

Archivierung digitaler Unterlagen im Staatsarchiv Basel-Stadt Rückblick auf den Infrastrukturaufbau und erste Erfahrungen

Von LAMBERT KANSY und MARKUS LOCH

Der Beitrag beschreibt rückblickend den Aufbau der Infrastruktur für die digitale Archivierung im Staatsarchiv Basel-Stadt in den Jahren 2010–2012 sowie erste Erfahrungen mit den realisierten Werkzeugen. Die Darstellung fokussiert auf die Implementierung einer standardisierten Ablieferungsschnittstelle (SIP), die Neudefinition der Struktur der Archivpakete (AIP) sowie die Entwicklung des digitalen Magazins auf technischer Ebene. Abschließend werden mögliche Ausbauschritte skizziert.

Einleitung

Das Staatsarchiv Basel-Stadt führte von 2008 bis 2011 das Projekt Informatisierung III durch, das sich neben der Digitalisierung von Archivgut und der Integration des Archivinformationssystems in die kantonale Systemlandschaft mit dem Aufbau von Infrastrukturen und Geschäftsprozessen zur Archivierung digitaler Unterlagen befasste. 2010 wurde der damalige Projektstand an der 14. Tagung des Arbeitskreises für die Archivierung von Unterlagen aus digitalen Systemen vorgestellt.¹ Der nachstehende Beitrag beschreibt die seither erreichten Ergebnisse und stellt die gemachten Erfahrungen vor, bevor er mit einem Ausblick auf die weiteren Entwicklungslinien schließt.

Ergebnisse im Überblick

Im zweiten Quartal 2010 wurden die in der ersten Etappe des Projekts entwickelten neuen Komponenten des Archivinformationssystems, das Modul Ingest und der Ingestserver sowie die Anbindung des Repositorysystems FedoraCommons, im Rahmen eines Releasewechsels von

¹ Vgl. Lambert Kansy: Archivierung digitaler Unterlagen im Staatsarchiv Basel-Stadt. Rückblick auf den Infrastrukturaufbau und erste Erfahrungen. In: Neue Entwicklungen und Erfahrungen im Bereich der digitalen Archivierung: von der Behördenberatung zum Digitalen Archiv. 14. Tagung des Arbeitskreises *Archivierung von Unterlagen aus digitalen Systemen* vom 1. und 2. März 2010 in München. Hg von Susanne Wolf, München 2010, S. 5562, http://www.staatsarchiv.sg.ch/home/auds/14/_jcr_content/Par/downloadlist/DownloadListPar/download_6.ocFile/Text%20Kansy.pdf. Alle Links in diesem Beitrag wurden am 20. September 2012 überprüft.

scopeArchiv eingeführt. Mit dem damit erreichten Stand konnten 2010 und 2011 im Rahmen eines Pilotversuchs in zwei Ablieferungen erstmals originär digitale Unterlagen aus dem System ARAP (elektronisches Polizeijournal der Kantonspolizei Basel-Stadt) übernommen werden.² Die dabei gesammelten – geringen – Erfahrungen sind in den Ausbau der Ingestwerkzeuge eingeflossen. Sie betreffen die Handhabung der Werkzeuge sowie die Kommunikation zwischen Aktenbildner resp. abliefernder Stelle und dem Staatsarchiv. Darüber hinaus wurden grundsätzliche Weiterentwicklungswünsche festgehalten, zum Beispiel der Aufbau einer generalisiert anwendbaren Ablieferungsschnittstelle.

Insgesamt wurden acht Arbeitspakete sehr unterschiedlichen Umfangs definiert:

1. Festlegung eines Persistent Identifier (PID) für die Archival Information Packages (AIP)
2. Ausbau der Protokollierung des Ingestprozesses
3. Verknüpfung zwischen der Ingest- und der Ablieferungsverwaltung
4. Betrieb der FedoraCommons-Datenbank auf einer Oracle-Datenbank statt wie bisher auf einer PostgreSQL³-Datenbank
5. Implementierung einer Standardschnittstelle für SIP und Anpassung der AIP-Definition
6. Verbesserter DIP-Viewer als Alternative zum Standard-Disseminator von FedoraCommons
7. Schaffung einer Möglichkeit mit hoher *usability*, um innerhalb des SIP browsen zu können, zum Zwecke der stichprobenhaften Kontrolle bei der manuellen Validierung des SIP
8. Realisierung eines SIP-Generators.

Unabhängig von diesem Ausbau der Infrastruktur wurden die 2009 definierten Geschäftsprozesse – von der Bewertung über Anbieten und Abliefern bis zur Erschließung und Magazinierung – weiter ausgearbeitet und in Form eines Organisationshandbuchs mit der bestehenden Aufbauorganisation abgestimmt. Dabei wurde an der Zerteilung der prospektiven Bewertung festgehalten: Die Makrobewertung hat die Bewertung der Aufgaben des Aktenbildners zum Gegenstand, die Mikrobewertung stellt eine Bewertung der angebotenen Unterlagen resp. des aktenbildenden Systems dar. Im Rahmen des Übernahmeprozesses findet bei vorliegender prospektiver Bewertung lediglich noch eine Kontrolle des Angebots und der Ablieferung statt, mit der die Übereinstimmung von zu übergebenden Unterlagen mit in der Bewertung festgelegten Kriterien überprüft wird.

Im Folgenden wird das zentrale Arbeitspaket AP5 genauer dargestellt.

² Zum Pilotprojekt der ARAP-Übernahmen vgl. Lambert *Kansy*: Pilot-Übernahme elektronischer Unterlagen durch das Staatsarchiv Basel-Stadt, 2011, <http://www.archiversum.com/wp-content/uploads/2011/07/Kansy-02.pdf>.

³ PostgreSQL (oft kürzer als Postgres bezeichnet) ist ein Open Source Datenbankmanagementsystem.

Implementierung einer Standard-Schnittstelle für Übernahmepakete (SIP)

Bei der Entwicklung der Ingestwerkzeuge hat das Staatsarchiv Basel-Stadt als Pilotübernahme die Ablieferung von Unterlagen aus dem System ARAP der Kantonspolizei definiert. Dafür wurde ein spezifisches SIP definiert, das auf dem XML-Schema des Übernahme-Assistenten von scope solutions basiert. Dieses Übernahmepaket wird bis auf Weiteres bei der Ablieferung aus ARAP verwendet. Von Beginn an wurde jedoch diese Pilotlösung nicht als Basis für eine durch verschiedene Quellsysteme zu beliefernde SIP-Schnittstelle verstanden. Die Standard-Ablieferungsschnittstelle sollte auf bestehenden Standards aufbauen, um die Chancen zu vergrößern, dass die Quellsysteme der Aktenbildner diese Übergabeschnittstelle bedienen können. Ziel ist es, die Anzahl der durch das Archiv zur Verfügung gestellten Ablieferungsschnittstellen möglichst gering zu halten, um den Bewirtschaftungsaufwand langfristig überhaupt bewältigen zu können.

Ende 2009 stellte das Schweizerische Bundesarchiv eine von ihm entwickelte Ablieferungsschnittstelle vor.⁴ Eine eingehende Überprüfung dieser SIP-Definition durch die KOST im Rahmen ihres Projekts bento ergab 2010, dass nur wenige Modifikationen notwendig waren, um die Schnittstelle auch für andere Archive nutzbar zu machen. So wurde die Spezifikation um die Möglichkeit ergänzt, zusätzliche Metadaten, die nicht in den Elementen von metadata.xsd abgebildet werden können, dennoch in strukturierter und damit auch maschinell verarbeitbarer Form im SIP abliefern zu können. Im Rahmen des bento-Projekts wurden zwei Wege definiert: zum einen die Definition von Key-Value-Paaren und zum anderen die Integration von Elementen, die in fremden XML-Schemata definiert sind. Auf diese Weise können beliebige weitere Metadaten in der Metadatenfile metadata.xml des SIP übernommen werden.⁵ Die SIP-Spezifikation des Bundesarchivs, ergänzt um die Erweiterung der Key-Value-Paare, soll 2012 als eCH-Standard verabschiedet werden.⁶ Aufgrund der sich abzeichnenden Akzeptanz dieser Ablieferungsschnittstelle erscheint deren Implementierung als Standard-Ablieferungsschnittstelle gerechtfertigt.

Neudefinition der Metadatenstruktur der Archivpakete (AIP)

Die Neudefinition des Standard-SIP konnte sich auf die Implementation eines – in Entwicklung befindlichen – Standards abstützen. Bei der Bildung der Archivpakete (AIP) hingegen wurde mit EAD ein bestehender und verbreiteter Metadatenstandard durch das – ebenfalls im Rahmen des KOST-Projekts bento entwickelte – XML-Schema xIsadg ersetzt.⁷ Grund hierfür ist die mehrdeutige Abbildbarkeit von ISAD(G)-Metadaten in EAD und die Komplexität dieses Standards, die auf seine ursprüngliche Zielsetzung (Aufbereitung gedruckter Findbücher für die Präsentation im Internet) zurückgeht. Beides macht EAD aus Sicht des Staatsarchivs nicht zum idealen Standard für die Abbildung von beschreibenden archivischen Metadaten innerhalb

⁴ Siehe <http://www.bar.admin.ch/themen/00876/00877/>.

⁵ Zum Projekt bento der KOST siehe http://kost-ceco.ch/cms/index.php?bento_de.

⁶ Zum Verein eGovernment Schweiz eCH siehe <http://www.ech.ch>.

⁷ Zur Spezifikation von xIsadg siehe http://kost-ceco.ch/cms/index.php?bento_de.

eines AIP, die aufgrund der Granularität der AIPs auf den Verzeichnisstufen Dossier und Dokument (sowie allenfalls Serie) keinesfalls Umfang und Struktur eines Bestand beschreibenden Findbuchs haben.

Das xIsadg-Schema basiert auf dem im Projekt eDavid des Stadtarchivs Antwerpen entwickelten XML-Schema isad.xsd und erweitert es vor allem hinsichtlich der Möglichkeit hierarchischer Schachtelung von Verzeichniseinheiten.⁸ Zudem wurde die Definition einiger ISA-D(G)-Felder im XML-Schema angepasst. Auch die Frage der Redundanzvermeidung wurde in xIsadg umzusetzen versucht, indem Vererbungsregeln definiert werden, die freilich nicht durch das Schema validiert werden können. Es wurde in xIsadg die Möglichkeit der Übernahme beliebiger zusätzlicher Metadaten in Form von key-value-Paaren geschaffen, ähnlich wie bei der Erweiterung der SIP-Spezifikation. Zusätzlich können hier auch fremde XML-Schemata integriert werden. Zwecks Referenzierbarkeit von Primärdaten, weiteren Metadaten Dateien sowie Archivpaketen wurden entsprechende Elemente definiert:

- `aipReference` – für die Referenzierung von Archivpaketen
- `primaryDataLocator` – für die Referenzierung von Primärdaten
- `secondaryDataLocator` – für die Referenzierung von Primärdaten über deren Beschreibung in anderen XML-Dateien.

Während die beiden letzten Referenzierungsmöglichkeiten für die Verwendung innerhalb eines AIP geschaffen wurden, wird mit dem Verweis auf ein AIP die Nutzung von xIsadg für den Austausch von beschreibenden Metadaten zwischen Archivinformationssystemen möglich. Die für die Nutzung von xIsadg innerhalb eines AIP notwendigen Anpassungen wurden vom Staatsarchiv Basel-Stadt gemeinsam mit der Geschäftsstelle der KOST definiert.

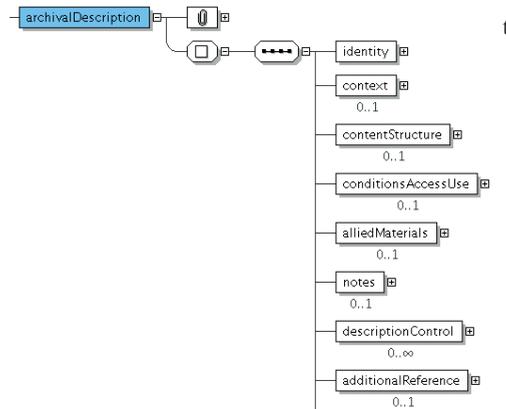


Abb. 1: Grundstruktur von xIsadg.

⁸ Zum Projekt eDavid siehe <http://www.edavid.be/eng/index.php>.

Die übrige Struktur des AIP wurde unverändert belassen. Es ist als JAR-Container definiert, der Metadaten und Primärdaten in drei Verzeichnissen getrennt enthält. Die Metadaten bestehen aus zwei Dateien im Verzeichnis metadaten: den technischen und administrativen Metadaten gemäß dem PREMIS-Standard, Version 2.1, und den beschreibenden Metadaten gemäß dem xIsadg-Schema.

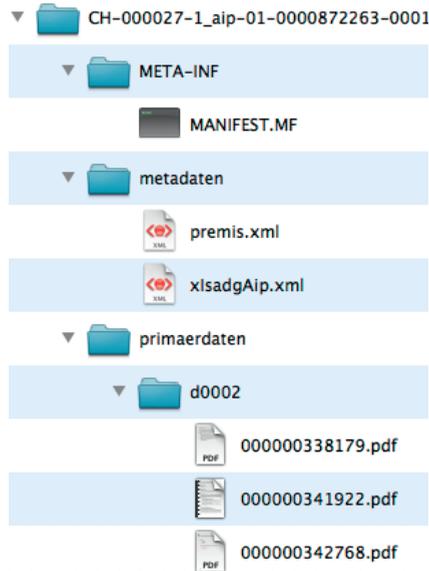


Abb. 2: AIP-Struktur.

- Das Verzeichnis META-INF enthält mit der Datei MANIFEST.MF Metadaten über das AIP und dessen Verzeichnis- und Dateiinhalte.
- Das Verzeichnis primaerdaten enthält in beliebiger Schachtelung in Verzeichnissen die Primärdaten.
- Der Zugriff auf die Primärdaten erfolgt von den beschreibenden Informationen in der Datei xIsadgAip.xml aus über eine xPointer-Verknüpfung mittels der Datei premis.xml.

Die JAR-Datei des Archivpakets erhält als Dateinamen den Persistent Identifier (PID), der mit dem ISIL-Code des Staatsarchivs beginnt, und nachfolgend die Art des Informationspakets angibt. Die weiteren Abschnitte des PID setzen sich aus der Kennung des Archivinformationssystems zusammen, gefolgt von der dort vorhandenen ID für die mit dem AIP referenzierte Verzeichnungseinheit sowie einer Versionsnummer des AIP.



Abb. 3: Beispiel für PID des AIP.

Infrastrukturausbau

Parallel zu dem Ausbau der Ingestwerkzeuge erfolgte ein Ausbau der Infrastruktur des digitalen Magazins (digiMag).

Im Rahmen des Projekts Informatisierung III wurde somit die vom Staatsarchiv Basel-Stadt im Bereich der Archivinformatik eingesetzte IT-Infrastruktur deutlich ausgebaut. Bestand zuvor das Archivinformationssystem scopeArchiv aus einem Datenbankserver, einem Applikationsserver für den Online-Archivkatalog sowie den lokalen Client-Anwendungen, so wurden die folgenden drei- virtualisierten – Server im Rahmen des Projekts zusätzlich beschafft:

- FTP-Server für die Entgegennahme der Ablieferungspakete (SIP)
- scopeIngest-Application-Server (mit IIS-Webserver für den SIP-Browser)
- createBento-Webserver (Apache/Tomcat-Webserver für den SIP-Generator).

Daneben wurde auch das digitale Magazin, sowohl in Hinblick auf seine Speicherkapazität als auch in Hinblick auf seine Architektur, massiv ausgebaut. Die neuen Komponenten des digiMag sind:

- FedoraCommons-DB-Server
- FedoraCommons-Application-Server (mit Apache/Tomcat)

Mit Blick auf das OAIS-Referenzmodell deckt die Infrastruktur (Hardware und Software) des Staatsarchivs Basel-Stadt nach Abschluss des Projekts die Prozesse Ingest, Storage, Data Management und – wenn auch nur in rudimentärer Form – Access ab. Gegliedert nach Software-Schichten ergibt sich die nachstehende Übersicht.

Infrastrukturlayer	OAIS-Prozess	StABS-Prozess	Anwendung/Lösung
Archival Information System	Access, Data Management	Erschließung, Benutzung	scopeArchiv, scopeQuery
Digital Archiving Software	Ingest, Storage	Übernahme, Magazinierung	FTP-Server, scopeIngest und FedoraCommons
Storage Management Software	Storage	Magazinierung	FedoraCommons, Primärspeicher
Storage Hardware	Storage	Magazinierung	Primärspeicher und arcun

Tabelle 1: Infrastruktur-Systematik

Ausbau des digitalen Magazins

Mit Beginn des Informatisierungsprojekts Info III gab es im Staatsarchiv Basel Stadt ein Netzwerklaufwerk (Protokoll: SMB/CIFS) zur Speicherung von Dateien. Dieses bestand aus einer handelsüblichen NAS-Speicherbox mit 5 Festplatten, die als RAID 5 konfiguriert waren und netto 3,5 TB Platz enthielten. Dieses Laufwerk war an allen Desktop Rechnern des Archivs gemountet. Ein kleiner Kreis von Archivmitarbeitern war mit Schreibrechten ausgestattet und konnte dort Digitalisate und digitales Archivgut in Form von einzelnen Dateien ablegen. Die übrigen Mitarbeitenden verfügten nur über Lesezugriff. Zur Sicherung der Daten stand ein weiteres gleichartiges Laufwerk als Spiegel zur Verfügung, auf das der Verzeichnis-Ast des ersten Laufwerkes repliziert wurde. Auf diesen Spiegel konnte ausschließlich der Systemadministrator zugreifen; er entsprach also in etwa einem Schattenarchiv (dark archive).

Beim Replizieren wurde überprüft, ob Dateien auf dem Spiegel verwaist waren, weil die Ursprungsdatei auf dem Urbild gelöscht oder verändert worden war. Falls das vorkam, wurde der Administrator über eine E-Mail informiert, und es war seine Aufgabe, dieser Abweichung nachzugehen: Durch Nachfragen bei den entsprechenden Archivmitarbeitern musste er herausfinden, ob die Löschung der Datei bewusst oder aus Versehen geschah.

Mit dieser einfachen Bewirtschaftung wurde eine sehr hohe Sicherheit gegen Ausfall der Dateien erzielt. Insbesondere das Risiko einer versehentlichen Löschung wurde so vermindert. Ein Backup hilft im Falle einer versehentlichen Löschung nur dann, wenn der Irrtum innerhalb der Backup-Frist bemerkt wird.

Als die Datenmenge im Laufe des Informatisierungsprojekts Info III zunahm, war schnell klar: Es wird ein größerer Speicher benötigt. Die neue Lösung sollte aber nicht laufwerkbasierend sein, sondern in einer Netzwerkschnittstelle bestehen, der die Dateien übergeben werden (Repositoryum). Ein solches Repositoryum funktioniert ähnlich wie eine Garderobe mit Garderobenmarken:⁹ Man bringt ein Kleidungsstück – in unserem Fall eine Datei – und erhält dafür eine Garderobenmarke. Dem entspricht in unserem Fall eine Zeichenkette als Persistent Identifier. Diese PID wird im Archiv Informations System abgelegt. Möchte man die Datei wieder haben, geht man (übers Netzwerk) zum Repositoryum, übergibt die PID und erhält dafür die Datei, so wie man an der Garderobe sein Kleidungsstück gegen Vorlage der Garderobenmarke zurück erhält.

Um es vorwegzunehmen: Die produktive Umgebung arbeitet immer noch (März 2012) mit einer laufwerkbasierenden Lösung. Allerdings ist das Laufwerk (der SMB-Server) auf einen größeren Host mit mehr Festplatten umgezogen. Das Testsystem befindet sich auf demselben Host und ist bereits auf Repositoryum-Betrieb umgestellt. Als entsprechende Repositoryum Software dienen FedoraCommons und eine Oracle Datenbank. Die Umstellung des Produktivsystems wird im zweiten Quartal 2012 erfolgen.

Die Vorbereitung des Umzugs vom alten System mit den beiden NAS-Speicherboxen auf ein neues, größeres System begann mit der Evaluierung eines neuen Speichers samt Rechner

⁹ Diese von Kai *Naumann* an der Tagung des AK AUdS 2012 gebrauchte Analogie haben wir bei der Ausarbeitung des Vortrags gerne aufgegriffen, weil sie den Sachverhalt anschaulich schildert.

(Host). Die zugrunde liegenden drei Komponenten Betriebssystem – Volume Manager – Dateisystem standen im Fokus der Evaluierung.

Die Anforderungen an das neue System lauteten grob umrissen wie folgt:

1. Eigener Host mit der Möglichkeit, Shellscrippts auszuführen
2. Speichervolumen deutlich grösser als 10 TB und weiter ausbaufähig
3. Möglichst preiswert (<1000 SFr/TB in der Anschaffung, weitere 1000 SFr/TB und Jahr im Betrieb)
4. Garantie der Bitstream Preservation möglichst schon im Betriebssystem
5. Thin Provisioning¹⁰ einfach und elegant möglich, da mehrere Virtuelle Systeme den Speicher nutzen
6. Gutes Zusammenspiel der Hard- und Softwarekomponenten.

Keine dieser Vorgaben war für sich alleine zwingend erforderlich; ein Plus im einen Bereich konnte ein Minus in einem anderen Bereich kompensieren. Es gab auch Parameter, die das Disk-Subsystem nicht zu haben brauchte, die allerdings oft bei Offerten als Vorteil angepriesen werden:

1. Mandantenfähigkeit (das Archiv ist und bleibt einziger Mandant des Disk-Subsystems)
2. Geschwindigkeit (wegen der höheren Ausfallwahrscheinlichkeit der Komponenten eher störend, deshalb SATA-Festplatten statt SAS).

Das neue System sollte die beiden Netzwerklaufwerke ablösen, und Digitalisate und originär-digitales Archivgut sollten über eine Netzwerkschnittstelle eingeliefert werden können. Es war aber absehbar, dass diese Repositorium-Funktionalität nicht rechtzeitig fertiggestellt werden konnte. Die bereits laufenden Digitalisierungsaktivitäten brachten zudem große Mengen an Digitalisaten, die vorläufig auf einem Netzwerklaufwerk zwischengelagert werden mussten. Wegen dieser Übergangszeit musste das anzuschaffende System das SMB-Protokoll beherrschen, eine Bedingung, die einfach erfüllt werden konnte, da eine gut funktionierende OpenSource Implementierung des SMB Protokoll (SAMBA) existiert. Diese kam entsprechend zum Einsatz. Das Umkopieren kann später stattfinden. Dabei werden die Dateien vom Netzwerklaufwerk geholt, mit Metadaten versehen, in ein SIP verpackt und so ins digitale Magazin eingeschleust. Damit ist sichergestellt, dass die Digitalisate korrekt im Archiv-Informationen-System verzeichnet sind und dass keine unverzeichneten Digitalisate auf dem Repositorium liegen.

¹⁰ Zum Begriff *Thin Provisioning* weiter unten mehr.

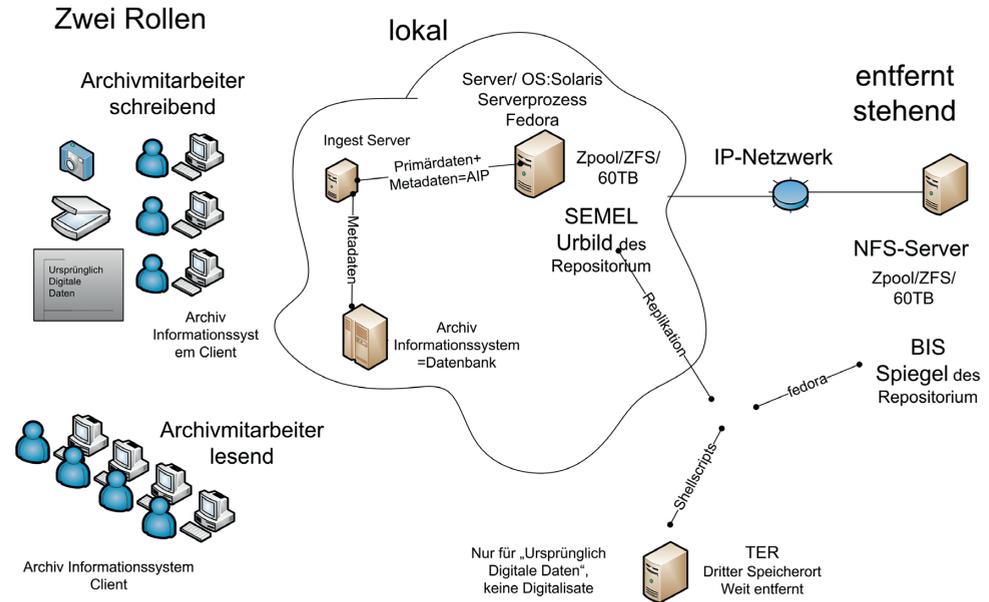


Abb. 4: Repository-Architektur des digiMag.

Eine gute *Thin Provisioning* Fähigkeit des Gesamtsystems war unter anderem wegen dieses später geplanten Umkopierens wichtig. Hierbei wird frei werdender Platz auf der einen Seite (Netzwerklaufwerk) der anderen Seite (Repository) zur Verfügung gestellt. So können beispielsweise auf einem Speicher mit 20 TB verfügbarem Platz, wovon 16 TB mit Daten belegt sind, diese 16 TB umkopiert werden, ohne dass vorübergehend ein zweiter Speicher mit mindestens der Größe 16 TB benötigt wird.

Aus Kostengründen ausgeschlossen waren von vornherein große Hardware RAID's für SAN-Netzwerke, wie zum Beispiel EMC, Hitachi oder Netapp. Diese Systeme sind sowohl in der Anschaffung wie im Betrieb viel zu teuer (mindestens um den Faktor 5), und der Vorteil der Mandantenfähigkeit bringt keinerlei Gewinn für ein Archiv. Ebenfalls überflüssig ist die hohe Geschwindigkeit dieser Systeme. Die Vorgabe des Speichervolumens auf über 10 TB schloss den Einsatz von preisgünstigen Speicherboxen aus dem SOHO Umfeld aus. 2009 hatten die größten erhältlichen Festplatten 1 TB und die größten SOHO Speicherboxen acht Einschübe, sodass damit netto eine Speicherkapazität von 6,5 Terabyte – bei Einsatz von RAID 5 – zu erzielen gewesen wäre.

Die Lösung sollte von wenigen Herstellern mit aufeinander abgestimmten Komponenten bezogen werden können. Im Idealfall ist das ein einzelner Hersteller, der alles – Hardware und Software – liefert. Gerade beim Einsatz von RAID Systemen ist das korrekte Zusammenspiel von Volume Manager, Treibersoftware und Firmware von Disk und Controller unerlässlich. Die häufigsten Bugs finden sich dort, ohne dass man als Anwender bei zwei Lieferanten einen davon haftbar machen könnte. Die Hersteller schieben im Zweifelsfall die Schuld hin und her.

Der lokale Informatikdienstleister des Kantons wurde aufgefordert, ein Angebot abzugeben, und offerierte ein großes Hardware RAID mit SAN-Anschluss. Die Kosten dafür waren allerdings für das Staatsarchiv Basel-Stadt untragbar hoch (43.000 SFr/TB). Angeschafft wurde letztendlich ein kleiner SUN sparc Rechner mit einem JBOD Disk-Subsystem. Dies brachte in Kombination mit zfs als Volume Manager und Dateisystem die maximale Erfüllung der vorgegebenen Kennwerte. Mit 24* 1 TB Festplatten wurde ein Netto-Speichervolumen von 20 Terabyte erzielt.

zfs brachte als Filesystem und Volume Manager eine Reihe von Vorteilen mit, die anderen Filesystemen fehlen, aber gerade in einer Archivlandschaft nützlich sind. Das sind:

1. Scrubbing: Die Möglichkeit, auf Betriebssystemebene alle Daten auszulesen und so die Lesbarkeit der Daten auf der Festplatte periodisch sicherzustellen.
2. Self healing: Wird beim Lesen ein Fehler detektiert, so werden die Daten der betroffenen Festplatte umgehend auf andere Platten kopiert.
3. Hashtrees: Beim Schreiben einer Datei wird ein Hashwert gebildet, der bei allen zukünftigen Lesevorgängen im Hintergrund überprüft wird. Wird ein Fehler entdeckt, so wird er falls möglich korrigiert, ansonsten wird eine Fehlermeldung an die Applikation geschickt. Beim Scrubbing wird der Hashwert ebenfalls geprüft.
4. Thin Provisioning: zfs bietet mit dem Konzept von Disk-Pools ein sehr elegantes Thin Provisioning ohne kompliziertes Umbauen eines RAID. Filesysteme können regelmäßig vergrößert und neu aufgeteilt werden und so in den freien Disk-Pool hineinwachsen.

All diese Kennwerte kennzeichnen aus unserer Sicht zfs gegenüber anderen Filesystemen und Volume Managern als überlegenes System. Als wichtigstes Merkmal seien die Hashtrees genannt, die kein anderes Filesystem anbietet. Damit kann auf einfache Art und Weise die von DIN 31644 unter 5.3.2 geforderte Integrität (K15 Unversehrtheit) der Datenströme sichergestellt werden.

Rückblick und erste Erfahrungen

Zu konstatieren ist, dass einige anfangs gesetzte Ziele des Teilprojekts 2 Digitale Archivierung nicht erreicht werden konnten. Verantwortlich hierfür sind zu einem kleineren Teil Veränderungen im Projektverlauf selbst; zum größeren Teil jedoch die Konkurrenz um die gleichen Personalressourcen durch andere Projekte. So mussten aus Termingründen Projekte wie die Einführung der Online-Bestellung auf Mitte 2011 und das Teilprojekt Digitalisierung von Archivgut zeitweise gegenüber dem hier dargestellten Projekt priorisiert werden. Daraus resultierten entsprechende Verzögerungen bei der Projektabwicklung. Es zeigte sich überdies, dass die Ausbaumaßnahmen einen größeren konzeptionellen Aufwand mit sich brachten als anfänglich geplant. Die Entwicklung verlief daher stärker iterativ als geplant, in engem Kontakt zwischen scope solutions als Auftragnehmer und dem Staatsarchiv als Auftraggeber.

Nicht umgesetzt wurde die 2010 skizzierte Anwendung *Bewerten und Anbieten*, mit der die vorarchivischen Geschäftsprozesse der Bewertung und des Anbietens abgewickelt werden sollten.

Grund für den Verzicht auf die Realisierung waren knappe finanzielle Ressourcen sowie zahlreiche offene Punkte in Bezug auf die Form der Umsetzung (Anwendung mit lokalem Client oder webbasierte Anwendung) und die Integration mit dem Archivinformationssystem. Auch wurden die Informationsangebote über die SIP-Schnittstelle und generell den Ansatz des Staatsarchivs Basel-Stadt bei der Archivierung digitaler Unterlagen aus Ressourcenmangel zurückgestellt. Im Bereich des Storage wurde die Anbindung von arcun als Teil des digitalen Magazins sowie der Umbau desselben zur Repository-Lösung nicht im Projektrahmen durchgeführt. Diese Arbeiten werden 2012 nachgeholt.

Trotz der (noch) nicht umgesetzten Ziele kann die um Projektauftrag 2008 formulierte Zielsetzung *Aufbau von Prozessen und Bereitstellung der erforderlichen Infrastruktur zur Archivierung digital erzeugter und archivierter Unterlagen der kantonalen Verwaltung* als erreicht gelten. Mit Abschluss des Projekts verfügt das Staatsarchiv Basel-Stadt über eine funktionsfähige Infrastruktur für die Archivierung digitaler Unterlagen, die 2012 produktiv gestellt wird. Zudem sind die archivischen Geschäftsprozesse von der Bewertung bis zur Magazinierung auf die digitale Archivierung vorbereitet und mit der Aufbauorganisation abgestimmt worden. Die geschaffene Infrastruktur stellt eine Erweiterung der bereits bestehenden Systeme des Archivinformationssystems und des digitalen Magazins dar. Auch bedeutet die Implementierung der KOSTbento-SIP-Schnittstelle und die Verwendung von xIsadg für die beschreibenden Metadaten des AIP eine Orientierung an Standards resp. die Nutzung von allgemein zugänglichen Spezifikationen und senkt so in beiden Fällen die Risiken von kompletten Eigenentwicklungen.

Im Bereich des digitalen Magazins konnte die dauerhafte bitstream preservation als Anforderung umgesetzt werden. Somit ist hier die notwendige Basis vorhanden, auf der Maßnahmen der digitalen Bestandserhaltung (digital preservation) aufgebaut werden können.

Aus der Projektmanagementsicht handelte es sich um ein ambitioniertes Vorhaben, in dem zahlreiche IT-Komponenten erst einmal entwickelt und anschließend zu einem Gesamtsystem integriert werden mussten. Es zeigte sich bereits zu Projektbeginn, dass es im Sinne der Minimierung von Aufwand und Risiko sinnvoll war, auf der vorhandenen Infrastruktur aufzubauen und das Archivinformationssystem scopeArchiv als führendes System beizubehalten und zu erweitern. Somit setzte das Staatsarchiv weiter auf die Nutzung einer bewährten Standardsoftware und in Teilbereichen der Entwicklung auch auf die Nutzung von offenen Standards. Zugleich resultierten daraus aber im Projektverlauf inhaltliche und terminliche Abhängigkeiten von Fahrplänen des Herstellers und archivübergreifenden Arbeiten im Rahmen des KOSTbento-Projekts.

In Bezug auf die einzelnen Komponenten wurde anfänglich die Bedeutung eines standalone-Werkzeugs zur Erzeugung schnittstellenkonformer SIP unterschätzt. Es zeigte sich im vorarchivischen Bereich seit 2009 deutlich, dass zahlreiche Aktenbildner archivwürdige digitale Unterlagen nur in Form von Dateiablagen anbieten können – sei es, weil diese in Ermangelung eines Records Management-Systems so abgelegt werden oder weil Fachapplikationen und legacy applications nur in dieser Form Unterlagen exportieren können und der Einbau einer SIP-Schnittstelle unverhältnismäßig ist. Es musste auch für diese Situation eine allgemeine Lösung gefunden werden, da die manuelle Generierung von SIP fehleranfällig und sehr aufwendig ist. Die Alternative hätte in der Öffnung des Ingestprozesses dahin gehend bestanden, dass beliebige Formen der Ablieferung hätten verarbeitet werden müssen. Dieser Ansatz führt jedoch dazu, dass das

Archiv Aufgaben der Aktenbildner übernimmt und die Anforderungen an die Übernahmewerkzeuge in Bezug auf deren Flexibilität in massiver Weise gesteigert werden, da dann jede Übernahme digitaler Unterlagen aus einem Quellsystem faktisch auf die Realisierung einer eigenen SIP-Schnittstelle hinausliefe. Dies ist sowohl mit Blick auf den zu leistenden Aufwand im Archiv als auch die Zuständigkeiten der Beteiligten im Übernahmeprozess abzulehnen. Auch führt eine Vielzahl von Übernahmeschnittstellen zu einem hohen Aufwand bei der Bewirtschaftung derselben im Verlauf des Lebenszyklus der Quellsysteme wie auch der Übernahmewerkzeuge des Archivs. Die Entwicklung eines webbasierten SIP-Generators war die Antwort hierauf, die nunmehr die Flexibilität des Archivs bei Übernahmen massiv erhöht hat.

Über alles betrachtet stehen Prozesse und Infrastruktur in einer ersten Version zur Verfügung und es gilt nun, damit weitere Erfahrungen im produktiven Einsatz zu sammeln, die zur Weiterentwicklung und Ausbauten führen werden.

Ausblick

Unter Einbeziehung der nicht realisierten Ziele und der bisherigen Erfahrungen sind in folgenden Bereichen weitere Ausbau-, respektive Entwicklungsschritte in Planung begriffen:

- Im Bereich Übernahme wird der bestehende ARAP-Ingest soweit angepasst werden, dass die gleiche AIP-Struktur generiert wird, wie im neu definierten Standard-Ingest (bento-Ingest). Sollte sich das Quellsystem ARAP in eine Richtung weiterentwickeln, die größere Anpassungen bei der SIP-Schnittstelle erforderlich macht, ist der Einsatz des SIP-Generators zu prüfen.
- Generell ist der Ingest-Workflow zu flexibilisieren. So soll der Schnittpunkt des SIP bei der Generierung der Archivpakete flexibel werden und zusätzlich zur heutigen Situation nicht nur auf Dossierstufe, sondern auch auf Bestands-, Serien- und Dokumentstufe erfolgen können. Auch muss die Einfügeposition der abgelieferten Unterlagen in die Archivtekonik zwischen den Verzeichnissstufen Bestand und Dokument frei wählbar sein.
- Schließlich ist der Ingest-Workflow so weit zu modifizieren, dass er auch für die dauerhafte Speicherung von Digitalisaten genutzt werden kann.

Der Umbau des digitalen Magazins zur Version 2.0, d.h. die Anbindung von arcun an FedoraCommons und den Wechsel des Zugriffs vom Netzwerklauferwerk zum Repository-Zugriff erfolgt im Anschluss an den Abschluss dieses Projekts. Auch wird die Nutzung von Zeitsignaturen und Checksummen zur Sicherstellung der Anforderungen Integrität und Authentizität angedacht. In betrieblicher Hinsicht wird die Minimierung des Stromverbrauchs durch den Einsatz von MAID-Systemen und damit eine Senkung der Betriebskosten sowie die Erhöhung der Ausfallsicherheit bei Systemstillständen angestrebt. Mittel- bis langfristig ist schließlich der Gesamtbetrieb des digitalen Magazins auf die zentralen Informatikdienste zu übertragen, um archivintern Ressourcen für die arbeitsintensive und nicht delegierbare Kernaufgabe der digitalen Bestandserhaltung freizusetzen.

Der Aufbau von Komponenten für die digitale Bestandserhaltung und die Schaffung eines digitalen Lesesaals, mit dem die im OAIIS vorgesehenen Zugangs- und Benutzungsfunktionen umgesetzt werden, bildet für die nächsten Jahre den Kernbereich der konzeptionellen Weiterentwicklung.