

Archivierung digitaler Unterlagen der Verwaltung

**Ein Projekt der Generaldirektion der Staatlichen Archive Bayerns
gefördert von der Deutschen Forschungsgemeinschaft**

2002-2005

Schlussbericht

1. Ziele und Arbeitsprogramm

Ziel des Projekts war die Entwicklung eines Modells für die Übernahme, dauerhafte Speicherung und Nutzung digitaler Unterlagen der Verwaltung. Das Projekt beschränkte sich im wesentlichen auf digitale Sachakten, da zu anderen Überlieferungsformen, wie z. B. Geoinformationssystemen, Kataster- oder Statistikenunterlagen, in anderen Bundesländern Forschungen liefen bzw. beabsichtigt waren.

Das Arbeitsprogramm sah ursprünglich folgende Teile (Arbeitspakete) vor:

1. Bestandsaufnahme der weltweit laufenden Forschungen zum Thema "Electronic Records".
2. Ermittlung der essentiellen und damit konstitutiven Merkmale eines digitalen Akts; Abgrenzung der akzidentiellen und damit für eine Speicherung irrelevanten Elemente.
3. Ermittlung der für die Speicherung von Kontext und Historie eines digitalen Akts unverzichtbaren Metadaten. Die Ergebnisse sollten in detaillierte Empfehlungen für die elektronische Schriftgutverwaltung bei den Behörden eingebracht werden.
4. Entwicklung des Modells eines selbstdokumentierenden und gekapselten - also für sich allein existenzfähigen - digitalen archivierbaren Akts (Objekts), der plattformunabhängig und dauerhaft nutzbar ist.
5. Suche nach Lösungen für die Speicherung oder - auch unter dem Gesichtspunkt der Authentizität vertretbarer - Wandlung bzw. Auflösung von digitalen Signaturen in den Unterlagen.
6. Entwicklung einer "Schnittstelle" für eine soweit wie möglich automatisierbare Übernahme digitaler Unterlagen von der Verwaltung ins Archiv.
7. Entwicklung einer nach Möglichkeit automatisierbaren Erschließung der gespeicherten Unterlagen mit dem Ziel, der Forschung möglichst vielseitige Möglichkeiten der Recherche und des Zugriffs - lokal wie über das Internet - zu eröffnen. Hierbei waren Maßnahmen zur Wah-

zung der gesetzlich vorgeschriebenen Einschränkungen der Benützung (Schutzfristen, Persönlichkeitsschutz) vorzusehen.

Zur Unterstützung des Projekts wurde die seit 1998 bei der Generaldirektion eingerichtete Arbeitsgruppe "Archivierung digitaler Unterlagen der Verwaltung" (AG AdUV) herangezogen, die aus fünf Archivarinnen und Archivaren (z.T. mit IT-Fachwissen, z.T. Spezialisten für die Aktenaussonderung) aus der Generaldirektion, dem Bayerischen Hauptstaatsarchiv und dem Staatsarchiv München bestand.

2. Verlauf des Projekts

Vorbemerkung:

Der Projektstart war für Oktober 2000 vorgesehen. Da sich auf die damalige Stellenausschreibung kein geeigneter Bewerber meldete, wurden der Projektumfang und damit auch die Anforderungen an den Mitarbeiter reduziert und die Stelle neu ausgeschrieben. Zum 1. Januar 2002 konnte das Projekt mit Dr. Rodrigo Readı Nasser als wissenschaftlichem Mitarbeiter beginnen. Da er auf eigenen Wunsch halbtags beschäftigt war, lief das Projekt bis zum 31.12.2005. Der Projektverlauf hat gezeigt, dass für das Projektmanagement (insbesondere für die laufende Dokumentation und die Öffentlichkeitsarbeit) zu wenig Ressourcen eingeplant waren. Es wäre daher sinnvoll gewesen, beim Projektantrag neben dem wissenschaftlichen Mitarbeiter auch eine Stelle für einen Projektmanager vorzusehen.

Zum Arbeitspaket 1:

Ausgangspunkt für die Bestandsaufnahme der Forschung war die Arbeit von Alf Erlandsson,¹ eine kommentierte Literaturübersicht aus dem Jahre 1997, die in vielen Teilen heute noch lesenswert ist. Darauf aufbauend wurden thematisch einschlägige Webseiten regelmäßig gesichtet und ein Literaturapparat aufgebaut. Inzwischen kümmert sich das Projekt PADI² um die Dokumentation und z. T. auch Archivierung der wichtigen Projekte auf internationaler Ebene; in Deutschland nimmt das Projekt nestor³ diese Aufgabe wahr. Mit dem Werk von Uwe M. Borghoff et al. liegt seit 2003 ein einführendes Handbuch für die Langzeitarchivierung vor,⁴ das durch das Buch von Ross Harvey⁵ ergänzt wird. Eine Fundgrube speziell für den Archivbereich bietet die Literatursammlung des niederländischen Projekts Digitale Du-

¹ Alf Erlandsson: Electronic Records Management. A Literature Review, Paris 1997 (International Council on Archives - Studies, 10).

² <http://www.nla.gov.au/padi/>.

³ <http://www.langzeitarchivierung.de>. PADI ist mittlerweile auch über nestor erreichbar und umgekehrt.

⁴ Uwe M. Borghoff - Peter Rödıg - Jan Scheffczyk - Lothar Schmitz: Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg 2003. Maßgebend ist inzwischen die überarbeitete englische Fassung: Long-Term Preservation of Digital Documents, Berlin-Heidelberg-New York 2005.

urzaamheid,⁶ und mit dem Arbeitsbuch⁷ des Internationalen Archivrats liegt nun auch eine beachtenswerte Handlungsanweisung für die Archivierung digitaler Unterlagen vor.

Nach wie vor lesenswert ist die Studie "Gesamtschweizerische Strategie zur dauerhaften Archivierung von Unterlagen aus elektronischen Systemen", auch wenn einige Schlussfolgerungen - wie z. B. mit der Verknennung der Möglichkeiten des Mikrofilms als Langzeitspeichermedium für digitale Unterlagen - heute nicht mehr zutreffen.⁸

Zum Arbeitspaket 2:

Ein digitaler Akt ist logisch gesehen ein Container, der neben Metadaten zum Akt eine Liste der zum Akt gehörenden Vorgänge enthält.⁹ Jeder Vorgang stellt seinerseits einen Container dar, der wiederum vorgangsspezifische Metadaten und eine Liste der zu ihm gehörenden Dokumente enthält. Dokumente bestehen aus Metadaten und den eigentlichen inhaltlichen Informationen; sie können mit weiteren Schriftstücken (z.B. Anlagen) verbunden sein. Des Weiteren sind zumindest auf der Vorgangs- und Dokumentenebene Protokollinformationen vorhanden (Geschäftsgangvermerke, Abzeichnungen, Versionierungen). Alle diese Elemente sind mitsamt ihrer hierarchischen Struktur für den digitalen Akt konstitutiv und unverzichtbar.

Da ein digitaler Akt zwar einen Anfang hat (Datum der z.d.A.-Verfügung des ältesten Vorgangs), aber zumindest theoretisch kein Ende, da sein Umfang nicht (wie beim Papier) durch physische Restriktionen begrenzt ist, werden bei der Archivierung Aktenschnitte gebildet, bei denen Unterlagen aus dem behördlichen Ablagesystem (elektronische Altregistrierung) herausgelöst und dem Archiv übergeben werden.

Bei der technischen Betrachtung muss zwischen dem Teil des Lebenszyklus der Unterlagen, der sich in der Verwaltung abspielt, und dem Teil ab der Archivierung unterschieden werden. Die schriftgutproduzierenden Systeme (Dokumentenmanagement- bzw. Vorgangsbearbeitungssysteme) in der Verwaltung legen Dokumente zunächst in ihren Ursprungsformaten

⁵ Ross Harvey: Preserving Digital Materials, München 2005. Leider wird nur englischsprachige Literatur berücksichtigt.

⁶ <http://www.digitaleduurzaamheid.nl> . Von allen Projekten zur Archivierung digitaler Unterlagen ist dieses auch dank massiver finanzieller Ausstattung wohl am weitesten gediehen. Neben Empfehlungen zur Archivierung unterschiedlicher Typen elektronischer Unterlagen wurde Software für die Behandlung von E-Mail entwickelt. Auch liegen erstmals in Europa konkrete Kostenschätzungen für Einrichtung und Betrieb eines "elektronischen Archivs" vor. Zur Betrachtung der Speicherkosten sollten allerdings die Kalkulationen von Jonas Palm, The Digital Black Hole [2006] (http://www.tape-online.net/docs/Palm_Black_Hole.pdf) unbedingt herangezogen werden.

⁷ International Council on Archives - Committee on Current Records in an Electronic Environment: Electronic Records. A workbook for Archivists, Paris 2005 (International Council on Archives - Studies, 16). Das Buch ist im wesentlichen eine Gemeinschaftsarbeit europäischer Archivare.

⁸ Zu finden unter <http://www.vsa-aas.org/index.php?id=110&L=0> .

⁹ In manchen Verwaltungen werden keine Akten mehr gebildet, sondern nur Vorgänge. In diesen Fällen fehlt also die Aktenebene.

(z.B. Word oder Excel) in einem Dateisystem ab. Die Metadaten und Protokollinformationen werden von den meisten Produkten in einer Datenbank geführt.

Bei der Archivierung werden Akten und Vorgänge (die oben erwähnten Container) durch Textdateien dargestellt, die hierarchisch aufgebaut sind und Metadaten sowie Verweise auf die inhaltlichen Informationen enthalten. Diese Textdateien können nach dem SGML-Standard oder nach den vom W3C entwickelten XML-Regeln kodiert sein.¹⁰ Es ist grundsätzlich unerheblich, ob man für Akt, Vorgang und Dokument separate Metadaten-Dateien vorsieht oder eine komplette Datei pro Aktenschnitt. Wir haben uns für letztere Lösung entschieden, um die Zahl der Dateien - und damit der sie verknüpfenden Verweise - überschaubar zu halten.

Für die Inhaltsinformationen hatten wir ursprünglich, den Empfehlungen des Bundesarchivs folgend, ein standardisiertes, offenes Bildformat (TIFF) vorgesehen. Die Entwicklung eines speziellen Archivformats (PDF/A) und dessen Festschreibung als Standard durch die ISO¹¹ haben uns jedoch bewogen, von TIFF abzugehen und uns für PDF/A zu entscheiden. Dateien in diesem Format sind erheblich kleiner als vergleichbare TIFF-Dateien; dies gilt insbesondere, wenn Farbinformationen berücksichtigt werden müssen. Solche sind auch in Verwaltungsunterlagen in steigendem Maß anzutreffen.

Diese Entscheidung hat Konsequenzen für die Erhaltung der Eigenschaften eines Dokuments nach der Archivierung. So gehen Interaktivität (Ausführung von Sprungmarken, Programmaufrufe, Makros) und auch Elemente des Aussehens (z. B. Farbtreue oder absolute Schriftgrößen, u.U. auch Fonts) verloren. Derartige Eigenschaften halten wir bei Textdokumenten nicht für essentiell; ihr Verlust ist in Kauf zu nehmen, zumal er nicht erst bei der Archivierung, sondern schon bei der Speicherung in der behördlichen Altregistratur auftritt. Eine Konservierung der originären Dateiformate ist in Anbetracht der Aufbewahrungsfristen schon dort nicht möglich, daher werden die Dokumente in ein so genanntes finales Format migriert, das eine zumindest mittelfristige Lesbarkeit gewährleistet. Zweckmäßigerweise sollte dieses finale Format dem für die Archivierung gewünschten entsprechen.

Problematisch ist das Bildformat bei Tabellenkalkulationsblättern, in denen Berechnungsformeln hinterlegt sind. Diese Formeln gehen bei der Formatumwandlung verloren. Das Problem ließe sich lösen, wenn die Hersteller von Office-Software das XML-basierte Open-Document-Format (ODF) als Dokumentformat verwenden würden; hierüber wird seit kurzer Zeit wieder intensiv diskutiert.¹²

Multimediale Dokumente lassen sich auf diese Art begrifflicher Weise weder ablegen noch archivieren. Für bestimmte Aktentypen (z. B. polizeiliche Ermittlungsakten mit Asservaten in

¹⁰ In unserem Projekt haben wir uns für ein stark reduziertes SGML entschieden, da hierdurch die Textdateien im Umfang verringert werden und für die geplante Ausbelichtung auf Film besser geeignet sind. Eine Umsetzung auf XML (DTD oder Schema) ist ohne Probleme möglich.

¹¹ ISO 19005.

¹² http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office .

Form von Video- oder Schallaufzeichnungen) ergeben sich daher Schwierigkeiten, für deren Lösung jedoch ein vielversprechender Ansatz vorliegt (s. u. im Abschnitt Mikroverfilmung). Für die Definition eines digitalen Akts ist auf die Darstellung der niederländischen Arbeitsgruppe "Digital Preservation Testbed" zu verweisen.¹³

Zum Arbeitspaket 3:

Der Metadatenkatalog für Sachakten der Verwaltung, entwickelt von der AG AdUV und gemeinsam mit Dr. Readi überarbeitet, ist 2004 erschienen¹⁴ und auch online¹⁵ verfügbar (die Printausgabe ist vergriffen). Er konnte nicht nur nahezu vollständig in das vom Freistaat Bayern mit einer Landeslizenz erworbene Vorgangsbearbeitungssystem "Fabasoft eGovSuite+Bayern" integriert werden, sondern hat darüber hinaus auch Eingang in die zur Zeit auf der Basis von DOMEA in Entwicklung befindliche Aussonderungsschnittstelle XArchiv gefunden. Diese Arbeit wird von der "Arbeitsgruppe ESys", einem Spiegelgremium der Archivreferentenkonferenz zu einschlägigen Bund-Länder-Arbeitsgruppen, geleistet und soll 2006 abgeschlossen werden.

Zum Arbeitspaket 4:

Der Gedanke, einen sich selbst dokumentierenden, gekapselten und damit autonomen digitalen Akt modellhaft zu entwickeln, entstand bei der Beschäftigung mit dem Projekt "Victorian Electronic Records Strategy" (VERS), das bereits 1998 abgeschlossen war.¹⁶ Bei allen Vorteilen, die dieses Modell bietet, dürfen die Nachteile nicht übersehen werden: Da die Dokumentinhalte, die vorher als PDF gespeichert wurden, innerhalb einer XML-Datei (Container, Metadaten und Signaturen) mit Base64 kodiert werden, entstehen bei Akten (records) mit zahlreichen Dokumenten außerordentlich umfangreiche Dateien. Für die Recherche müssen die Dokumente "entkapselt", also extrahiert und als PDF geladen werden (alternativ kann man die Dokumente vor der "Kapselung" indizieren und den Index in einer Datenbank führen). Drittens fehlt eine auf Dauer praktikable Lösung für den langfristigen Umgang mit der digitalen Akt "versiegelnden" Signatur. Denn in VERS müssen die archivierten Akten in regelmäßigen Abständen mit neuen Signaturen versehen werden, um ihre Authentizität zu erhalten. Dadurch wachsen die Dateien weiter an, da die ungültig gewordenen Signaturen nicht entfernt werden dürfen.

¹³ From digital volatility to digital permanence. Preserving email, Den Haag 2003, S. 13-17.

¹⁴ Digitale Unterlagen Nr. 2: Metadaten für die Aussonderung und Archivierung digitaler Sachakten. Empfehlungen für die Behörden des Freistaates Bayern, hrsg. von der Generaldirektion der Staatlichen Archive Bayerns, München 2004.

¹⁵ <http://www.gda.bayern.de/digeinl.htm> .

¹⁶ Victorian Electronic Records Strategy - Final Report. Public Record Office Victoria 1998. Online verfügbar unter <http://www.prov.vic.gov.au/vers/published/final.htm> . Gute Zusammenfassung bei Borghoff (wie Anm. 4), S. 155-161 und 187-193.

Diese Nachteile haben uns bewogen, auf die Eigenschaften Selbstdokumentation und Kapselung zu verzichten. Das in unseren Publikationen (s.u. 3.1) beschriebene Modell ist von Hard- und Softwareplattformen unabhängig. Für eine Dokumentation der Struktur und der Metadaten muss natürlich gesorgt sein.

Zum Arbeitspaket 5:

In den Projekten ArchiSig¹⁷ bzw. ArchiSafe¹⁸ wurde eine Lösung entwickelt, die eine Speicherung und Aktualisierung elektronisch signierter Dokumente für eine begrenzte Zeit (ca. 30 Jahre) auf wirtschaftliche Weise ermöglicht. Für die Endarchivierung jedoch wird auch hier vorgeschlagen, die Signaturen in Klartext aufzulösen und so nur die Informationen: Faktum der Signierung, Datum, Signierer, zertifizierende Stelle beim Dokument zu speichern.

Da bis heute nicht geklärt ist, wie die künftig zum Einsatz kommende digitale Signatur aussehen wird (rechtlich wie technisch) und wie das Umfeld (PKI-Struktur) gestaltet wird, halten wir weitere Arbeiten in dieser Richtung zur Zeit nicht für sinnvoll.

Zum Arbeitspaket 6:

Eine Aussonderungsschnittstelle besteht nach unserem Verständnis aus drei Teilen:

1. einer vom Archiv vorgegebenen Dateistruktur, einem Metadatenkatalog und Vorgaben für die Formate der dem Archiv zu übergebenden Dateien,
2. einem vom Archiv entwickelten Regelwerk für den Ablauf des Aussonderungsverfahrens,
3. einem Softwaremodul, mit dem die in der behördlichen Registratur abgelegten und dem Archiv zu übergebenden Daten in die in Punkt 1 genannte Form gebracht und gemäß den in Punkt 2 spezifizierten Regeln übergeben werden.

Eine elektronische Registratur wird mit Dokumentenmanagement-Systemen oder speziellen Registraturverwaltungssystemen geführt. Das in Punkt 3 angesprochene Softwaremodul muss also in diesen Systemen realisiert werden. Das ist Sache der Behörden bzw. ihrer Systemlieferanten, die hierfür mit dem Archiv zusammenarbeiten müssen.

Voraussetzung für das Funktionieren der Aussonderungsschnittstelle ist die Hinterlegung eines Fristen- und Bewertungskatalogs im elektronischen Aktenplan des behördlichen Systems. Da dieser Katalog behörden- oder behördenspartenspezifisch ist, muss die Hinterlegung spätestens bei der Konfiguration des Systems zwischen Behörde und Archiv vereinbart werden. Die technische Einrichtung hängt von den Gegebenheiten des jeweiligen Systems ab. Änderungen sowohl der Aufbewahrungsfristen als auch der Bewertungskriterien müssen jederzeit möglich sein.¹⁹

¹⁷ <http://www.archisig.de>

¹⁸ <http://www.archisafe.de>

¹⁹ Diese Notwendigkeit wird von den Systemproduzenten häufig nicht gesehen, wie Gespräche im Rahmen des bayerischen DMS-Projekts gezeigt haben.

Als Output liefert die Aussonderungsschnittstelle Metadaten-Dateien und Dateien mit Primärinformationen in den vom Archiv spezifizierten Strukturen und Formaten. Die Speicherung im Archiv erfolgt nach einer Vollständigkeits- und Qualitätskontrolle zunächst parallel auf einem dedizierten Server und offline. Die Unterlagen werden für einen Zeitraum von mindestens fünf Jahren²⁰ digital vorgehalten, um Rückgriffe der abgebenden Behörden zu erleichtern. Innerhalb dieses Zeitraums ist die Umsetzung auf das Langzeitspeichermedium Mikrofilm (s.u.) vorzunehmen.

Das Bayerische Staatsministerium der Finanzen hatte vor einigen Jahren das Dokumentenmanagement-System SER DOMEA der Firma SER eGovernment Deutschland GmbH eingeführt. Daraufhin wurde mit dem Ministerium und der Firma die Anpassung des DMS an die archivischen Anforderungen für die Aussonderung gemäß dem von Dr. Readi entwickelten Modell und die Programmierung einer entsprechenden Schnittstelle vereinbart. Nach hoffnungsvollem Beginn konnte dieser Projektteil jedoch aus verschiedenen Gründen nicht fortgeführt werden.

Die in unserem Projekt entwickelten Anforderungen an eine Aussonderungsschnittstelle waren mittlerweile in die Ausschreibung eines landesweiten bayerischen DMS eingegangen. Mit dem im bayerischen DMS-Projekt (ELDORA) federführenden Innenministerium und der Firma Fabasoft wurde vereinbart, diese Schnittstelle nach Abschluss von X-Archiv technisch zu realisieren.

Zum Arbeitspaket 7:

Für die Recherche nach und in elektronischen Unterlagen ist ein stufenweises Vorgehen zu empfehlen. Grundsätzlich sollen diese Unterlagen keinen besonderen Zugangsweg benötigen, sondern in die Findmittellandschaft eines Archivs integriert sein. In einem ersten Schritt werden daher die Metadaten, die man für ein bis auf die Aktenebene greifendes Findmittel benötigt, aus den Metadaten-Dateien extrahiert (Betreff, Laufzeit, Klassifikationsstufe, Signatur). Das Archiv kann diese Metadaten bei Bedarf um weitere ergänzen; auch kann aus ihnen ein Index erzeugt werden.

Hierzu wird immer wieder die Frage gestellt, ob denn die in der behördlichen Vorgangsbearbeitung entstehenden Metadaten hinreichend zuverlässig sind; das gilt besonders für jene, die manuell eingegeben werden müssen (Betreffe). Das Archiv dürfte kaum in der Lage sein, diese über Stichproben hinaus zu überprüfen (das ist bereits jetzt bei umfangreichen Papierunterlagen der Fall). Da die Behörde selbst ihre Akten jederzeit finden können muss und die Unzulänglichkeiten einer Volltextrecherche (zu umfangreiche Treffermengen bei unvollständigen Ergebnissen) auch dort bekannt sind, steht zu hoffen, dass die Qualität der Schriftgutverwaltung ein hinreichendes Niveau haben wird. Eine Überprüfung abzugebender Unterlagen

²⁰ Schätzwert, der sich, wenn einmal Erfahrungen vorliegen, durchaus ändern kann.

durch das Archiv mittels einer Online-Recherche in der Registratur muss jedoch stets möglich sein.

Vom Findmittel aus kann nun - das ist ein Vorteil elektronischer Unterlagen - in den Metadaten-Dateien selbst recherchiert werden, sofern keine Zugangseinschränkungen bestehen. Der Forscher kann dadurch die Suche in den Primärinformationen auf jene beschränken, die er wirklich braucht, und sich stundenlanges "Blättern" in den elektronischen Akten weitgehend ersparen. Diese zweite Stufe des Recherchierens wird vorerst im Archiv vor Ort erfolgen müssen, da eine zweifelsfreie Identifikation des Forschers über das Internet zur Zeit noch nicht gewährleistet werden kann.²¹

Das Modell einer Findmitteldatenbank, die sowohl Papierunterlagen als auch elektronische Unterlagen gleichermaßen erschließt, hat Dr. Readi entwickelt;²² ein Prototyp wird zur Zeit getestet.

3. Weitere Besonderheiten des Projekts

3.1. Berücksichtigung besonderer Formen digitaler Unterlagen:

Der aufgrund des vorläufigen Scheiterns der Schnittstellenprogrammierung eingetretene Zeitverlust konnte anderweitig ausgeglichen werden, denn inzwischen hatte das Polizeipräsidium Niederbayern/Oberpfalz aus dem Bestand des aufgelösten Präsidiums der Bayerischen Grenzpolizei 83 Magnetbänder und 25 Disketten mit polizeistatistischen Daten aus den 90-er Jahren an das Bayerische Hauptstaatsarchiv abgegeben. Begleitunterlagen, insbesondere technischer Art, fehlen. Schon das Polizeipräsidium konnte die Bänder mangels geeigneter Laufwerke nicht mehr lesen. Dr. Readi hat vom Mathematischen Institut der Universität München ein altes QIC-150-Laufwerk als Dauerleihgabe beschafft und die Bänder ausgelesen. Sie enthalten tar-Archive (ca. 10.000 Dateien) mit Statistiktabelle, die mit einem Editor wie Emacs problemlos gelesen werden können. Die Disketten (5,25 Zoll) ließen sich dank der Hilfe des Instituts für Softwaretechnologie der Universität der Bundeswehr München (Prof. Borghoff) bis auf eine ebenfalls auslesen. Da die knapp 70.000 Tabellen mit der Textverarbeitung HIT (Siemens) erstellt waren und dessen Format nur unvollständig dokumentiert ist, bedurfte es zeitraubender Experimente, bis eine skriptgesteuerte automatisierte Umsetzung in ASCII-kodierten Text gelang. Die Tabellen liegen nun auf CD-ROM vor, der Zugriff erfolgt über Inhaltsverzeichnisse in HTML.

²¹ Zugangskennung und Passwort genügen hierfür u.E. nicht; um den Einsatz qualifizierter Signaturen wird man nicht herunkommen.

²² Rodrigo Readi Nasser, Findmitteldatenbanken. Ein Beitrag zur Erschließung hybrider Akten. In: Archivalische Zeitschrift 87 (2005), S. 235-248.

Statistiktabellen sind zwar per se keine Akten, kommen aber durchaus als Anlagen zu diesen vor (meistens als Excel-Kalkulationsblätter), so dass die Beschäftigung mit dieser Überlieferung nicht projektfremd war. Auch zeigt der Fall, welche Probleme bei der Übernahme großer Mengen von Altdaten außerhalb eines geregelten Aussonderungsverfahrens auftreten können.

3.2. Mikrofilm als langzeitstabiles Archivierungsmedium:

Als Dr. Readi und der Projektleiter anlässlich der 7. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" im März 2003 auf den Mikrofilm als Speichermedium für digitale Unterlagen hinwiesen, waren die Reaktionen eher skeptisch. Inzwischen gilt der Film jedoch als durchaus konkurrenzfähig zur digitalen Speicherung, was unter anderem dem Erfolg des Projekts ARCHE²³ zu verdanken ist.

Unser Lösungsweg sieht vor, die Metadaten-Dateien einer Aktenabgabe (bestehend aus einer die Abgabe beschreibenden Datei, Dateien für jeden Aktenschnitt und Dateien mit Primärinformationen) in einer bestimmten Abfolge auf Mikrofilm auszubelichten. Dabei ist es im Hinblick auf die spätere Rückdigitalisierung für die Benützung notwendig, den in den Metadaten-Dateien verwendeten Font in einen OCR-geeigneten Font (z.B. OCR-A oder OCR-B) umzuwandeln. Dadurch kann - sofern die Schriftgröße nicht zu klein gewählt wurde - eine fehlerfreie Rückdigitalisierung gewährleistet werden.

Geräte für eine Ausbelichtung sind inzwischen auf dem Markt; die Kosten sind jedoch beträchtlich, und die Bedienung ist nicht einfach. Daher empfiehlt es sich, die Ausbelichtung von einem Dienstleister vornehmen zu lassen. Für die Rückdigitalisierung können handelsübliche oder auf diese Aufgabe spezialisierte Scanner (z.B. von der Firma Zeutschel) verwendet werden. Was noch fehlt, ist eine Software, die aus den Scans der Metadaten-Dateien im Anschluss an die OCR-Lesung diese Dateien vollständig (ohne Unterbrechungen, die von der Aufteilung einer Datei auf mehrere Filmaufnahmen herrühren) rekonstruiert.

Zwei Probleme bleiben jedoch: Selbst wenn es gelingt, 16 oder mehr Einzelaufnahmen auf einem Filmbild (Frame) unterzubringen, kann die Menge der Filme in Anbetracht des Umfangs digitaler Unterlagen (man denke nur an die oft zahlreichen Versionen eines Schriftstücks, die während der Vorgangsbearbeitung entstehen und aus Gründen der Nachvollziehbarkeit des Verwaltungshandelns sämtlich aufgehoben werden müssen) ein Maß erreichen, das den Umgang mit diesem Medium aufwändig und damit kostspielig macht. Des Weiteren können multimediale Überlieferungsformen so nicht gespeichert werden. Es gibt jedoch Ansätze,²⁴ den Film, und zwar Mikrofarbfilm, als digitalen Speicher zu verwenden. Hierbei wird nicht ein analoges Bild erzeugt und ausbelichtet, sondern der Bitstrom "verpixelt". Mittels einer verlustlosen Komprimierung und der Nutzung der drei Farblayer sollen sich Speicherkapazitäten von ca. 700 MB auf einem Mikrofilm bei gleichzeitiger Redundanz erreichen

²³ <http://www.landesarchiv-bw.de/arche/>

²⁴ <http://www.peviar.ch> .

lassen. Dieser Lösungsweg verlangt jedoch eine hinreichende Dokumentation der zum Interpretieren des Bitstroms notwendigen Algorithmen und Dateiformate.

4. Kooperationen

Projektpartner waren das Bayerische Staatsministerium der Finanzen und der Bayerische Verwaltungsgerichtshof.

Das Kompetenzzentrum Germanistik an der Universität Trier (Prof. Gärtner) unterstützte Dr. Readi bei Fragen zu SGML. Das Institut für Softwaretechnologie der Universität der Bundeswehr München (Prof. Borghoff) half nicht nur bei Hardwareproblemen, sondern brachte auch Erfahrungen aus einem Archivierungsprojekt mit der Bayerischen Staatsbibliothek ein, zu deren Digitalisierungszentrum (Frau Dr. Dörr) ebenfalls fruchtbare Kontakte aufgebaut wurden. Der Seniorchef der Firma mfm Hofmaier in München, Dipl.-Ing Dietrich Hofmaier, half uns mit wertvollen, anderweitig kaum zu beschaffenden Fachinformationen, bestärkte unsere Sicht des Mikrofilms als langzeitstabilem Träger und vermittelte den Kontakt zur Firma Zeuschel. Wertvolle Informationen brachten auch die Diskussionen im Projekt nestor, an dem die Staatlichen Archive Bayerns als Partner beteiligt waren.

5. Publikationen

Rodrigo Readi Nasser und Karl-Ernst Lupprian, Ein Modell für die elektronische Darstellung hybrider Verwaltungsakten in Archiven. In: Archivalische Zeitschrift 85 (2003), S. 147-173.

Karl-Ernst Lupprian, Anforderungen an Metadaten für die Aussonderung und Übernahme digitaler Akten. In: Atlanti. Review for Modern Archival Theory and Practice 2003, S. 25-37.

Digitale Unterlagen. Nr. 2: Metadaten für die Aussonderung und Archivierung digitaler Sachakten. Empfehlungen für die Behörden des Freistaates Bayern, hrsg. von der Generaldirektion der Staatlichen Archive Bayerns, München 2004.

Rodrigo Readi Nasser, Karl-Ernst Lupprian und Lothar Saupe, The Key Role of Metadata for Permanent Preservation of Digital Records in the Archival Environment - A Message from the Far Future. In: Metadata in Preservation. Selected Papers from an ERPANET Seminar at the Archives School Marburg, 3-5 September 2003, hrsg. von Frank M. Bischoff, Hans Hofmann und Seamus Ross, Marburg 2004 (Veröffentlichungen der Archivschule Marburg - Institut für Archivwissenschaft - Nr. 40), S. 185-203.

Karl-Ernst Lupprian, Expertengespräch "Archivierung digitaler Unterlagen der Verwaltung". In: Nachrichten aus den Staatlichen Archiven Bayerns, Nr. 48/2004, S. 6-8.

Rodrigo Readi Nasser, Findmitteldatenbanken. Ein Beitrag zur Erschließung hybrider Akten. In: Archivalische Zeitschrift 87 (2005), S. 235-248.

6. Öffentlichkeitswirksame Präsentationen

Eine Präsentation von Ergebnissen der Arbeitspakete 2, 3 und 4 fand anlässlich der 7. Tagung des Arbeitskreises für die Archivierung von Unterlagen aus digitalen Systemen im März 2003 in Berlin statt. Der seinerzeit vom Veranstalter geplante Tagungsband ist jedoch nicht erschienen. Der hierfür vorbereitete Aufsatz wurde überarbeitet und erschien im ERPANET-Tagungsband der Archivschule Marburg.

Die AG AdUV veranstaltete im Berichtszeitraum drei Round-Table-Gespräche mit Vertretern der bayerischen Justiz, der Polizei und der Landratsämter zu Fragen der Schriftgutverwaltung und Archivierung im elektronischen Umfeld, in deren Rahmen das Projekt ebenfalls vorgestellt wurde.

Am 22. Juni 2004 stellte der Projektleiter im Rahmen eines Workshops zum Projekt ARCHE in Birlinghoven das Projekt mit besonderer Berücksichtigung der Mikrofilmlösung kurz vor.

Am 9. November 2004 fand in München ein Expertengespräch zum Projekt statt. Die 38 Teilnehmer vertraten elf Archive, eine Bibliothek, sieben Verwaltungseinrichtungen, fünf Fachfirmen und drei Forschungseinrichtungen. Nach einem Bericht des Projektleiters über den bisherigen Verlauf des Projekts stellte Dr. Readi seinen Lösungsansatz und erste Versuche zur Umsetzung vor. Ferner wurden offene Fragen wie Speicherung, Erschließung und Benützung diskutiert und vom Fraunhofer-Institut für Physikalische Messtechnik ein Bericht über den Stand des Projekts ARCHE (Ausbelichtung von digitalen Daten auf Mikrofilm mittels eines Laser-Ausbelichters) vorgetragen.

7. Weitere Arbeiten und Planungen

Das Auslaufen der Förderung des Projekts zum 31.12.2005 bedeutet keineswegs sein Ende. Insbesondere die Bereiche Erschließung (Findmittel) und das Umsetzen der Mikrofilmlösung werden als Dienstaufgaben weiter bearbeitet. Der Aufbau einer Findmitteldatenbank und der Zugriff darauf über Webbrowser werden von der IT-Leitstelle bei der Generaldirektion der Staatlichen Archive Bayerns (bis Ende 2006 mit ehrenamtlicher Unterstützung durch Dr. Readi) vorangetrieben. Sobald die Schnittstelle XArchiv verfügbar ist, wird die Anpassung des landesweiten DMS wie auch des im Finanzministerium eingesetzten Produkts angegangen bzw. wieder aufgenommen.