

Aufbau einer Infrastruktur für die digitale Archivierung im Staatsarchiv Basel-Stadt. Werkstattbericht

Lambert Kansy

Einleitung

Die Archivierung digitaler Unterlagen im Staatsarchiv Basel-Stadt ist neben der Digitalisierung von Archivgut und der Integration des Archivinformationssystems scopeArchiv in die kantonale Systemlandschaft Gegenstand der dritten Informatisierungsstufe des Staatsarchivs, die im Juni 2008 vom Grossen Rat des Kantons bewilligt worden ist und Ende 2010 abgeschlossen wird.¹

Das Teilprojekt zur Archivierung digitaler Unterlagen verfolgt das Ziel, Prozesse und Infrastruktur für eine geregelte Archivierung digitaler Unterlagen aus der Verwaltung des Kantons aufzubauen. Die abzudeckenden Geschäftsprozesse reichen von der prospektiven Bewertung beim Aktenbildner bis zur Bereitstellung archivierter Unterlagen für die Benutzung. Das Ergebnis der Arbeiten muss den Erhalt der Authentizität und Integrität digitalen Archivguts sicherstellen und dabei den Erhalt der Bitfolge der Digitalen Objekte auf der einen und die Sicherstellung ihrer Interpretierbarkeit durch Anwendungen sowie der menschenlesbaren Darstellung auf der anderen Seite berücksichtigen. Als Erhaltungsstrategie wird von der Migrationsstrategie ausgegangen, in der der Fokus auf einen applikationsunabhängigen Erhalt der Informationsinhalte und weniger auf äussere Eigenschaften der archivierten Objekte gelegt wird.

Im Rahmen des Projekts werden Ergebnisse aus Projekten der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST)² berücksichtigt. Zudem beteiligt sich das Staatsarchiv an einem Pilotprojekt der Firma scope solutions zur Entwicklung einer Komponente für die Übernahme und Archivierung digitaler Unterlagen mit scopeArchiv.

Grundlagen

Die Erarbeitung der Projektlandkarte, mit der die einzelnen Arbeitsfelder abgesteckt wurden, orientierte sich an dem *Reference Model for an Open Archival Information System* (OAIS; ISO-Standard 14721)³ und für die vorarchivischen Prozesse ergän-

¹ Weitere Informationen zur dritten Informatisierungsstufe unter <http://www.grosserrat.bs.ch/suche/geschaeftedetails/?idurl=08.0524> (Abruf 28.4.2010) sowie unter <http://www.staatsarchiv.bs.ch/ueber-uns/projekte/projekte-informatisierung3.htm> (Abruf 28.4.2010).

² Informationen zur KOST unter <http://kost-ceco.ch> (Abruf 28.4.2010).

³ Vgl. <http://public.ccsds.org/publications/archive/650x0b1.pdf> (Abruf 28.4.2010).

zend an dem *Producer-Archive Interface Methodology Abstract Standard* (PAIMAS; ISO-Standard 20652). Es wurden auf dieser Basis folgende Arbeitsfelder abgeleitet:

- Geschäftsprozesse,
- Schnittstellen und Datenobjekte,
- Infrastruktur,
- Informationsangebote.

Prozesse

Betroffen sind die archivischen Kernprozesse Bewerten, Übernehmen, Sichern, Erschliessen und Bereitstellen.

Der Prozess Bewertung von digitalen Unterlagen wurde definiert und gliedert sich in eine Makrobewertung, die den Aktenbildner als solchen bewertet, und eine Mikrobewertung, die Informationsbestände eines Aktenbildners bewertet. Grundsätzlich wird hierbei von einer prospektiven Bewertung ausgegangen, so dass im Rahmen des Anbietungsschritts bei der Übernahme eine Überprüfung des Bewertungsentscheids erfolgt, um sicherzustellen, dass die bei der prospektiven Bewertung getroffenen Annahmen auch der Praxis der Geschäftsverwaltung entsprechen.

Als Ergebnis dieses mehrstufig konzipierten Bewertungsablaufs resultiert neben den Bewertungsentscheiden zur Makro- und Mikrobewertung und mit letzterem verbundene Registraturpläne mit Archivwerten, Aufbewahrungsfristen und Aussonderungsregeln auch eine Ablieferungsvereinbarung, in der die Periodizität der Übernahme und die Struktur der Ablieferung festgehalten sind.

Die Modellierung der weiteren Prozesse Übernehmen, Sichern, Erschliessen und Bereitstellen verwendet die Ergebnisse des KOST-Projekts AUGev.⁴

Der dort formulierte Übernahmeprozess gliedert sich in mehrere, auch wiederholbare Schritte: Anbieten (Metadaten und teilweise auch Primärdaten gemäss Ablieferungsvereinbarung), Prüfen des Angebots, Übernahmeentscheid, Erstellung und Übergabe der Ablieferung (*Submission Information Package* (SIP)), Prüfen derselben und anschließende Verarbeitung, d. h. Erstellung von *Archival Information Packages* (AIPs) und der *Descriptive Information* (DI) mit anschließender Einlagerung im Digitalen Magazin respektive Übergabe der beschreibenden Metadaten an das Archivinformationssystem. Dabei wird der Übernahmeprozess eng mit dem der Sicherung verbunden.

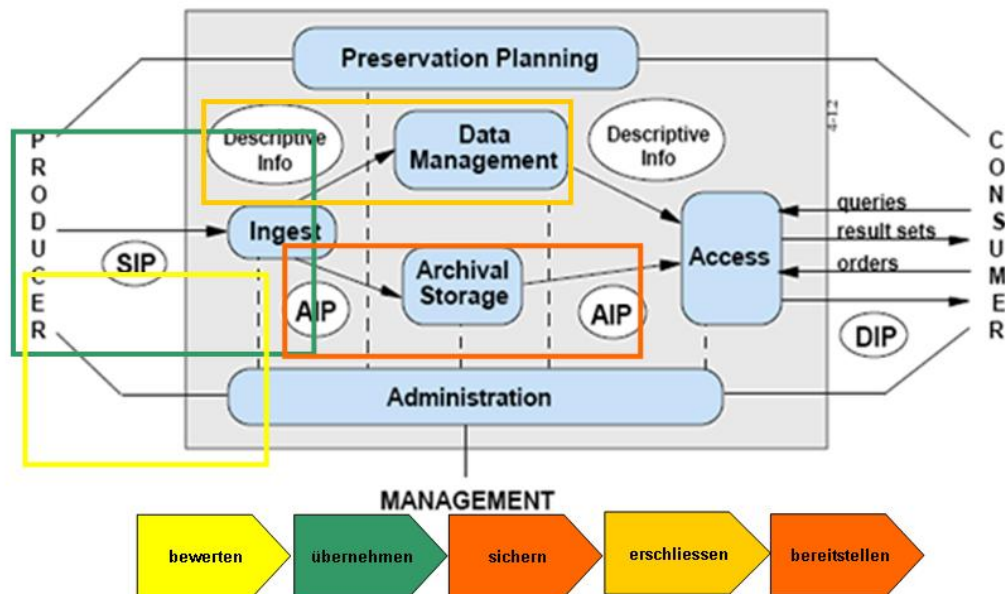
Die klassische Erschliessung im Sinne des Ordnen und Verzeichnens entfällt weitgehend. Neu wird Erschliessung im Wesentlichen die Anreicherung, Kontextualisie-

⁴ Informationen zu diesem Projekt sind zu finden unter: http://kost-ceco.ch/cms/index.php?augev_public_de (Abruf 28.4.2010).

zung und Vernetzung mit anderen Archivinhalten der ins Archivinformationssystem übernommenen *Descriptive Information* bedeuten.

Die Bereitstellung schliesslich umfasst die Anforderung eines *Archival Information Packages* aus dem Archivinformationssystem und dessen Ausgabe aus dem Digitalen Magazin an den Benutzer, der ein Archivmitarbeiter sein wird, solange die (Online-)Benutzung im eigentlichen Sinne des OAIS-Modells nicht realisiert ist.

Die folgende Grafik beschreibt die Abbildung der betrachteten Geschäftsprozesse auf die Funktionsbereiche des OAIS-Modells.



Grafik: 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

Betrachtete Geschäftsprozesse und OAIS-Modell

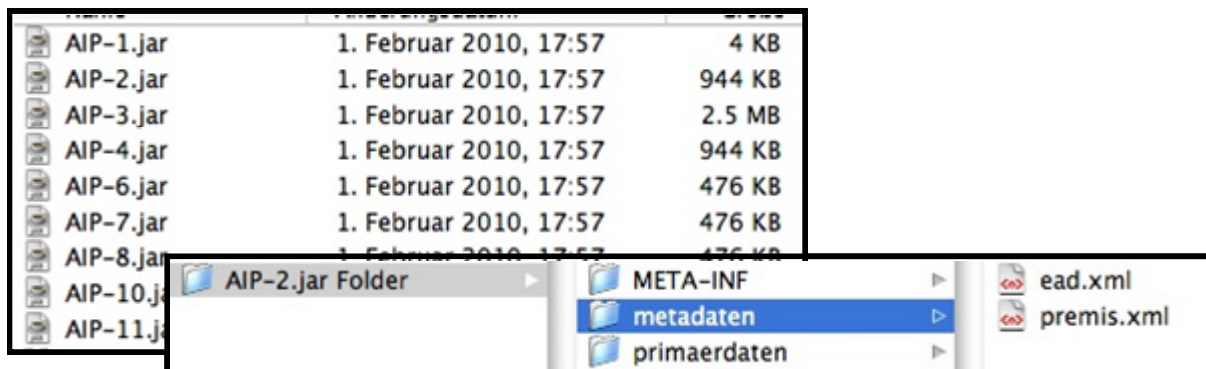
Das Projekt befasst sich explizit nicht mit den Funktionsbereichen *Administration*, *Preservation Planning* und *Access*. Daher wird auch das Benutzungsobjekt, das *DIP*, nicht behandelt. Anstelle der Benutzung wird mit der Bereitstellung ein Abschluss gesetzt. Streng genommen handelt es sich um einen Teil des *Archival Storage*, da lediglich das *AIP* aus dem Digitalen Magazin geholt wird, nicht aber ein *DIP* daraus gebildet wird. Auf diese Weise ist jedoch dennoch die Gewährleistung einer einfachen manuell durchzuführenden Benutzung digitalen Archivguts möglich.

Schnittstellen

Das zweite Handlungsfeld befasst sich mit den Objekten der Prozesse: dem Übernahmeobjekt, dem Archivobjekt und den Verzeichnungsinformationen.

Das Übernahmeobjekt (*Submission Information Package* (SIP)) umfasst eine Ablieferung und wird in Form einer Containerdatei im ZIP-Format geliefert. Diese enthält zwei Verzeichnisse für Metadaten und Primärdaten. Letztere kann weitere Verzeichnisstrukturen beinhalten.

Das Metadatenverzeichnis enthält eine Datei mit administrativen (Provenienz und Angaben zur Ablieferung), technischen und deskriptiven Metadaten zu den Primärdaten.



Archival Information Package-Struktur des Pakets

Aus dem SIP werden während des *Ingest* die Archivobjekte (*Archival Information Packages* (AIPs)) generiert. Dabei wird aus einem SIP eine Vielzahl von AIP – in der Regel auf Stufe Dossier – generiert. Doch muss die Generierung auch andere Granularitätsrelationen zwischen SIP und AIP zulassen.

Auch das Archivobjekt wird als Containerdatei realisiert. Dabei handelt es sich um eine JAR-Datei, die – analog zum SIP – die beiden Verzeichnisse „Metadaten“ und „Primärdaten“ enthält. Während die Primärdaten aus dem SIP lediglich kopiert werden, wird die Metadatenfile des SIP umgeformt und ergänzt. So werden technische Metadaten aus den Primärdaten mittels JHOVE und DROID extrahiert und zusammen mit Angaben über den Ingestprozess (Datum und Name des durchführenden Benutzers) in einer XML-Datei gemäss PREMIS gespeichert. Diese enthält folgende technische Metadaten für alle Primärdateien:

- Dateiformat,
- Dateigrösse in Byte,
- MD5-Checksumme,
- URL auf den Speicherort der Primärdatei.

Die deskriptiven Metadaten aus dem SIP werden in das AIP übernommen und ergänzt um die Archivsignatur und die Einfügeposition des AIP in der Archivtekonik unter Angabe des Pfades von der Einfügeposition bis zur Abbildung der Provenienzstelle in der Archivtekonik – in der Regel auf Stufe der Fonds. Diese Angaben werden in Form einer XML-Datei gemäss EAD gespeichert.

Der Zugriff auf Primärdaten erfolgt aus der EAD-Datei mit den beschreibenden Metadaten via Referenzierung mit xpointer auf die entsprechenden URL-Angaben in der PREMIS-Datei.

Auf diese Weise entsteht ein AIP, das relativ weitgehend selbstbeschreibend ist und durch Angabe eines eindeutigen Identifikators (*Persistent Identifier*) referenziert werden kann.

Aus den im SIP gelieferten und im Zuge des *Ingest* ergänzten beschreibenden Metadaten wird die Verzeichnungsinformation (*Descriptive Information (DI)*), gebildet, die in das Archivinformationssystem scopeArchiv übertragen wird und folgend Elemente umfasst:

- Titel, Signatur, Zeitraum, Umfang, Archivalienart, Verzeichnungsstufe und Provenienz,
- Angaben betreffend Geheimhaltung, Datenschutz und Öffentlichkeitsstatus,
- Verortung in der Archivtektonik unter Angabe des Pfades bis zur Provenienzstelle auf Stufe der Fonds,
- Angabe des *Persistent Identifiers* für den Zugriff auf das AIP im Digitalen Magazin.

Die SIP- und AIP-Definition – v. a. die Struktur der Metadaten – erfolgt abschliessend erst unter Berücksichtigung der Ergebnisse des KOST-Projekts bento.⁵

Infrastruktur

Die Implementierung der Prozesse und Schnittstellen geht davon aus, dass das Archivinformationssystem scopeArchiv weiterhin das führende System im archivischen Workflow bleibt. Daher wurde im Rahmen des oben erwähnten Pilotprojekts mit scope solutions eine Komponente zur Durchführung eines OAIS-konformen Ingestprozesses entwickelt.

Die folgende Gegenüberstellung stellt dar, mit welchen Infrastrukturkomponenten die einzelnen Prozesse unterstützt, respektive abgewickelt werden.

Prozess	Komponente
Bewertung (Appraisal)	„Bewerten & Anbieten“
Übernahme (Ingest)	<ul style="list-style-type: none"> ➤ „Bewerten & Anbieten“ ➤ FTP-Server „Anlieferung“ ➤ scopeIngest
Sicherung (Archival Storage)	Digitales Magazin <ul style="list-style-type: none"> ➤ Fedora, ➤ Primärer Speicher, ➤ arcun
Erschliessung (Data Management)	scopeArchiv
Bereitstellung (Archival Storage)	Digitales Magazin

Prozesse und Infrastrukturkomponenten

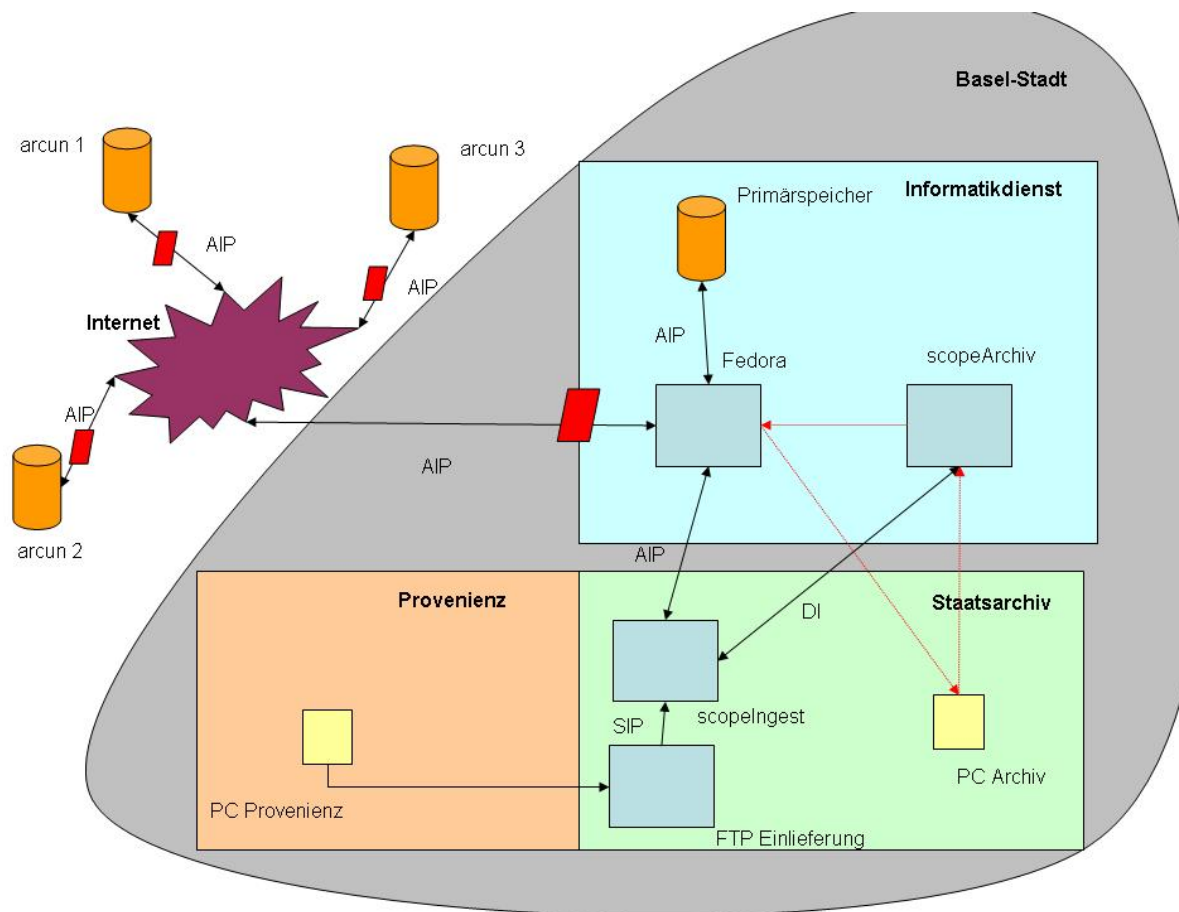
⁵ Informationen zu diesem Projekt sind zu finden unter http://kost-ceco.ch/cms/index.php?bento_de (Abruf 28.4.2010).

Die generierten AIP werden in das Digitale Magazin eingelagert. Dieses liefert einen *Persistent Identifier* zurück, der zusammen mit der deskriptiven Information in das Archivinformationssystem übernommen wird.

Die bestehende Infrastruktur des Digitalen Magazins wird erweitert um das Fedora Repository, einem open source digital Repository.⁶ Fedora übernimmt die generierten AIP, lagert diese im Speichersystem ein und generiert einen *Persistent Identifier*.

Der lokale und aus zwei NAS-Systemen bestehende Speicher wird erweitert um die Anbindung des entfernten Speichersystems arcun, das aus einer dreifach redundanten Speicherung an drei Standorten in der Schweiz besteht und zurzeit von mehreren Archiven im Rahmen der KOST aufgebaut wird.⁷

Neben diesen Komponenten wird in Gestalt eines FTP-Servers eine digitale Anlieferung geschaffen, mit der Ablieferungen innerhalb des Verwaltungsnetzes abgewickelt werden können.



Schematische Sicht auf die Infrastruktur

Der Geschäftsprozess „Bewerten“ soll durch die Entwicklung der Anwendung „Bewerten & Anbieten“ erstmals im Staatsarchiv informatikgestützt abgewickelt werden. Die notwendigen Spezifikationen sind im Gange.

⁶ Informationen zu Fedora Repository siehe unter <http://www.fedora-commons.org/> (Abruf 28.4.2010).

⁷ Vgl. http://kost-ceco.ch/cms/index.php?project_arcun_de (Abruf 28.4.2010).

Die konzipierte und pilotmässig im Rahmen des Projekts realisierte Komponente *scopeIngest* beruht auf einem Applikationsserver, der über das Modul *Ingest* in *scopeArchiv* angesprochen und gesteuert wird. Sie bildet folgenden Workflow ab:

- Einlesen des SIP,
- technische Validierung gegenüber der SIP-Definition,
- manuelle Validierung durch stichprobenhafte Prüfung der Metadaten und der Primärdaten auf inhaltliche Korrektheit,
- Generierung der AIP und der DI gemäss der Definition,
- Abspeicherung der AIP im Repository und Ergänzung der DI um den *Persistent Identifier*,
- Übergabe DI an *scopeArchiv*.

Informationsangebote

Durch die Neugestaltung der Prozesse, der Definition von Schnittstellen und den Aufbau der Infrastruktur schafft das Archiv die Voraussetzungen für eine operative Umsetzung der digitalen Archivierung – aber nur seinerseits.

Oft vergessen, im archivischen Alltag jedoch unverzichtbar, sind die Informationsangebote, die die Partner des Archivs (Aktenbildner und deren Informatikprovider) überhaupt erst in die Lage versetzen, dessen Anforderungen zu entsprechen. Ohne entsprechende Bereitstellung von Informationen über Abläufe und technisch-organisatorische Voraussetzungen ist eine erfolgreiche Archivierung digitaler Unterlagen kaum zu erwarten.

Daher umfasst dieses Arbeitsfeld des Projekts die Erarbeitung klarer Vorgaben und Hilfestellungen für die aktenbildnerseitigen Aktivitäten in der Bewertung und der Übernahme.

Hierzu sollen im Rahmen des Projekts folgende Ergebnisse erarbeitet werden:

- Grundsätze zur Archivierung digitaler Unterlagen im StABS (Policy),
- Erläuterungen zu Abläufen und Rollen; inkl. Muster für eine Ablieferungsvereinbarung,
- Publikation und Erläuterung der Übernahmeschnittstelle SIP mit Darstellung der Struktur, der erforderlichen Metadaten und einem Katalog akzeptierter Dateiformate.

Projektverlauf und Ausblick

Die Geschäftsprozesse sind weitgehend definiert und bilden die Grundlage für die Arbeiten im Bereich Infrastruktur, in dem – Stand März 2010 – folgende Ergebnisse vorliegen:

- FTP-Server „Anlieferung“ eingerichtet,
- Ingesttool scopeArchiv entwickelt,
- Repository Fedora im Test,
- Einbindung arcun im Test,
- Ausbau Primärer Speicher in Beschaffung.

Eine abschliessende Definition der Schnittstellen SIP und AIP wird erst nach Abschluss des KOST-Projekts bento Mitte 2010 erfolgen, das die Erarbeitung dieser beiden Schnittstellen leistet.

Das sich im Aufbau befindende Gesamtsystem wird 2010 im Rahmen einer Pilotübernahme aus dem elektronischen Polizeijournal der Kantonspolizei erstmals getestet. Hierzu erfolgte vorgängig die Definition von spezifisch auf diese Übernahme zugeschnittenen SIP und AIP. Im Februar 2010 wurden erstmals die Prozesse Übernahme, Erschliessung und Bereitstellung getestet – mit vielversprechendem Erfolg, aber noch grossem Verbesserungspotential.

Bevor das Projekt Ende 2010 abgeschlossen wird, sind noch folgende Arbeiten zu erledigen:

- Entscheidung, ob „Bewerten und Anbieten“ realisiert wird und gegebenenfalls Realisierung,
- generalisierte Schnittstellen SIP und AIP festlegen und in scopelngest implementieren,
- produktive Einrichtung scopelngest und Digitales Magazin (Fedora, Primärer Speicher und arcun),
- Integration aller vorhandenen Komponenten,
- Aufbau der Informationsangebote und interne Schulung.