

Pilotprojekt KOSTPROBE Arbeitsgruppe Aufbewahrung

Zusammenstellung provisorischer Lösungen für die Aufbewahrung archivierter elektronischer Unterlagen

Dieses Dokument dient dazu, Möglichkeiten zur Aufbewahrung digitaler Unterlagen zu beschreiben, die in verschiedenen Mitgliedsarchiven der KOST bereits in Gebrauch sind. Diese verstehen sich sämtlich als provisorische Lösungen. Dabei werden zugunsten einer kurzfristigen Realisierbarkeit Abweichungen von gängigen Vorgaben für Archivierungssysteme bewusst in Kauf genommen. Es wird erwartet, dass alle hier beschriebenen Lösungen mittelfristig durch ein Archivierungssystem ersetzt werden.

Die Informationen in diesem Dokument wurden im Rahmen des Pilotprojekts KOSTPROBE zusammengestellt, zur Hauptsache in der Diskussion der Arbeitsgruppe Aufbewahrung.

Mitarbeit: Hedi Bruggisser (StATG), Christian Eugster (KOST), Olivier Debenath (StAZG), Markus Lischer (StALU), Claudia Schmucki (StAZH), Peter Witschi (StAAR)

Redaktion: Georg Büchler (KOST)

Disposition

1	Konzept Fileserver	1
1.1	Staatsarchiv Appenzell-Ausserrhoden	1
1.2	Staatsarchiv Zürich.....	2
1.3	Beurteilung.....	2
1.4	Vorgehensskizze zur Implementierung.....	3
2	Konzept Speicherung in der Archivdatenbank.....	3
2.1	Staatsarchiv Zug	3
2.2	Beurteilung.....	4
2.3	Vorgehensskizze zur Implementierung.....	4
3	Allgemeine Vorbereitungsarbeiten bei Implementierung.....	4

1 Konzept Fileserver

1.1 Staatsarchiv Appenzell-Ausserrhoden

Im Rahmen des Projekts ELAR hat das StAAR für die Primärdatenhaltung einen dedizierten File-Server eingerichtet. Zur Zeit sind Regierungsratsprotokolle, GIS-Querschnitte, digitalisierte Familienregister und die Datenbank zur Bauernhausforschung darauf abgelegt. Diese belegen 40 GB von insgesamt 311 GB Speicherplatz auf dem Server. Sie sind im Archivsystem scopeArchiv verzeichnet.

Die Dateistruktur und Dateibenennung folgen einem einheitlichen Schema. Kernelement ist die Identität von Dateiname und Archivsignatur. Das erste Element, eine Kombination dreier Buchstaben, gibt dabei die Abteilung an. So liess sich eine flache Struktur mit maximal vier Ebenen realisieren. Berechtigte Personen können die auf dem Server abgelegten Primärdaten direkt aus dem Archivsystem scopeArchiv aufrufen. Die Benutzerversionen sind in scopeArchiv beim jeweiligen Dossier unter „Dateien“ verlinkt. Der File-Server ist vor direktem Zugriff geschützt; nur zwei Personen können ihn ohne den „Umweg“ über scopeArchiv ansprechen.

1.2 Staatsarchiv Zürich

Das StAZH verzeichnet bereits übernommene digitale Objekte in seiner Archivdatenbank DACHS. Die Verzeichnung soll einem Benutzer den Entscheid über die Relevanz der Primärdaten für seine Fragestellung ermöglichen. Der Verzeichnungseintrag ist identifiziert durch die Archivsignatur mit dem Anhängsel „EL“.

Die Primärdaten werden inklusive Dokumentation und Metadaten auf einem Fileserver abgelegt. Ein Ordner (bezeichnet mit der Bestandessignatur und dem Bestandesnamen) enthält die AIPs (Archival Information Packages gemäss OAIS-Terminologie). Diese sind identifizierbar über die Archivsignatur. Die Bildung der Ordnernamen ist also bis auf Stufe AIP vorgegeben. Die AIPs sind weiter unterteilt in Datenobjekte, Dokumentation und Struktur/Layout. Unter Datenobjekte sind die Primärdaten (bis anhin in den Formaten Text, PDF, TIFF und XML), unter Dokumentation die Metadaten (ein XML-File gemäss der alten Version des PREMIS-Standards) und die nicht strukturierte Dokumentation (PDF-Files, evtl. XML-Files) abgelegt. Unter Struktur/Layout befinden sich allfällige DTD (Document Type Definitions) bzw. XML-Schemas für die Primärdaten, sowie CSS (Cascading Style Sheets) zur Ansicht von Primärdaten und Metadaten.

Einige Archivare verfügen über die Berechtigung Ändern, der Rest der Mitarbeitenden hat Leserecht auf dem Fileserver.

Grundsätzlich sind alle bisher archivierten Daten nicht zur Benutzung freigegeben. Diese müsste durch Zugriff auf den Fileserver geschehen; eine Verlinkung zwischen dem Erschliessungssystem (DACHS) und den AIPs existiert nicht, da in DACHS keine Links möglich sind. Die bisher verwendeten Formate sind weit verbreitet und stellen für die Benutzung auch für ungeübte Anwender keine grundsätzlichen Probleme dar.

Anzumerken ist schliesslich, dass Daten in einer Anfangsphase teilweise redundant gehalten werden (Papier, CD, elektronisches Archiv).

1.3 Beurteilung

1.3.1 Stärken

- Der Zugriff auf die Primärdaten durch Verlinkung aus dem Findmittel heraus ist sehr bedienungsfreundlich.

- Die Verwendung der Verzeichnungsstruktur des Archivs als Ordnerstruktur des Filesystems trägt durch gegenseitige Referenzierung zur Sicherheit bei.
- Die redundante Datenhaltung kommt der Forderung nach Speicherung auf mehreren Datenträgern entgegen.

1.3.2 Schwächen:

- Der Dateiname steht nicht in direktem Zusammenhang mit der Datei, sondern nur via Dateisystem, ist also mithin kein der Datei inhärentes Merkmal. Daher birgt er als Identifikationsmerkmal Risiken. Das StAAR hat diese erkannt und arbeitet an einer besseren Identifikation der archivierten Dateien. Mit der Version 4.2 von scopeArchiv werden zudem Änderungen im Ablagepfad oder im Namen einer Datei durch FileWatcherService erkannt und die Verknüpfungen automatisch nachgeführt.
- Wenngleich eingeschränkt auf sehr wenige Personen, birgt doch der mögliche direkte Zugriff auf den File-Server das Risiko von Inkonsistenzen zwischen Primär- und Verzeichnungsdaten. Wenn Pfad oder Dateiname geändert werden, müssen diese Veränderungen in allen Links im Findmittel nachgetragen werden. Andernfalls resultieren broken links.

1.4 Vorgehensskizze zur Implementierung

- Hardware einrichten (Server, Harddisk oder Partition)
- Zugriffsbeschränkungen auf die Hardware definieren. Dabei ist zu bedenken, dass im Normalfall neben den ArchivarInnen auch Mitarbeitende und Software (Backup- und Restoredienste etc.) der Informatikzentren Zugriff haben.
- Ablagestruktur bestimmen (dabei auch provisorische und temporäre Ordner vorsehen für Testdaten sowie für Daten in Bearbeitung)
- Dateinamenkonzept erstellen
- Primärdaten auf der Hardware gemäss Struktur ablegen
- Primärdaten im Findmittel verzeichnen, allenfalls durch Import von ausgewählten Primär- bzw. Metadaten
- Link zu den Primärdaten im Findmittel eintragen
- Falls vorgesehen: Hash-Werte ermitteln und archivieren

2 Konzept Speicherung in der Archivdatenbank

2.1 Staatsarchiv Zug

Im StAZG werden archivierte Unterlagen provisorisch direkt in der Archivdatenbank (STAR) abgelegt. Am Beispiel des Landwirtschaftlichen Informationssystems LAWIS lässt sich aufzeigen, wie dabei vorgegangen wird: Die LAWIS-Betriebsdaten bilden einen Bestand mit jährlichen Teilbeständen. Auf Dokumentenstufe werden die einzelnen Unterlagen verzeichnet. Dabei handelt es sich um die Betriebsdaten eines Jahres in der Gesamtübersicht (als csv-File) und um die digitalen Aequivalente der Betriebsblätter. Zum entsprechenden Eintrag in

der Archivdatenbank kann ein Link auf ein digitales Objekt abgelegt werden. Mit dem digitalen Objekt können verschiedene Versionen (sprich Migrationen) der Primärdaten verbunden werden. Diese werden vorläufig und provisorisch direkt in der Archivdatenbank abgelegt.

2.2 Beurteilung

2.2.1 Stärken

- Der Zugriff auf die Primärdaten durch Verlinkung aus dem Findmittel heraus ist sehr bedienungsfreundlich.
- Dadurch, dass es keinen direkten Schreibzugriff auf die Primärdaten gibt (ausser für Datenbankadministratoren), wird die Sicherheit erhöht – gerade im Vergleich mit der Variante Fileserver (AR, ZH).
- Bei jedem Update der Archivdatenbank werden die Primärdaten mitgenommen.

2.2.2 Schwächen

- Die Speicherung der Primärdaten im Findmittel führt dazu, dass diese in einem proprietären Umfeld gehalten werden. Damit entsteht eine zusätzliche Abhängigkeit vom Hersteller des Findmittels und mithin ein zusätzliches Risiko für die Zugänglichkeit der Primärdaten.

2.3 Vorgehensskizze zur Implementierung

STAR/3 bietet die Möglichkeit, Dateien direkt in das Findmittel zu importieren.

3 Allgemeine Vorbereitungsarbeiten bei Implementierung

Abgeklärt werden müssen ferner einige allgemeine Punkte. Die folgende Übersicht ist nicht abschliessend:

- Die zu erwartenden Mengen und Grössenordnungen müssen mindestens grob geschätzt werden.
- Authentizitätsanforderungen müssen geprüft und gegebenenfalls umgesetzt werden: redundante Speicherung auf anderem, physisch getrenntem Datenträger; Massnahmen zur Datensicherung; etc. Siehe dazu das entsprechende Papier der KOST.
- Dateiformatfragen sind zu klären: Zu verwendende Formate definieren, Verwaltung dieser Formate gewährleisten.