

Elektronisch archivieren?

Ein Modellprojekt zur Archivierung von Unterlagen aus Geschäftsverwaltungssystemen

Wie die meisten kantonalen Archive setzt sich auch das Staatsarchiv Zug schon seit Längerem mit der Überlieferung nur noch digital gespeicherter Unterlagen auseinander. Das grosse mediale Interesse, das diesem Thema beschieden ist, rührt von offenen Fragen her, die von den Archiven noch zu beantworten sind. Es gehört allerdings zu den typisch archivischen Gegebenheiten, dass sich erst sehr viel später herausstellen wird, ob die gegebenen Antworten richtig oder falsch waren. Die kaum kalkulierbaren Konsequenzen von digitalen Umstellungen führen – trotz grundsätzlicher Sympathie und grossen Erwartungen an digitale Technologien – zu Skepsis gegenüber der Ablösung herkömmlicher Medien. (Die Volatilität in der Bewertung des Mediums Mikrofilm, die zwischen Ablehnung und Hochschätzung hin und her pendelt, stellt das antizyklische Pendant zu dieser Beobachtung dar.)

So sehr die digitale Datenverarbeitung den betrieblichen Alltag vereinfachen mag – für die Archivierung ist sie zunächst einmal ganz einfach eine neue und gewaltige Herausforderung. Digitale Daten unterscheiden sich von analogen, also augenlesbaren Unterlagen in zwei ganz wesentlichen Merkmalen: einerseits in ihrer binären Verschlüsselung und andererseits in ihrer nur mittelbaren Bindung an eine materielle Trägersubstanz. Selbstverständlich brauchen auch digitale Daten in irgendeiner Form einen Datenträger. Dieser Träger ist aber austauschbar: Digitale Daten können beliebig zwischen Festplatten, Magnetbändern, optischen Datenträgern oder Speicherbausteinen kopiert oder verschoben werden. Die beiden genannten Merkmale erlauben es, dass elektronisch gespeicherte Daten maschinell bearbeitet und verlustfrei kopiert werden können. Ihre absolute Unabhängigkeit von *spezifischen* Datenträgern und ihre maschinelle Interpretierbarkeit führen dazu, dass sie räumlich beliebig in Echtzeit verfügbar sind – ein Umstand, an den sich unsere Informationsgesellschaft sehr schnell gewöhnt hat. Leider ist damit auch die Wahrnehmung des spezifischen Wertes einer materiell verbrieften Information ein Stück weit verloren gegangen. Mit dem Wegfall der unmittelbaren Bindung zwischen Information und materieller Trägersubstanz steht allerdings nichts weniger als der klassische Gegenstand der Archivierung zur Disposition: Das Archiv hat sich bei der Bestan-

deserhaltung stets um materielle Fragen gekümmert. Darüber hinaus stehen dem Archiv aus den Trägersubstanzen analoger Unterlagen oft indirekte Metadaten (zum Beispiel Zeit-, Versions- und Herkunftsangaben) zur Verfügung, die bei digitalen Daten zuerst einmal explizit mitgespeichert werden müssen.¹ Dementsprechend lässt sich in der digitalen Welt die Frage nach dem Original nicht so ohne Weiteres beantworten. Die Sicherheit, die analoge Unterlagen durch ihre feste Verbindung von Trägersubstanz und Information unmittelbar vermitteln und meistens auch gewährleisten können, entfällt mit der Digitalisierung generell. Dieses virtualisierte Verhältnis zwischen der eigentlichen Datenspeicherung und der Datenrepräsentation dürfte – zusammen mit der zwingenden und absoluten Abhängigkeit von hochkomplexer, anfälliger Technologie und dem extrem raschen Wandel dieser Informationstechnologien – dafür verantwortlich sein, dass die umfassende Ablösung der Papierüberlieferung bislang ausgeblieben ist.

Die vorsichtige Haltung gegenüber digitalen Daten soll nun aber nicht heissen, dass im Bereich der elektronischen Archivierung nichts geschieht. Ganz im Gegenteil: Schon sehr Vieles ist geleistet worden. So steht mit dem OAI-Modell (Open Archival Information System)² ein theoretischer ISO-Standard zur Verfügung, der die elektronische Archivierung umfassend organisiert. Ferner gibt es mit PDF/A eine weitere ISO-Norm, welche schon heute für sequentiell organisierte Dateien ein taugliches Archivverpackungsformat zur Verfügung stellt. Und entsprechend gibt es auch Normen, welche die für die Archivierung relevanten semantischen und technischen Metadaten vorgeben,³ wobei die semantischen Metadaten den «Geschäftszusammenhang», die technischen Metadaten die Beschaffenheit der digitalen Dateien beschreiben – Angaben, die für einen künftigen Zugriff wichtig sind.

Die Schwierigkeit der digitalen Archivierung liegt also nicht darin, dass die Einzelprobleme grundsätzlich ungelöst wären. Sie besteht vielmehr darin, dass eine konzeptionelle Verbindung dieser Einzellösungen noch fehlt.

KOST-Modellprojekt 2007/08

Als die schweizerische Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) Ende 2006 die Mitgliederarchive aufforderte, Vorschläge für ein Modellprojekt 2007/08 zu unterbreiten, plädierte das Staatsarchiv Zug dafür, das Problem der Speicherung elektronischer Primärdaten anzugehen, und zwar sowohl theoretisch durch Beschreibung und Modellierung der nötigen Prozesse wie auch praktisch durch die Realisierung einer Pilotlösung. Der Vorschlag wurde von der KOST noch im gleichen Jahr für gut befunden und zusammen mit ähn-

¹ Die Sicherung solcher Metadaten muss im Lebenszyklus digitaler Daten vor der Archivierung durchgeführt werden. Archive engagieren sich darum verstärkt in der vorarchivischen Betreuung.

² <http://public.ccsds.org/publications/archive/650x0b1.pdf>

³ Für semantische Metadaten gibt es *EAD* (<http://www.loc.gov/ead/>), *GEVER i017* (<http://www.isb.admin.ch/themen/standards/alle/03228/?lang=de>) und *Dublin Core* (<http://dublincore.org/>). Für technische Metadaten gibt es *PREMIS* (<http://www.oclc.org/research/projects/pmwg/>) und *LMER* (<http://www.d-nb.de/standards/lmer/lmer.htm>).

lichen Anträgen aus den Staatsarchiven Basel-Stadt und St. Gallen als Grundlage für das Projekt «Archivierung von Unterlagen aus Geschäftsverwaltungssystemen» (AUGev) verwendet. In einer ersten Phase wurden von allen Beteiligten die Archivprozesse und Datenmodelle identifiziert, modelliert und dokumentiert und in einem Schlussbericht festgehalten. Das Ziel ging dahin, auf theoretischem Niveau, unabhängig von Institutionen und verwendeten Produkten, die digitale Archivierung zu begreifen und standardisiert in UML (Unified Modelling Language) zu beschreiben, um so das notwendige Wissen für deren Beherrschung in den Archiven zu verankern.

In der zweiten Phase sollten die erarbeiteten Prozessbeschreibungen und Datenmodelle in die Praxis umgesetzt und damit deren Funktionstüchtigkeit bewiesen werden. In dieser Phase wählten die drei beteiligten Staatsarchive unterschiedliche Lösungsansätze, bedingt durch jeweils spezifische Prämissen und Vorgaben (zum Beispiel Praxisunterschiede in der Ablageorganisation, in der Signaturbildung oder in der Bewertung von Unterlagen).

Für die Umsetzung der theoretischen Modelle, die in der Phase 1 erarbeitet worden waren, wurden im Staatsarchiv Zug die folgenden sieben Kernpunkte als zentral gesetzt:

1. Prozesse und Datenstrukturen müssen transparent sein.
2. Die digitale Archivierung muss unabhängig von einzelnen Plattformen und Programmen funktionieren.
3. Die digitale Archivierung muss den Grundanforderungen der OAIS-Norm entsprechen.
4. Die digitale Archivierung umfasst elektronische Daten, keine Programme.
5. Das Archivgut wird in Datenkapseln, so genannten «archival information packages» (aip), gespeichert. Diese sollen nicht nur während der Laufzeit eines Systems, also ad hoc, existieren, sondern persistent sein.
6. Die Datenkapseln sind autonom. Sie beschreiben sich und die übergeordnete Archivstruktur.
7. Das digitale Archiv ist ein Online-System.⁴

Ausgehend von diesen sieben Thesen entwickelte das Staatsarchiv Zug das offene Software-Framework *Java Open Archive Implementation (JOAI)*. Das Gerüst basiert aus verschiedenen Gründen auf der Java-Technologie. Java ist objektorientiert, plattformunabhängig und durch das Internet stark verbreitet. Zudem sind die Entwicklungswerkzeuge sowie der Compiler von Sun Microsystems unter der GNU General Public License 2006 veröffentlicht worden. Das Gerüst JOAI ist objektorientiert angelegt und besteht aus einer Vielzahl von Programmkomponenten, die je eine ganz spezifische Aufgabe lösen und einen eigenen Lebenszyklus besitzen und daher gezielt austauschbar sind. Darüber hinaus ermöglicht die Objektorientierung, Probleme deduktiv – vom Allgemeinen zum Speziellen hin – zu lösen, und zwar deshalb, weil die verwendeten Programmkomponenten in einem hierarchischen Verhältnis zueinander stehen, sich also gewissermassen auseinander ableiten.

Damit wird durchaus denkbar, dass andere Archive bestehende Teile aus dem Gerüst JOAI für sich wieder verwenden und nur noch Lösungen für ihre spezifischen Anforderungen – zum Beispiel für den Bewertungsvorgang oder die Signierung – entwickeln müssten. So könnten aus dem virtuellen Werkzeugkasten JOAI heraus beliebige digitale Magazine gebaut respektive weiter gebaut werden – getreu der Überzeugung, dass sich Software in solchen «akademischen», öffentlichen Diskursen am effizientesten weiter entwickeln lässt.

Das digitale Magazin, welches das Staatsarchiv Zug für sich selbst entwickelt, heisst gemäss der Konvention umgekehrter Domännennamen *ch.stazg.joai*. Darin werden die Aufgaben der digitalen Archivierung in den vier autonomen Bereichen *Submission*, *Dissemination*, *Migration* und *Packages* gelöst. Die *Submission* umfasst die Übernahme, Bewertung und Aufbereitung von digitalen Daten durch das Archiv. Sie beinhaltet Prozesse, die jeweils nur einmal, nämlich beim Übergang ins Archiv, von den zu archivierenden Daten durchlaufen werden. Die *Dissemination* beinhaltet alle Aspekte der Benutzung von digitalen Daten. Es handelt sich um wiederkehrende, lesende Prozesse, die bei jedem Zugriff ausgeführt werden. Die *Migration* beinhaltet ebenfalls wiederkehrende Prozesse, die der Erhaltung digitaler Bestände dienen. Dazu gehören Dateiformat- und Datenträgermigrationen. Diese werden zyklisch alle drei Jahre durchgeführt. Die drei genannten Bereiche decken Prozesse ab, der vierte Bereich der *Packages* hingegen beinhaltet Datenstrukturen wie etwa Archivpakete, Archivjournale oder Logbücher. Für die Beschreibung dieser Datenstrukturen werden *XML/XSD/XSL*-Technologien⁵ eingesetzt, für die Verpackung des digitalen Archivguts das offene Archivformat *JAR*.⁶ Als Testdaten werden für den Betrieb des Pilots ausgewählte Dossiers aus der elektronischen Geschäftsverwaltung KONSUL verwendet.

Im September 2008 wird an der Schlussveranstaltung des KOST-Modellprojekts 2007/08 der Pilot *ch.stazg.joai* offiziell vorgestellt werden. Ob dieser Pilot weiterentwickelt werden wird oder ob er in der Versenkung halbfertiger Softwarekonstrukte verschwindet, ist heute noch völlig offen. Auf jeden Fall diente er dazu, die Wahrnehmung archivischer Bedürfnisse im Zusammenhang mit digitalen Medien entscheidend zu schärfen. Und so ist seine Entwicklung ein weiterer Schritt auf dem Weg, von welchem die Zukunft zeigen wird, ob er zu digitalen oder aber zurück zu analogen Archivmedien führt.

Olivier Debenath

⁴ Im Gegensatz dazu wären auch externe Datenträger denkbar, welche – wie herkömmliches Archivgut – in einem Magazin überliefert würden. Die kurze Lebenszeit von Datenträgern und Kodierungen spricht jedoch gegen einen solchen Lösungsansatz.

⁵ S. dazu <http://www.w3.org/XML/> bzw. <http://www.w3.org/XML/Schema> und <http://www.w3.org/Style/XSL/>

⁶ <http://java.sun.com/j2se/1.5.0/docs/guide/jar/jar.html>