim Ingest-Prozess gewonnenen. Zur Vereinfachung soll durchwegs das Dossier als AIP-Granularität gelten, ob es nun Dokumente, Subdossiers oder beides beinhaltet. Nicht präjudiziert ist damit die Frage nach der Granularität der Benutzung. Aus den vorliegenden AIP-Varianten können DIPs in der jeweils gewünschten Granularität (Dokument, Dossier, Ablieferung) generiert werden. Eine Rolle spielt dabei auch die Erschliessungstiefe.

Empfehlung für ein Containerformat:

Die einzelnen Dateien, welche zusammen ein AIP bilden, werden sinnvollerweise in einem Container zusammengefasst, damit sie in Form einer einzigen Datei von der Speicherinfrastruktur verwaltet werden können. Damit wird auch der Idee des Preservation Data Store (PDS)¹⁰², der Speicherung von Primärdaten zusammen mit der notwendigen Preservation

¹⁰² *Preservation Data Store* ist ein Begriff aus dem PLANETS-Projekt (Preservation and Longterm Access through Networked Services, <http://www.planets-project.eu/>); siehe im Detail: http://www.haifa.ibm.com/projects/storage/datastores/papers/Preservation_DataStores_MSST07_camera.pdf.

Description Information an einem physischen Speicherort, Rechnung getragen. Als Containerformat bieten sich TAR, ZIP und JAR sowie die XML-Formate METS und XFDU an (Letztere als Container mit base64-Primärdaten, nicht als Metadatenstandards).

Da es sich beim Containerformat letztlich um ein Dateiformat handelt, können die Kriterien aus dem Katalog archivischer Dateiformate der KOST zur Anwendung kommen. METS und XFDU, obwohl interessant als XML-basierte Lösungen, scheiden wegen der geringen Verbreitung als zu exotisch aus. TAR, ZIP und JAR weisen technisch viele Gemeinsamkeiten auf. TAR ist das älteste unkomprimierte Dateiformat, das Dateien und ihre Verzeichnisstruktur in einer "Archiv"-Datei zusammenfasst. ZIP ist eigentlich ein komprimiertes TAR und besitzt die grösste Verbreitung. Die meisten Tools können mit allen Formaten gleichermassen umgehen.

Wir empfehlen die Verwendung von JAR10 (Java Archive Format/zu Deutsch "Gefäss"). Es besteht zudem die Möglichkeit, alle Dateien zur Sicherung der Integrität mit einem Hash-Wert zu versehen und die gesamte JAR-Datei zu signieren.

Empfehlung für ein AIP:

Anhand der Bewertungskriterien für ein AIP gelangen wir bei der Betrachtung verschiedener AIP-Lösungen zu folgenden Erkenntnissen:

- METS alleine reicht zur Speicherung von inhaltlichen und technischen Metadaten nicht aus.
- METS in Kombination mit EAD führt zu Überschneidungen.
- Dublin Core ist sehr limitiert in der Beschreibung von archivischen Metadaten und Strukturen.
- Für administrativ-technische Metadaten ist trotz der Möglichkeiten von METS bzw. EAD ein eigenes Schema (PREMIS oder LMER) unerlässlich.

EAD kann ISAD(G) abbilden sowie zusätzliche (GEVER-)Metadaten aufnehmen. Einzelne EAD-Files lassen sich als Findmittel verwenden oder mit anderen zu einem umfassenden Archivinformationssystem konkatenieren. Wir empfehlen deshalb eine Lösung, bei der die inhaltlichen Metadaten zu einem Geschäft im EAD-MetadatenSchema gespeichert werden. Mit Metadaten sind hier das Ordnungssystem der GEVER-Ablage sowie die Metadaten aus dem Entstehungszusammenhang der einzelnen Dokumente gemeint. Soweit eine Abbildung Ordnungssystem → Verzeichnungsebenen möglich ist, werden die Metadaten nach dem archivischen Standard ISAD(G) im EAD-Schema abgebildet. Zusätzliche Metadaten aus der GEVER-Anwendung werden als Key-Value-Paare im EAD-Schema abgebildet. Für die Abbildung technischer Metadaten ist das EAD-Schema nicht geeignet. Wir empfehlen hierfür PREMIS.

Empfehlung für ein SIP:

Die Granularität der Ablieferung definiert sich über die Menge der abgelieferten Dossiers und Unterlagen. Das heisst, eine Ablieferung besteht aus den Dokumenten und den zusätzlichen Informationen zum Entstehungszusammenhang dieser Dokumente (insbesondere die Ord-

nungsstruktur der Ablage in der abliefernden Stelle), alles zusammen in einem geeigneten Container zusammengefasst. Eine geeignete standardisierte Exportmöglichkeit für Dossierdaten ist in der GEVER-Spezifikation bis jetzt noch nicht vorgesehen. Da auch noch keine Archivschnittstelle als solche für den Import von GEVER-Daten in Archivsysteme definiert ist, können wir hier noch nicht auf Standards zurückgreifen¹⁰³.

Pilotprojekt Staatsarchiv St.Gallen (LARIS I)

Ausgangslage

Das Pilotprojekt wurde innerhalb jedes Staatsarchivs selbstständig durchgeführt. Natürlich erfolgte immer noch ein Erfahrungsaustausch sowie eine Unterstützung und Zusammenarbeit mit der KOST. In der Pilotprojekt-Phase erfolgte die Umsetzung des vorher gemeinsam erstellten Archivierungsprozesses. Dabei sollten elektronische Daten (Primärdaten und Metadaten) aus dem RIS übernommen und archiviert werden. Die Übernahme der Daten dient der Verifikation des Archivprozesses. Allenfalls muss dieser nach Abschluss der Übernahme überarbeitet werden. Ein weiteres Ziel war die Benutzung der übernommenen Daten voraussichtlich via scopeArchiv (vorhandene Archivsoftware). Gleichzeitig werden Richtlinien für die Verwaltungsstellen sowie Schnittstellendefinitionen für die Lieferanten erstellt.

Beschreibung

Die Erarbeitung der Anforderungen erfolgte in drei Workshops. Die Ergebnisse können dem Dokument 'Pilot Specification for Staatsarchiv St.Gallen' entnommen werden. Als System wurde ein Unixserver Power 5 mit 4 GB RAM unter AIX 8.3 genutzt. Genauere Angaben sind Bestandteil des Vertrages zwischen dem Staatsarchiv St.Gallen und der IBM. Diese Angaben sind in diesem Bericht nicht freigegeben, können jedoch bei begründetem Interesse beim Staatsarchiv oder bei der IBM nachgefragt werden. In Abbildung 41 ist eine schematische Darstellung der Pilotlösung dargestellt. Für den Input (Übertragung von Meta- und Primärdaten in das System) gibt es zwei Wege. Die erste Möglichkeit bestand darin, die Daten auf dem eigenen PC mit kolibri¹⁰⁴ zu erweitern und dann via FTP dem System zuzuführen. Die Erweiterung der Lösung sah vor, mit FTP und Nutzung vom des BatchBuilders die Daten an DIAS Core zu liefern. Anschliessend erfolgte die Benutzung des DIP via einfachen webbasierenden Aufruf oder direkt über scopeArchiv.

¹⁰³ Eine eCH-Archivschnittstelle basierend auf den Resultaten des BAR ist in Arbeit.

¹⁰⁴ Weitere Informationen über kolibri unter http://kopal.langzeitarchivierung.de/index_koLibRI.php.de.

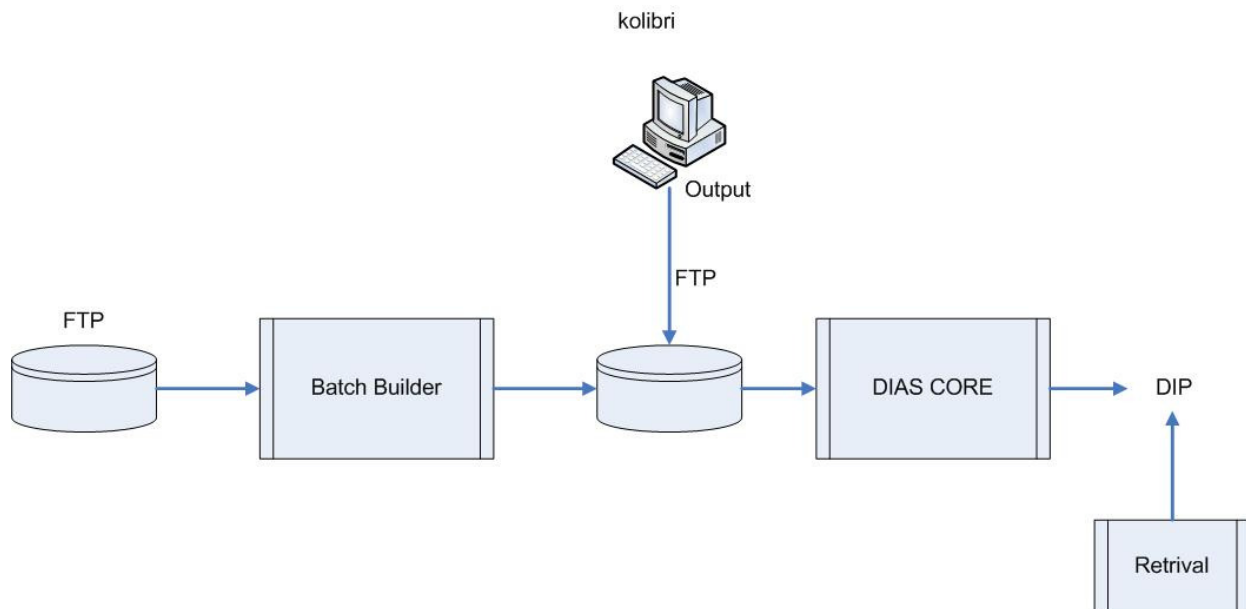


Abbildung 41: Schematische Darstellung Pilotlösung

Beschreibung RIS¹⁰⁵:

Unter <https://www.ratsinfo.sg.ch> veröffentlicht der Kantonsrat St.Gallen seine Geschäfte seit Ende Mai 2006 im Internet. Mit dem Ratsinformationssystem erreicht der Kanton St.Gallen eine neue Dimension bezüglich Information und Transparenz. Die Öffentlichkeit erhält einfach und schnell Einblick in staatliches Handeln. Der Internet-Auftritt des Ratsinformationssystems erlaubt den Zugriff auf die Beratungsunterlagen des Kantonsrates, geordnet nach Art des Geschäftes, Schlagwort, Thema oder Klassifikationsnummer. Zentrales Objekt des Ratsinformationssystems ist das Geschäft. Es wird mit bestimmten Metadaten erschlossen; ihm zugeordnet sind Dokumente, Referenzen auf weitere Geschäfte sowie Informationen zu beteiligten Personen und Personengruppen. Je nach Art des Geschäfts wird zusätzlich der Geschäftsablauf grafisch dargestellt. Die Orientierung über die laufenden Geschäfte erfolgt mittels Newsletter. Zu jeder Session stehen im Ratsinformationssystem auch die wichtigsten Sitzungsdokumente zur Verfügung. Ab der Junisession 2006 sind zudem die Abstimmungsergebnisse zu den behandelten Geschäften und das Stimmverhalten der einzelnen Ratsmitglieder dargestellt – deren Erfassung beansprucht allerdings einige Tage. Schliesslich stehen jeweils am Folgetag die Wortmeldungen als Audio-Dateien zur Verfügung, später auch in schriftlicher Form. Das integrale gedruckte Kantonsratsprotokoll entfällt; gedruckte Auszüge des Protokolls werden nur noch auf Verlangen zugestellt.

Beweggründe für RIS als Piloten:

- Die ausgewählten Daten stehen bereits heute im Internet jedermann zur Verfügung. Es müssen keine Vorsichtsmassnahmen in Bezug auf den Datenschutz getätigt werden.
- Es besteht eine Exportmöglichkeit. Dabei wird dieselbe Schnittstelle wie für den Export in das Intranet/Internet genutzt.

¹⁰⁵ Text aus dem Internetauftritt von RIS übernommen.

- Kultur: Das RIS wird durch die Staatskanzlei betreut. Die Mitarbeitenden in der Staatskanzlei stehen Neuerungen und neuen Ideen offen gegenüber.
- RIS ist eine der Applikationen, welche auf der für den Kanton St.Gallen festgelegten Produktstandards ECM-Plattform FileNet P8 betrieben wird. Es werden noch ca. zwölf weitere Applikationen auf dieser Plattform betrieben und fortlaufend neue in Betrieb genommen.

Datengewinnung aus RIS:

Für die Datengewinnung aus RIS konnte die bereits vorhandene Exportfunktion verwendet werden. Mit dieser Exportfunktion werden normalerweise die RIS-Daten in das Intranet/Internet exportiert (Schnittstelle Internet). Durch die Nutzung der vorhandenen Exportfunktion war die Auswahl der Daten begrenzt und eigene Anforderungswünsche waren nicht möglich. Ebenso wäre es mit Hilfe von externer Unterstützung auch möglich gewesen die Dateinamen sprechend zu gestalten. Aus diesem Grunde bestehen die Dateinamen aus Buchstaben und Zahlenkombinationen (eine interne Namensvergabe). Es wurde auf beide Umsetzungen durch fremde Unterstützung aus Kostengründen verzichtet. Dadurch wäre der Rahmen des Pilotprojektes gesprengt worden. Für den Pilotversuch waren die gelieferten Daten vollkommen ausreichend. Die Umsetzung wäre nur kosmetischer Art gewesen. Aus technischer Sicht und für die Zukunft stellt eine Umsetzung der Anforderungen jedoch kein Problem dar.

Ein Geschäft besteht aus einer XML-Datei (Metadaten) und aus einer variablen Anzahl PDF-Dateien (Primärdaten). Beim ersten Export der Daten wird eine sogenannte Masterdatei auch im XML-Format geliefert, welche alle weiterführenden Informationen zu den Geschäften, Beteiligten etc. beinhaltet. Für den eigentlichen Export wird diese jedoch nicht mehr benötigt. Falls in Zukunft mehr Inhalte aus dem RIS exportiert werden sollten, müssen die Daten zuerst aus der Masterdatei (oder dem System direkt) gelesen und daraus die abzuliefernden Objekte erstellt werden. Dies war jedoch keine Anforderung an das Pilotprojekt. Hierfür wäre auch eine eindeutig definierte Standardschnittstelle von Vorteil.

Datenerweiterung:

Für die Übernahme in DIAS musste die Geschäft-XML, falls nicht vorhanden, noch um die Angabe Datum (Abschlussdatum des Geschäftes) erweitert werden. Ebenso wurden für den eigentlichen Transfer mehrere Geschäfte in ein File (Datei) zusammengezogen. Dadurch entstand die Ablieferungsdatei. Eine Ablieferung besteht aus einer Ablieferung-XML, welche alle Informationen der Geschäft-XML sowie im Kopfteil weitere Angaben (woher?, durch wen?, wer ist verantwortlich? etc.) zur Ablieferung enthält. Der restliche Inhalt der Ablieferung enthält die einzelnen Primärdaten in PDF, welche pro Geschäft in einem Ordner abgelegt sind. Schlussendlich wird alles als eine ZIP-Datei gespeichert.

Datentransfer:

Der Datentransfer erfolgt mittels FTP. Mittels einfachem 'drag and drop' werden die ZIP-Dateien in den passenden Ordner gezogen. Die ZIP-Datei (Ablieferung) wird in den Ordner BBInBox auf dem Unixserver gelegt.

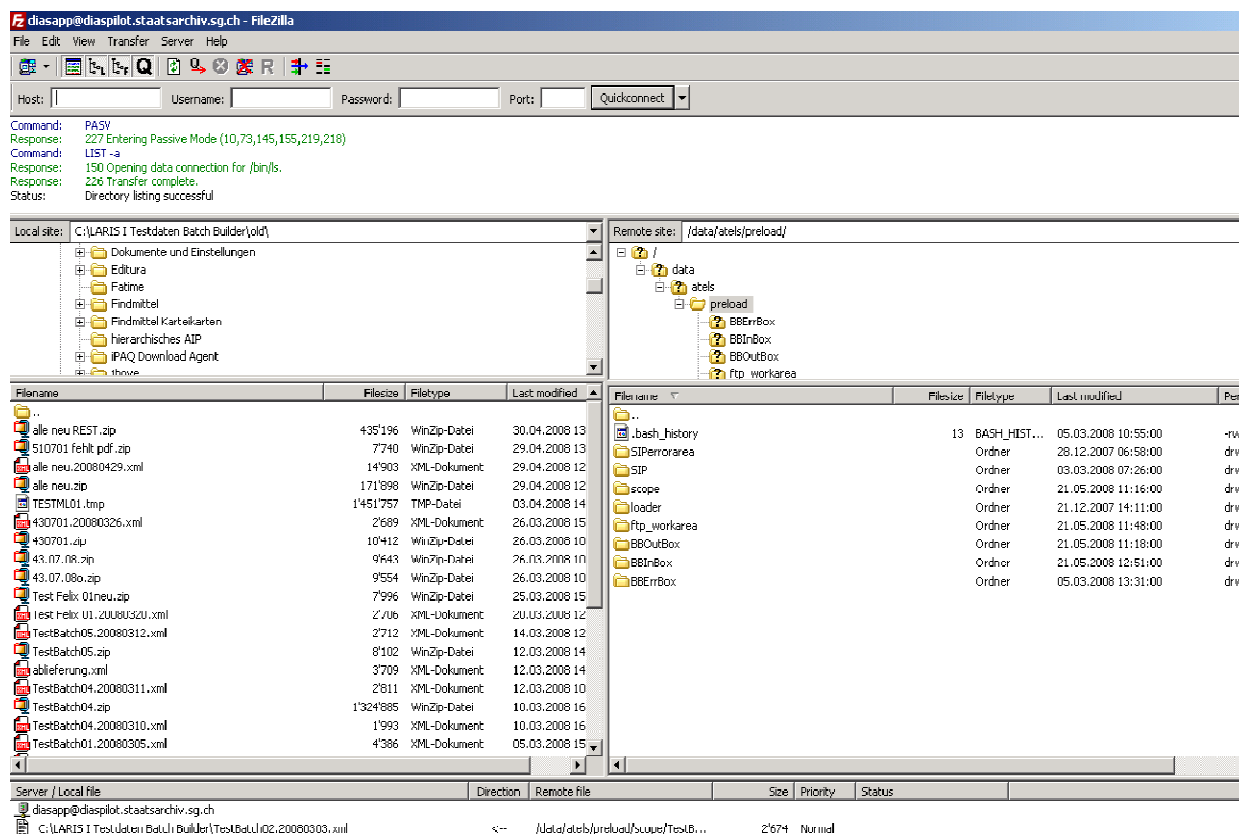


Abbildung 42: Datenübermittlung an DIAS

Funktion DIAS:

Nach Prüfung der Vollständigkeit der ZIP-Datei durch mehrmaliges Lesen und Prüfen der Grösse der Datei startet DIASBatchBuilder automatisch. Der DIASBatchBuilder ist die eigentliche 'Applikation' von DIAS. Hier können Parameter gesetzt, Anpassungen am Datenmodell getätigt, Schnittstellen definiert werden etc. Dabei werden im ersten Schritt die Inputdateien für DIAS mit den Ergänzungen aus kolibri (Dateiformat, Hashwert etc.) erstellt und in den Ordner BBOutBox gestellt. Diese Dateien könnten auch direkt in DIAS geladen werden (kann über einen Eintrag im Parameterfile definiert werden). Für Testzwecke und zur Kontrolle wurde dieser Zwischenschritt eingefügt. Gleichzeitig wird die Inputdatei für scopeArchiv (Übernahmeassistenten) erstellt und im Ordner scope gespeichert. Die scopeArchiv Übernahme-Datei wird gemäss Schema der Firma scope solutions ag im DIASBatchBuilder erstellt. Die scopeArchiv-Datei wird nachher mit dem Übernahmeassistenten von scopeArchiv in das System geladen. Dies entspricht einer normalen Übernahme von Metadaten und ist zum heutigen Zeitpunkt der Standardprozess für die Übernahme von elektronischen Metadaten in scopeArchiv. Anschliessend werden die Daten aus dem Ordner BBOutBox in den Ordner SIP/ORG10DIASMETS10

kopiert. Dieser Ordner wird durch einen Daemon¹⁰⁶ bzw. Scheduler überprüft. Sobald ein Input vorhanden ist, startet DIAS einen Prozess, welcher die Inputdaten verarbeitet und in DIAS ablegt. Die Metadaten werden dabei in Tabellen geschrieben. Die Primärdaten wurden für den Piloten auf übliche Disk geschrieben und täglich auf einer weiteren, an einem anderen Standort befindlichen Disk gesichert. Gleichzeitig verwaltet DIAS alle Geschäfte als eigene Objekte, welche aus den XML-Dateien Geschäft und Mets (erweiterte XML-Datei mit Angaben zu Datei etc.) sowie den Primärdateien besteht. Für den Pilotbetrieb wurden diverse Logfiles eingeführt. Ein Logfile (Status) diente der Feststellung, wer, was, wann auf dem System durchgeführt hat. Ein zweites Logfile (Ablieferung) enthielt Angaben zu den Ablieferungen. Für einen produktiven Betrieb müssen die Logfiles noch erweitert werden. Dies soll der besseren Überwachung der Prozesse dienen.

Benutzung mit scopeArchiv:

Die scopeArchiv-Datei wurde mit dem Übernahmeassistenten von scopeArchiv geladen. Dies ist das übliche Verfahren um Metadaten in XML zu laden. In diesem Projekt haben wir auf eine Automatisierung verzichtet. Eine automatisierte Schnittstelle müsste in Zukunft von scope solutions ag angeboten werden. Für das Pilotprojekt wurden die Daten in einen Testbaum geladen. Es besteht eine direkte Verlinkung von den Verzeichnungseinheiten zu DIAS.

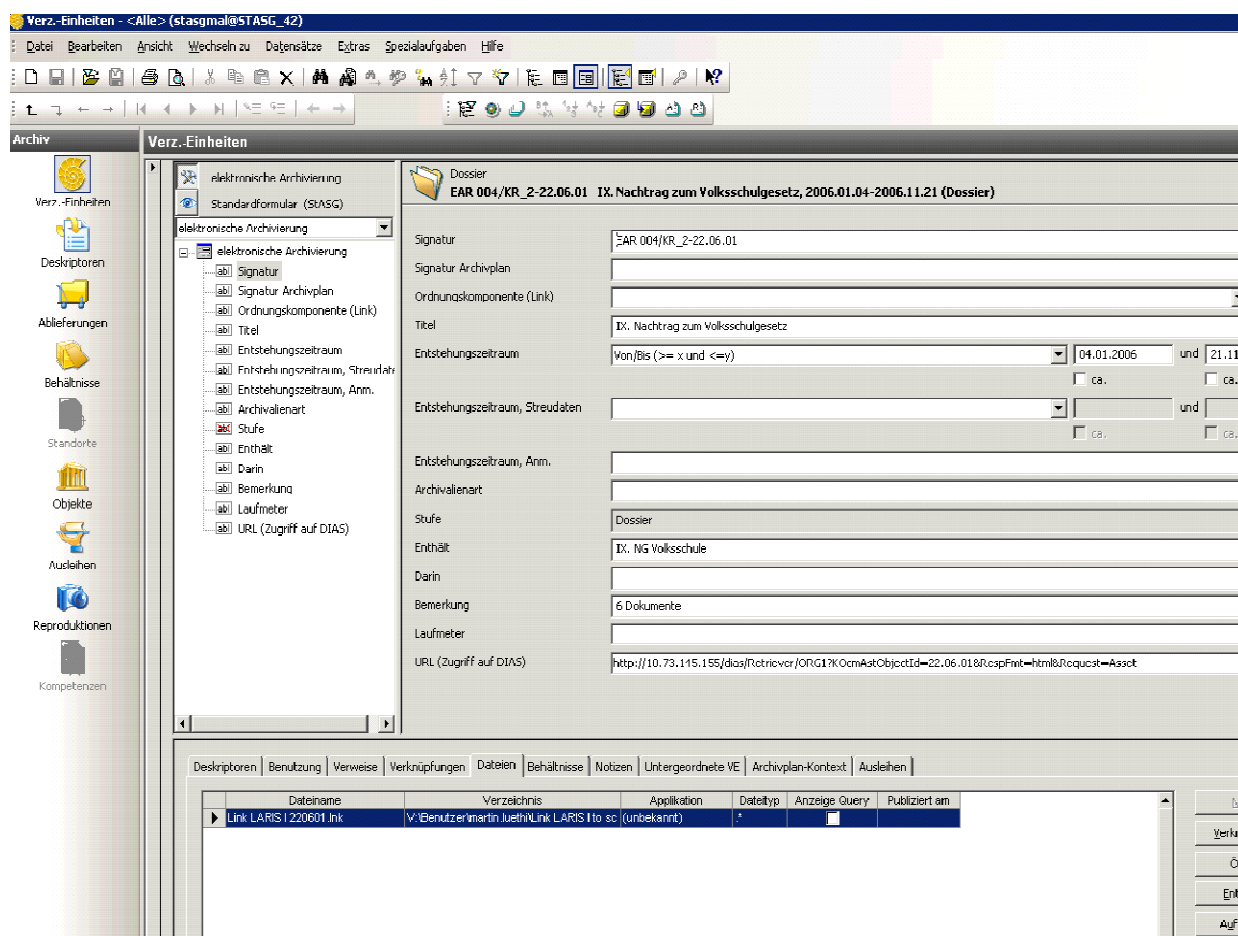


Abbildung 43: scopeArchiv-Verzeichnungseinheit Dossier URL-Link

¹⁰⁶ Ein Programm, das im Hintergrund abläuft und bestimmte Dienste zur Verfügung stellt oder Tätigkeiten durchführt.

Benutzung ohne scopeArchiv (Weblösung):

Im Projekt LARIS I sollte der Aufruf auf die Objekte nicht nur direkt aus scopeArchiv, sondern auch mit einer einfachen Weboberfläche durchgeführt werden. Auf der Weboberfläche besteht die Möglichkeit, Suchen nach Dossier-Nummer, interner und externer Adresse durchzuführen. Das gesamte Objekt enthält die Primärdaten (PDF) sowie die Metadaten (geschaeft und mets). Die Geschäft-XML entspricht 1:1 der gelieferten Metadatendatei aus RIS. Die METS-XML enthält sowohl die Informationen der Geschäft-XML sowie die Erweiterungen aus JHOVE und kolibri.

Fazit Pilotprojekt LARIS I

Nach der Fertigstellung des Piloten bestand die Möglichkeit die Daten via eine Webschnittstelle im Browser anzusehen oder den Aufruf direkt über die Archivsoftware scopeArchiv zu tätigen. Die Datenübernahme erfolgte aus dem Ratsinformationssystem. Für den Piloten wurden die Daten noch mit ergänzendem Inhalt angereichert. Da eine optimale Anpassung der Daten nur durch den externen Partner erfolgen konnte und dies kostenpflichtig gewesen wäre, wurde darauf verzichtet. Die Einbettung von scopeArchiv war ein grosser Erfolg. Es ist nun möglich, direkt aus scopeArchiv auf DIAS zuzugreifen und Dossiers anzusehen. Der Zugriff erfolgt entweder über eine Linkdatei oder eine URL in scopeArchiv. Diese Einbettung konnte ohne Unterstützung von scope solutions ag durchgeführt werden. Der Zugriffsprozess erfolgt wie oben erwähnt über einen Browser bzw. über scopeArchiv. Ein weiteres sehr wertvolles Ergebnis aus dem Projekt war die Bestätigung, dass durch ein konkret umgesetztes Projekt am 'lebenden' Objekt die Erfahrungssammlung viel höher liegt als bei einem Projekt nur auf Papier.

Stillstand

Eigentlich wäre als nächster Schritt die Fortsetzung des Piloten vorgesehen gewesen mit dem Ziel, den Betrieb von DIAS aufrechtzuerhalten und mit einer weiteren Dienststelle Daten bzw. Unterlagen zu übernehmen. Die Schwierigkeit bestand jedoch darin, geeignete Lieferanten von elektronischen Daten zu finden. Leider war es nicht möglich, einen internen Partner zu finden, welcher im Besitz von bewerteten und für die elektronische Langzeitarchivierung vorgesehenen Daten war. Wir haben uns auch aus diesem Grund entschieden, erst mit einer Langzeitarchivierungslösung weiterzumachen, nachdem die folgenden offenen Punkte geklärt sind: Zuerst möchten wir uns einen Überblick über die vorhandenen Applikationen und alten Datenbestände im Kanton St.Gallen machen, konkret: Was ist wo und in welchem Zustand überhaupt vorhanden. Als zweiter Punkt ist die Definition einer Archivschnittstelle massgebend. Ohne eine solche Schnittstelle wird kein Anbieter eine Möglichkeit der Auslesung von Daten selbstständig forcieren. Der dritte Punkt besteht darin, sich eine Übersicht der bereits vorhandenen Langzeitarchivierungslösungen zu erarbeiten und diese miteinander zu vergleichen (Marktvergleich).

Projekt Rubicon¹⁰⁷

Als Nachfolgeprojekt von LARIS wurde deshalb 2008/2009 im Staatsarchiv St.Gallen das Projekt Rubicon durchgeführt. Dabei standen stichwortartig folgende konzeptionellen Fragen im Vordergrund:

- Grundsätzlich dient Rubicon als Reflexionsphase: Sammeln der bisher getätigten Erfahrungen und deren Umsetzung und Auswirkung)
- Durchführung eines Marktvergleiches (was für Langzeitarchivierungslösungen gibt es)
- Schnittstellenproblematik / Speicherproblematik
- Vorarbeiten Bereich ECM (was sind notwendige Anpassungen/Erweiterungen)
- Aussonderungsprozess (der gesamte Lebenszyklus von der Entstehung der Unterlagen über die Nutzung bis zur Langzeitarchivierung im Staatsarchiv)

Nach Rubicon startet Horreum I¹⁰⁸ mit folgenden Zielen:

- Erarbeitung einer Strategie für das Staatsarchiv im Bereich Langzeitarchivierung
- Dienstleistungen des Staatsarchivs im Kanton aufzeigen
- Bedarfsanalyse: Überblick des Ablieferungspotentials Kanton St.Gallen und deren Ausprägung (DB, Systeme, ECM etc.)
- Staatsarchiv als Partner positionieren
- Vertiefung des Aussonderungsprozesses (Festlegung Austauschstandard, Formate, Validierung, Konvertierung, Schnittstellen)
- Betriebsmodell und Business Modell mit Staatsarchiv als Betreiber
- Aufbau Grundsystem (wenn möglich und sinnvoll)

¹⁰⁷ Der Abschluss des Projektes lag zur Zeit der Tagung noch nicht vor.

¹⁰⁸ Über die Ergebnisse wird an der nächsten Tagung berichtet.

Abkürzungsverzeichnis

Auswahl fachlicher Abkürzungen:

AIP	A rchival I nformation P ackage
BLOB	B inary L arge O bject
DC	D ublin C ore
DCMI	D ublin C ore M etadata I nitiative
DIMAG	D igitales M agazin
DIP	D issemination I nformation P ackage
DROID	D igital R ecord O bject I dentification
OAIS	O pen A rchival I nformation S ystem
GEVER	G eschäfts v erwaltung
ISAD/G	I nternational S tandard A rchival D escription (G eneral)
JHOVE	J ournal Sage/ H arvard O bject V alidation E nvironment
LZA/dLZA	L angzeit a rchiv, d igitales L angzeit a rchiv
PAIMAS	P roducer- A rchive I nterface M ethodology A bstract S tandard
PREMIS	P reservation M etadata: I mplementation S trategies
PURL	P ersistent U niform R esource L ocator
SIP	S ubmission I nformation P ackage
URL	U niform R esource L ocator