

## **Projekt „Aufbau eines Digitalen Archivs“ im Bundesarchiv Werkstattbericht über den „individuellen“ Umgang mit Massendaten**

KARSTEN HUTH, KATHRIN SCHROEDER

### ***1. Ausgangssituation und Herausforderungen***

Die Archivierung digitaler Überlieferungen aus Behörden ist ein wichtiger Schwerpunkt der Abteilung B (Bundesrepublik Deutschland) des Bundesarchivs. Die ältesten Daten stammen aus den 70er Jahren<sup>1</sup> und können heute noch interpretiert werden. Hingegen fehlen für die Übernahme, Speicherung, Erschließung und dauerhafte Erhaltung elektronischer Bestände aus Behörden praktische Erfahrungen. Der aktuelle Bedarf für ein „Digitales Archiv“ ergibt sich aus der voranschreitenden Umstellung der Bundesbehörden von einer papiergebundenen Aktenführung hin zum e-Government. Bereits im Februar 2006 sind Aussonderungen an das Bundesarchiv erfolgt, weitere wurden angekündigt.

Die Herausforderungen bestehen in der Neuartigkeit der Aufgabe und der Einführung eines mandantenfähigen, flexiblen Systems, das funktional den Anforderungen des Bundesarchivs langfristig gerecht wird und wirtschaftlich zugleich ist. Die Notwendigkeit zur Flexibilität des Archivs ergibt sich schon aus dem Umstand, dass Metadaten aus Vorgangsbearbeitungssystemen der Bundesbehörden in unterschiedlichen Formaten exportiert werden. Da diese Metadaten die Struktur und den Kontext der elektronischen Akten enthalten und somit auch zu primärem Archivgut werden, ist ihre Speicherung in einem einheitlichen Archivformat notwendig. Das Bundesarchiv hat sich deshalb entschlossen, ein Standard Archivierungsmodul (SAM) für die Übernahme von elektronischen Aussonderungen zu entwickeln. Eine Hauptaufgabe von SAM wird die Umsetzung der heterogenen Metadaten aus den Behördensystemen in ein einheitliches Archivformat sein. Die „Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung“ (KBSt) fördert das Bundesarchiv bei der Entwicklung und beim Ausbau von SAM als EfA-(Einer-für-Alle)-Dienstleistung<sup>2</sup>.

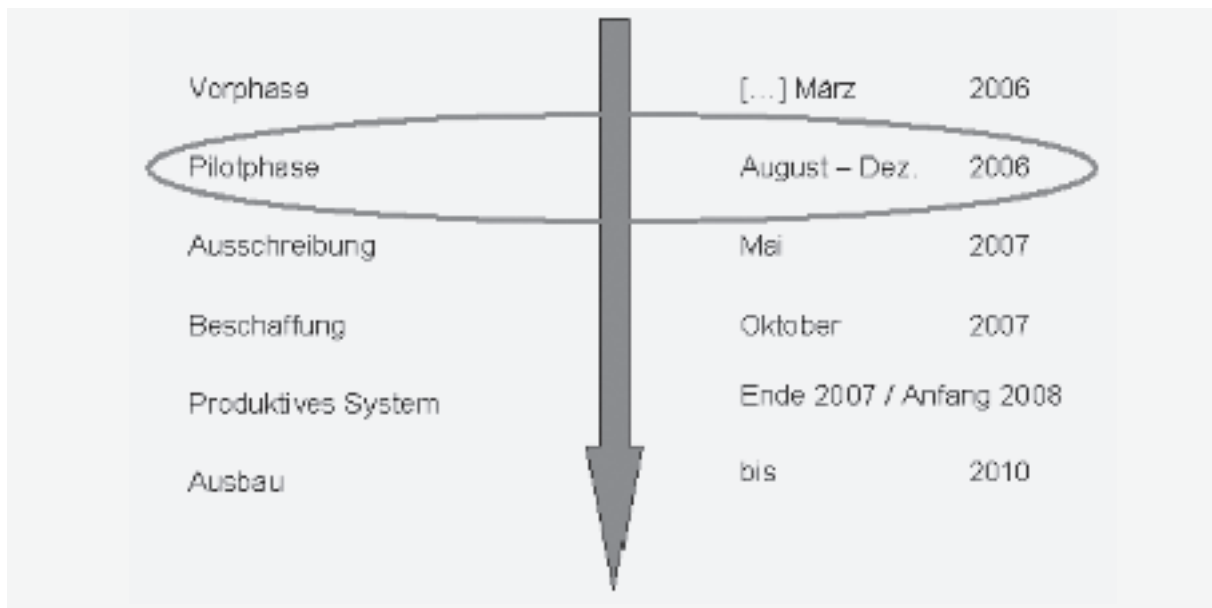
Ein inkrementelles Vorgehen im Projekt wurde gewählt, um die Funktionalitäten des „Digitalen Archivs“ dem aktuellen Bedarf anpassen zu können, soweit dies wirtschaftlich ist. Zunächst wurde in einer Pilotphase ein „End-to-End-Prozess“ erfolgreich realisiert. Während der Pilotphase wurden Erkenntnisse gewonnen, die in einen detaillierten Anforderungskatalog für das zukünftige, produktive System eingeflossen sind. Im Anschluss an die Einführung des Systems werden Erweiterungen in mehreren Ausbaustufen folgen.

---

<sup>1</sup> Siehe unter [http://www.bundesarchiv.de/aufgaben\\_organisation/abteilungen/bundesrepublik/00203/index.html](http://www.bundesarchiv.de/aufgaben_organisation/abteilungen/bundesrepublik/00203/index.html) (letzter Zugriff 24.4.2007).

<sup>2</sup> Siehe unter <http://www.bundesarchiv.de/service/behoerdenberatung/01435/index.html> (letzter Zugriff 3.4.2007).

## – Speicherung –



Unser Beitrag beschreibt den Workflow während der Testphase. Dieser umfasst das Anbieten und Bewerten, die Übernahme, Speicherung und Recherche von elektronischen Akten.

## 2. Darstellung des End-to-End-Prozesses

### 2.1 Der organisatorische Vorlauf

Bevor die erste Anbietung elektronischer Akten erfolgen kann, treten die Behörde und das Archiv in Kontakt. Dabei werden die technischen Einzelheiten der Datentransfers geklärt. Folgende Informationen müssen an das Archiv übergeben werden:

- Die Datenschemata der Metadaten aus dem Vorgangsbearbeitungssystem.
- Eine kurze technische Beschreibung der verwendeten Hardware/Software Struktur des Vorgangsbearbeitungssystems mittels eines Formulars des Bundesarchivs.

Darüber hinaus werden in Absprachen zwischen Behörde und Archiv die zulässigen Dateiformate der Primärdokumente festgelegt. Präferenzformate für die Archivierung elektronischer Akten sind zur Zeit TIFF 6.0 und PDF/A.

Die Behörde erhält anschließend eine eindeutige Behördenidentifikationsnummer und eine Profildnummer. Diese IDs dienen der eindeutigen Identifikation des Absenders und des im System hinterlegten technischen Profils. Es wurde berücksichtigt, dass eine Behörde mehrere Profile haben kann. Während der Pilotphase wurde die Behördenrolle vom Bundesarchiv übernommen.

### 2.2 Der Geschäftsprozess „Anbieten und Bewerten“

Während der Testphase wurde das vierstufige Verfahren des DOMEA-Konzeptes<sup>3</sup> umgesetzt. Dabei erstellt die Behörde eine Anbieterliste über die Unterlagen, deren Aufbewahrungsfrist abgelaufen ist. Diese Anbieterliste beinhaltet die Metadaten aus dem Vorgangsbearbeitungssystem der Behörde und kann im XML- oder Excel- (CSV) Format an das Bundesarchiv übermittelt werden. In der Pilotphase wurde dies per E-Mail-Anhang realisiert. Die Behörde sandte die Daten an eine nur für diesen Zweck

<sup>3</sup> Siehe unter DOMEA-Erweiterungsmodul „Aussonderung und Archivierung Elektronischer Akten“ im Internet: [http://www.kbst.bund.de/cln\\_047/nn\\_836802/Content/Standards/Domea\\_\\_Konzept/domea\\_\\_node.html\\_\\_nnn=true](http://www.kbst.bund.de/cln_047/nn_836802/Content/Standards/Domea__Konzept/domea__node.html__nnn=true) (letzter Zugriff 3.4.2007).

## – Speicherung –

bestimmte E-Mail-Adresse. Bei einer erfolgreichen Validierung der Legitimationsdaten (Behördenidentifikationsnummer/Profilnummer) initiierte das System anschließend den Geschäftsprozess.

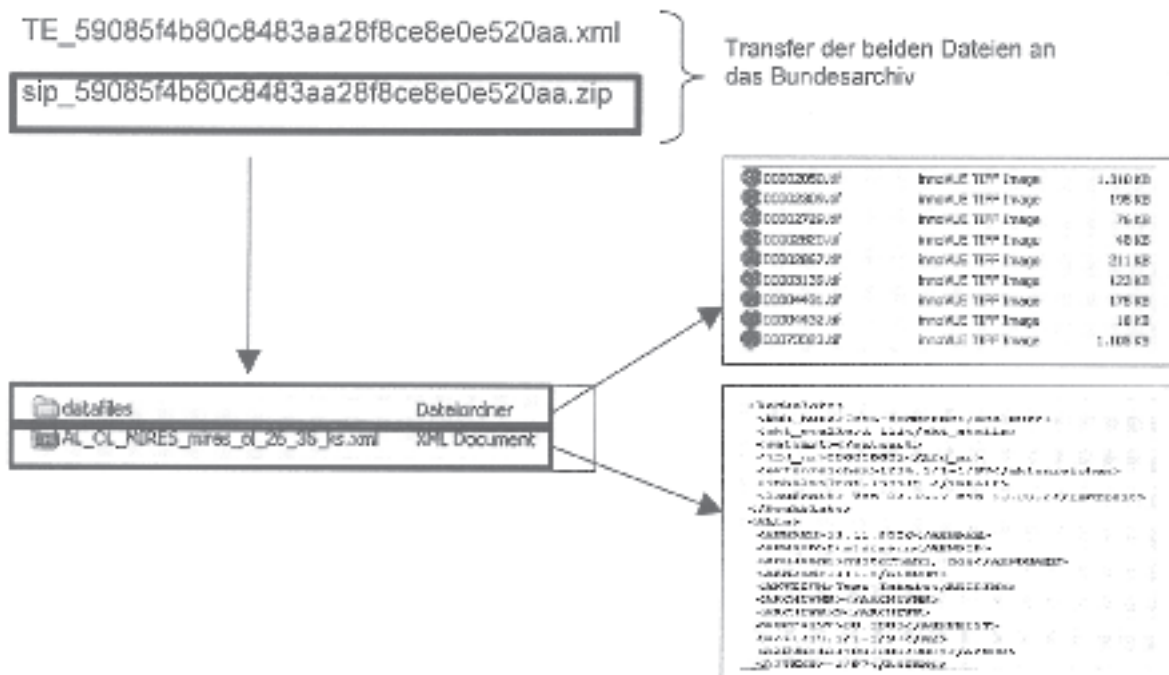
Zunächst wurden die eingehenden Metadaten gegen das im System hinterlegte Datenschema geprüft (s. 2.1). Das System generierte eine eindeutige Transaktionsnummer für den gesamten Archivierungsprozess und übermittelte sie per Mail an die Behörde. Die Transaktionsnummer wird für Prozesse, die in mehreren interaktiven Schritten ablaufen, benötigt. Nach erfolgreicher Validierung wurden die Daten in das vom Bundesarchiv definierte XML-basierte Datenformat „XBARCH“ umgesetzt.

Anschließend wurden die an der Testphase beteiligten Kolleginnen und Kollegen per E-Mail über den Eingang einer Anbieterliste informiert. Sie konnten sich beim System anmelden und auf die Anbieterliste zugreifen. Das System war vollständig web-basiert realisiert. Die archivfachliche Auswertung und Bearbeitung der Anbieterliste erfolgte über eine Eingabemaske, die mittels eines Standard-Web-Browsers aufgerufen wurde. Ziel der Bearbeitung war die Festlegung der Aussonderungsart, nachdem die Akten- und Vorgangsbetreffe auf ihre Archivwürdigkeit hin geprüft wurden. Als Parameter konnten „A“ (Archivieren) und „V“ (Vernichten) eingetragen werden. Nach Abschluss der Bewertung wurde das Bewertungsverzeichnis in das Metadatenformat der Behörde umgesetzt und per E-Mail der Behörde zugesandt.

### 2.3 Der Geschäftsprozess „Übernahme der Aussonderungsdatei“

Die Behörde erstellte anhand des Bewertungsverzeichnisses eine Aussonderungsdatei. Diese beinhaltete Metadaten und Primärdaten der elektronischen Akten, welche vom Archiv mit „A“ bewertet wurden. Vor dem Transfer der elektronischen Akten stand eine erneute Kontaktaufnahme zwischen der Behörde und dem Archiv an. Beide einigten sich auf einen Übertragungsweg und einen Zeitraum, in dem die Aussonderungsdatei an das Archiv transferiert werden musste. Der Datentransfer von der Behörde zum Archiv konnte Online über ein sicheres Übertragungsprotokoll oder Offline mittels Datenträger erfolgen. Die Aussonderungsdatei wurde in ein spezielles Verzeichnis übertragen, das vom System ständig auf neu eingehende Daten hin geprüft wurde. Die Aussonderungsdatei bestand aus einem ZIP-Container, in dem die Primärdaten und das Aussonderungsverzeichnis (mit den Metadaten des VBS) verpackt wurden. Zusätzlich zur Aussonderungsdatei war eine XML-Datei erforderlich, mit der sich die Behörde authentifiziert hat.

Beispiel für eine Aussonderungsdatei



## – Speicherung –

### 2.4 Der Geschäftsprozess

#### „Ingest, Qualitätskontrolle und Übergabe an die Archivierungslösung“

Nach einem Eingang der Aussonderungsdatei wurden zunächst die Authentifizierungsdaten validiert, anschließend der ZIP-Container entpackt und die Daten auf Vollständigkeit geprüft. Das System kontrollierte, ob zu jedem Dokument, das im Aussonderungsverzeichnis aufgeführt ist, auch ein entsprechendes Primärdokument mitgeliefert wurde.

Die Metadaten wurden in das XML-basierte Datenformat des Bundesarchivs (XBARCH) umgesetzt. Anschließend generierte das System die Archivsignaturen automatisch über einen implementierten Algorithmus. Die eingebundene Open-Source-Lösung JHOVE<sup>4</sup> ermittelte alle für die Archivierung relevanten technischen Metadaten (z. B. über das Dateiformat), die später den XBARCH-Metadaten hinzugefügt wurden. Die aufbereitete Aussonderungsdatei wurde an die Kolleginnen und Kollegen des zuständigen Fachreferats zur abschließenden Qualitätskontrolle weitergeleitet.

Nach der Freigabe durch das zuständige Fachreferat wurden die Dateien (Metadaten und Primärdokumente) in Archival Information Packages (AIPs)<sup>5</sup> strukturiert. Jedes AIP enthielt ein Primärdokument und die dazugehörigen Metadaten in Gestalt einer XBARCH XML-Datei. Eine XBARCH-Datei gibt Auskunft darüber, zu welcher Akte und zu welchem Vorgang das Primärdokument gehört. Zusätzlich sind auch Angaben zu den übrigen Dokumenten des Vorgangs enthalten. Weiterhin gibt XBARCH Auskunft über den Verlauf der Aussonderungsphase, das Archivformat der Primärdokumente und die notwendigen technischen Anforderungen (Hardware und Software) zur Darstellung der Primärdokumente. Um die Authentizität der Primärdokumente zu erhalten, enthält XBARCH einen eindeutigen Hash-Wert, der während der Pilotphase von der Archivierungslösung für jedes Primärdokument generiert wurde. Zudem beschreibt XBARCH die technische Umgebung (VBS) der Behörde, aus der die Aussonderung stammt.

Zum Abschluss des Ingest-Prozesses übergab die Softwarelösung, die für das Anbieten, Bewerten und Übernehmen der elektronischen Akten zuständig war, das Primärdokument und die XBARCH-Datei in einem unkomprimierten tar-Archiv an die Archivierungslösung zur endgültigen Speicherung. Bei diesem Vorgang wurden eine Auswahl von Metadaten in eine Archivdatenbank importiert, welche die Archivobjekte verwaltet und gleichzeitig die Grundlage für die Recherche bildete. Zusätzlich legte die Softwarekomponente des Speichersystems auf dem Datenträger ein Verzeichnis über die auf diesem Träger gespeicherten AIPs an. Auf Basis dieser Verzeichnisse könnte im Bedarfsfall die Archivdatenbank sowie der Kontext der Daten vollständig rekonstruiert werden.

Jedes AIP ist wie eine organische Zelle konzipiert, die Informationen über den Aufbau eines komplexeren Organismus enthält, zu dem sie gehört. Ein Nachteil dieser Methode ist die erhöhte Redundanz bei der Datenspeicherung. Jede XBARCH-Datei wird vermutlich im Durchschnitt 50 KB benötigen.

### 2.5 Recherche und Benutzung

Nach einem erfolgreichen Archivierungsvorgang wurde die Behörde informiert, ein Report erstellt und die Daten für die Benutzung freigeschaltet. Es war eine hierarchische Navigation (Provenienz-Akte-Vorgang-Dokument) und gleichzeitig eine Suche nach konfigurierbaren Kriterien möglich.

---

<sup>4</sup> Im Internet unter <http://hul.harvard.edu/jhove/> (letzter Zugriff 11.04.2007).

<sup>5</sup> Siehe Reference Model for an Open Archive Information System – NASA - Washington, DC; 2002 – S. 4-33 ff. Im Internet unter: <http://public.ccsds.org/publications/archive/650x0b1.pdf> (letzter Zugriff 11.04.2007).

## – Speicherung –



### 3. Ergebnisse des Pilotprojekts

Es ist gelungen, innerhalb von etwa zwei Monaten die Anforderungen des Bundesarchivs sowie alle wesentlichen Funktionen einer Archivierungslösung, d.h. vom Anfang (Produzenten-Archiv-Schnittstelle, Bewertung, Übernahme) bis zum Ende (Erschließung, Langzeitspeicherung, Access) mit zwei einfach zu integrierenden Standardlösungen umzusetzen.

Die Testphase erfolgte unter praxisnahen Bedingungen, wodurch nachgewiesen wurde, dass bei der Aufbereitung und Führung von digitalen Daten für die Archivierung eine genügende Kontrolle über die technischen Prozesse erlangt und die Anwender weitgehend von unnötigen manuellen und fehleranfälligen „Ad-hoc-Arbeitsschritten“ befreit werden können. Dadurch ist eine hohe Qualitätssicherung des eigentlichen Geschäftsprozesses möglich.

### 4. Ausblick

Die nächste Projektphase betrifft die Ausschreibung für die erste Ausbaustufe eines produktiven digitalen Archivs. Die Unterlagen für eine Ausschreibung werden zur Zeit am Bundesarchiv angefertigt.

In der Testphase konnten noch keine Geschäftsprozesse für die Migration von Dateiformaten erprobt werden. Nach dem heutigen Erfahrungsstand müssen Migrationen aber zwingend durchgeführt werden, um das elektronische Archivgut an die kommenden technischen Umgebungen anzupassen. Die dadurch unvermeidlichen Veränderungen am Archivgut müssen abgewogen, kontrolliert und sorgfältig dokumentiert werden. Nur so kann das Archiv die Authentizität seiner Objekte nachweisen und erhalten. Entsprechende Geschäftsprozesse sind für den Produktivbetrieb des Digitalen Archivs vorgesehen. Weiterhin sieht der Plan eine Erweiterung der Funktionalitäten des Digitalen Archivs bis 2010 vor.