

## Übernahme / Workflows

### Wie kommt zusammen, was zusammengehört?

### Übernahme und Speicherung digitaler Archivalien im Deutschen Bundestag

ANGELA ULLMANN

Die Frage, wie und ob etwas zusammenkommt, ist auf den ersten Blick vielleicht eher marginal, wenn man davon ausgeht, dass sich im Archiv am Ende alles irgendwie findet – Hauptsache, es gelangt überhaupt dorthin. Wir Archivare aber wissen, dass die Übernahme am Anfang einer Reihe von Arbeiten steht, die vonnöten sind, um unsere gesetzlichen und fachlichen Aufgaben zu erfüllen. Im konventionellen Bereich – wie ich die Archivierung analoger Überlieferung hier der Einfachheit halber bezeichnen möchte – hat sich diese Einsicht bereits vor Jahrhunderten durchgesetzt. Bei der digitalen Überlieferung jedoch schauten wir bis vor kurzem gebannt auf die technischen Probleme und konzentrierten uns auf Fragen, die im Rahmen einer Kooperation vielleicht andere Berufsgruppen besser und kompetenter beantworten können. Ich denke da zum Beispiel an die Haltbarkeit von Speichermedien, die Archivfähigkeit von Datenformaten oder die Entwicklung von Datensicherheitskonzepten.<sup>1</sup>

Rein archivfachliche Fragen fanden dagegen bislang weit weniger Berücksichtigung in der Diskussion. Diese sind jedoch nicht trivial. Bilden archivierte Webpräsentationen einen eigenständigen Provenienzbestand bzw. den Teil eines solchen oder Sammlungsgut? Nach welchen Prinzipien werden digitale Archivaliengattungen in einen Bestand eingeordnet? Bilden sie dort so etwas wie eine „Aktengruppe“ oder „Serie“? Wie sieht ihre tektonische Zuordnung aus?

Analoges Archivgut lagert grundsätzlich in der Obhut des verwahrenden Archivs. Im digitalen Bereich ist das nicht unbedingt so – das Archiv muss hier zwar einen Zugriff auf die Archivalien erhalten. Diese können physisch aber außerhalb des Archivs gespeichert sein. Hat das Archiv die Archivalien auch physisch übernommen, existieren wiederum zwei grundsätzliche Möglichkeiten der Speicherung: online oder offline – auf internen Datenträgern oder auf externen. Sicher können diese Aspekte bei der Verzeichnung und Magazinverwaltung berücksichtigt werden. Auch das DOMEA-Konzept<sup>2</sup> hat schließlich Wege für den Umgang mit Hybridakten gefunden. Nur wie ist dies dann im Archiv konzeptionell zu planen und praktisch umzusetzen? Nebenbei bemerkt steht „DOMEA“ zwar für „Dokumentenmanagement und elektronische Archivierung im IT-gestützten Geschäftsgang“, beschäftigt sich aber mit elektronischer Archivierung gar nicht, weil die Regelungen an der „Archivpforte“ enden. Es existieren somit verbindliche Empfehlungen für die öffentlichen Institutionen hinsichtlich der Übergabe von in DMS erzeugten Unterlagen an das zuständige Archiv – für die öffentlichen Archive hingegen gibt es nach wie vor nichts in der Form. Archivare fordern nachdrücklich verbindliche Regeln für den vorarchivischen Bereich und argumentieren mit dem berechtigten Schreckensszenario des drohenden Überlieferungsverlustes. Ich befürchte allerdings, dass dieser auch droht, wenn für die Archive solche Normen fehlen. Etwas globaler angelegte Richtlinien für vertrauenswürdige digitale Archive sind im letzten Jahr im Rahmen von nestor erarbeitet worden.<sup>3</sup> Abzuwarten bleibt, inwieweit diese durch Archive angenommen und umgesetzt werden.

Die nächste drängende Frage stellt sich dann bei der Erschließung und Recherche. Welche Werkzeuge und Rahmenbedingungen benötigen wir, um dem Benutzer analoges und digitales Archivgut über eine integrierte Recherche zugänglich zu machen? Die Zeiten, in denen alle Recherchen immer vom Archivar direkt beantwortet wurden, sind wohl vorbei. Der Archivar wird immer mehr vom unmittelbaren zum mittelbaren Vermittler. Viele Benutzer recherchieren zunächst über das Internet und bestellen die Archivalien, bevor sie dann das Archiv selbst aufsuchen.

All dies und noch einiges mehr verbirgt sich hinter dem Thema meines Vortrages. Für mich bietet er gleichzeitig die Möglichkeit einer Zwischenbilanz nach über fünf Jahren digitaler Überlieferungssicherung im Parlamentsarchiv des Deutschen Bundestages. Damit zwangsläufig verbunden ist die

---

<sup>1</sup> Mittlerweile scheint sich – im gesamten Bereich eGovernment und digitale Archivierung – eher eine prozessorientierte Sicht durchzusetzen, die nicht mehr die Technik in den Vordergrund stellt. Dies klang bei den Anwendervorträgen auf dem 13. Anwenderforum „Effizientes E-Government durch effiziente Prozesse“ am 6. und 7.2.2007 durch. Vgl. hierzu insbesondere Andrea HÄNGER. Der E-Governmentprozess Langzeitarchivierung. Vorstellung und Hintergründe des Pilotprojekts im Bundesarchiv. URL <http://www.anwenderforum2007.de/>.

<sup>2</sup> Vgl. Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung im Bundesministerium des Innern (KBSt). DOMEA-Konzept. URL <http://www.kbst.bund.de>.

<sup>3</sup> Vgl. Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Hrsg. von der nestor-Arbeitsgruppe Vertrauenswürdige Archive-Zertifizierung. Frankfurt am Main 2006. URL <http://nbn-resolving.de/urn:nbn:de:0008-2006060710>.

## – Übernahme/Workflows –

Frage, ob die zunächst völlig isoliert voneinander laufenden Initiativen und Projekte zur digitalen Überlieferungssicherung auf eine sinnvolle Art im Archiv zueinander finden.

Das Parlamentsarchiv hat erst Ende 2001 mit digitaler Überlieferungssicherung begonnen. Zu diesem Zeitpunkt gab es keinerlei Vorarbeiten. Zum Ende des Jahres 2006 – also fünf Jahre später – waren 1.722 digitale Bilddateien archiviert; 21.437 der im vorarchivischen Bereich des Systems „Digitaler Bilderdienst“ befindlichen digitalen Bilddateien erfüllten das Kriterium der Archivfähigkeit. Das Webarchivsystem enthielt 36 historische Webpräsentationen des Deutschen Bundestages. Beide Systeme sind online über das Internetangebot des Deutschen Bundestages verfügbar.<sup>4</sup>

Insgesamt verwahrte das Parlamentsarchiv zum Ende des Jahres 2006 145 GB digitale Bilder und Webpräsentation sowie 4.500 Stunden digitale Videoaufzeichnungen.

Diese Quellensicherungen setzten umfangreiche konzeptionelle Arbeiten voraus wie

- die Definition der Archivfähigkeit digitaler Bilddateien und Netzressourcen,
- die Festlegung eines Workflows für die Archivierung von Bilddateien und Webpräsentationen,
- die Beschreibung und Beschaffung des Systems „Digitaler Bilderdienst / Bildarchiv“,
- die Entwicklung eines Webarchivsystems,
- die Erarbeitung und Umsetzung von Erschließungsstandards für Bilder und Webpräsentationen.

Weitere Konzepte liegen zwar vor, haben aber den Praxistest noch nicht bestanden und auch noch keine Quellensicherung ermöglicht: so bspw. die Anpassung des DOMEA-Konzeptes an die Bundestagsverwaltung und die Entwicklung eines Workflows für die Übernahme elektronischer Akten. Auch bei der Archivierung digitaler Videoaufzeichnungen besteht noch großer Optimierungsbedarf.

Ich möchte im Folgenden die Konzepte und praktischen Verfahren für die Übernahme einzelner digitaler Quellengattungen vorstellen und dann auf die Frage eingehen, wie denn alles zueinander kommen könnte.

Die Archivierung digitaler Kamerabilder war das erste Vorhaben des Parlamentsarchivs zur digitalen Überlieferungssicherung.<sup>5</sup> Hier ist die Praxis am weitesten ausgereift. Bezüglich des Umfangs nimmt diese Quellengattung „nur“ den dritten Platz im digitalen Bereich ein. Bei Berücksichtigung der Qualitätskriterien hat eine Bilddatei im Mittelwert eine Dateigröße von 15 MB. Das digitale Bildarchiv beherbergt damit Dateien in einem Umfang von über 25 GB.

Die Übernahme und Archivierung läuft folgendermaßen ab:

Ein Auftragsfotograf oder ein Fotograf der hauseigenen Foto- und Bildstelle nimmt mit der Digitalkamera ein Motiv auf und berücksichtigt dabei bestimmte Qualitätsstandards. Diese sind für externe Fotografen auch Bestandteil des Vertrages. Die Bilddatei wird mit vom Parlamentsarchiv gelieferten und ständig aktualisierten Wertelisten und unter Anwendung einer ebenfalls vom Archiv gepflegten Beschreibungsrichtlinie mit inhaltlichen und technischen Metadaten versehen. Danach erfolgt die Lieferung an den jeweiligen Auftraggeber, also bspw. das Pressezentrum oder das Referat Öffentlichkeitsarbeit, wo das Bild in den „Bildeingang“ des Systems „Digitaler Bilderdienst / Bildarchiv“<sup>6</sup> gelangt. Der Auftraggeber (oder, wie wir sagen, der „Fachanwender“ bzw. das „Einstellerreferat“) kontrolliert anschließend die Beschreibungsangaben und korrigiert oder vervollständigt diese. Erst dann kann das Bild in den Bilderdienst selbst transferiert werden, ist für alle Fachanwender sichtbar und kann darüber hinaus auch via Intranet oder Internet bereitgestellt werden. Das Parlamentsarchiv hat als einzige Organisationseinheit jederzeit auf alle Bilder Zugriff. Alle Bilder verbleiben mindestens eine Legislaturperiode (also regulär vier Jahre) im Bilderdienst. Erst dann bewertet das Parlamentsarchiv die Bilder und übergibt sie bei Archivwürdigkeit an das digitale Bildarchiv. Alle anderen werden – mit einer automatischen Einspruchsmöglichkeit des Einstellerreferates – gelöscht.

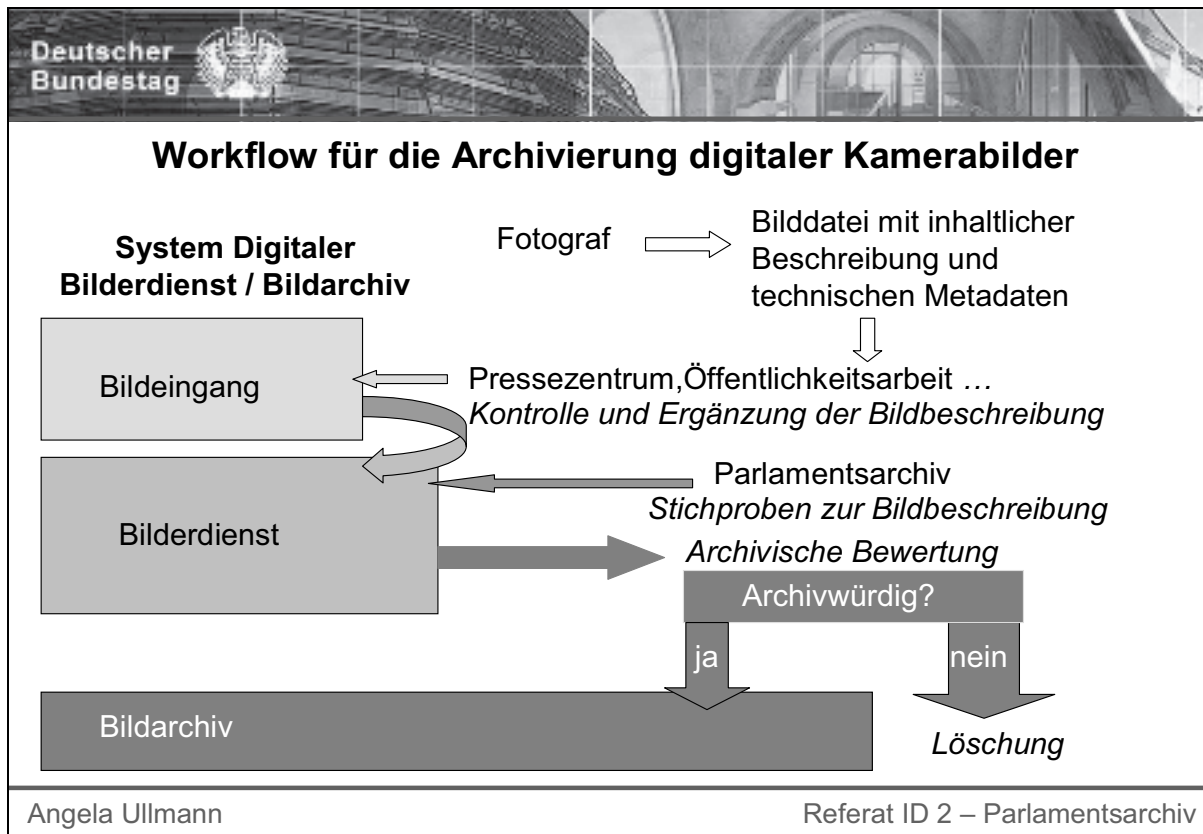
---

<sup>4</sup> Digitaler Bilderdienst online unter URL

<http://bilderdienst.bundestag.de>. Webarchiv online unter URL <http://webarchiv.bundestag.de>.

<sup>5</sup> Vgl. Angela Ullmann. Das Parlament in Pixeln - Digitale Fotos als neue Herausforderung. 2003. URL <http://www.uni-muenster.de/Forum-Bestandserhaltung/downloads/ullmann.pdf>.

<sup>6</sup> Vgl. Angela Ullmann. Das System „Digitaler Bilderdienst / Bildarchiv“ beim Deutschen Bundestag. In: Digitales Verwalten – Digitales Archivieren. Hrsg. von Rainer Hering und Udo Schäfer. (Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg; Bd. 19). Hamburg 2004, S. 131 – 140.

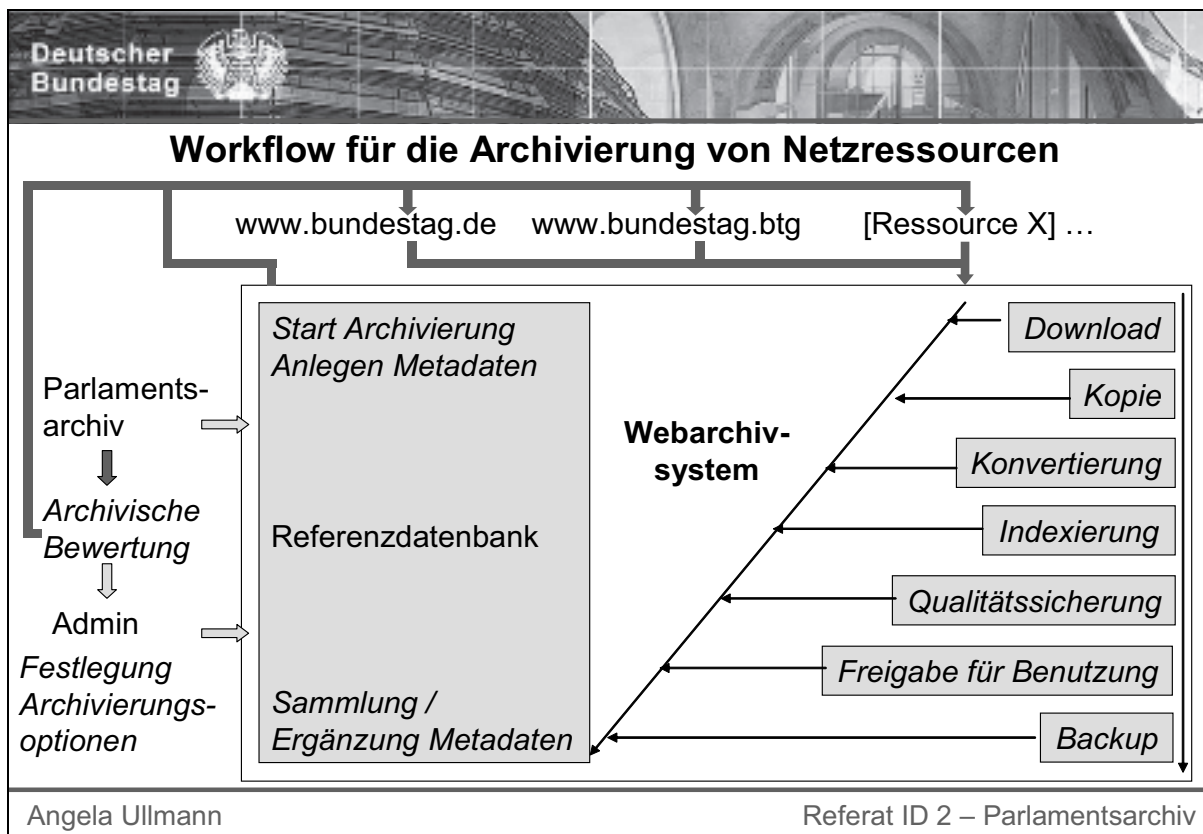


Archivierte Netzressourcen rangieren vom Umfang her im Parlamentsarchiv an zweiter Stelle. Ein Snapshot von [www.bundestag.de](http://www.bundestag.de) benötigt 3,5 GB Speicherplatz. Momentan hat das Webarchiv einen Umfang von ca. 120 GB<sup>7</sup>.

Der Workflow für die Archivierung von Netzressourcen beginnt – wie im konventionellen Bereich – mit der archivfachlichen Bewertung. Die Bewertung schlägt sich in so genannten Archivierungsoptionen nieder.<sup>8</sup> Diese Optionen werden im Webarchivsystem als Sets hinterlegt und bei einem Download berücksichtigt. Der eigentliche Archivierungsvorgang beginnt mit dem Anlegen der Metadaten in der Referenzdatenbank des Webarchivsystems. Nach deren Speicherung greift das System automatisch auf die ausgewählte Domäne zu und kopiert die Netzressource auf den Webarchivserver. Anschließend wird eine Bearbeitungskopie angefertigt, die konvertiert, indiziert und kontrolliert wird. Die dabei anfallenden Metadaten werden weitgehend automatisch in die Referenzdatenbank übernommen. Erst nach der Qualitätssicherung erfolgen dann die Freigabe für eine Benutzung und der Export auf den Webserver sowie die zusätzliche Datensicherung in Form eines Backups.

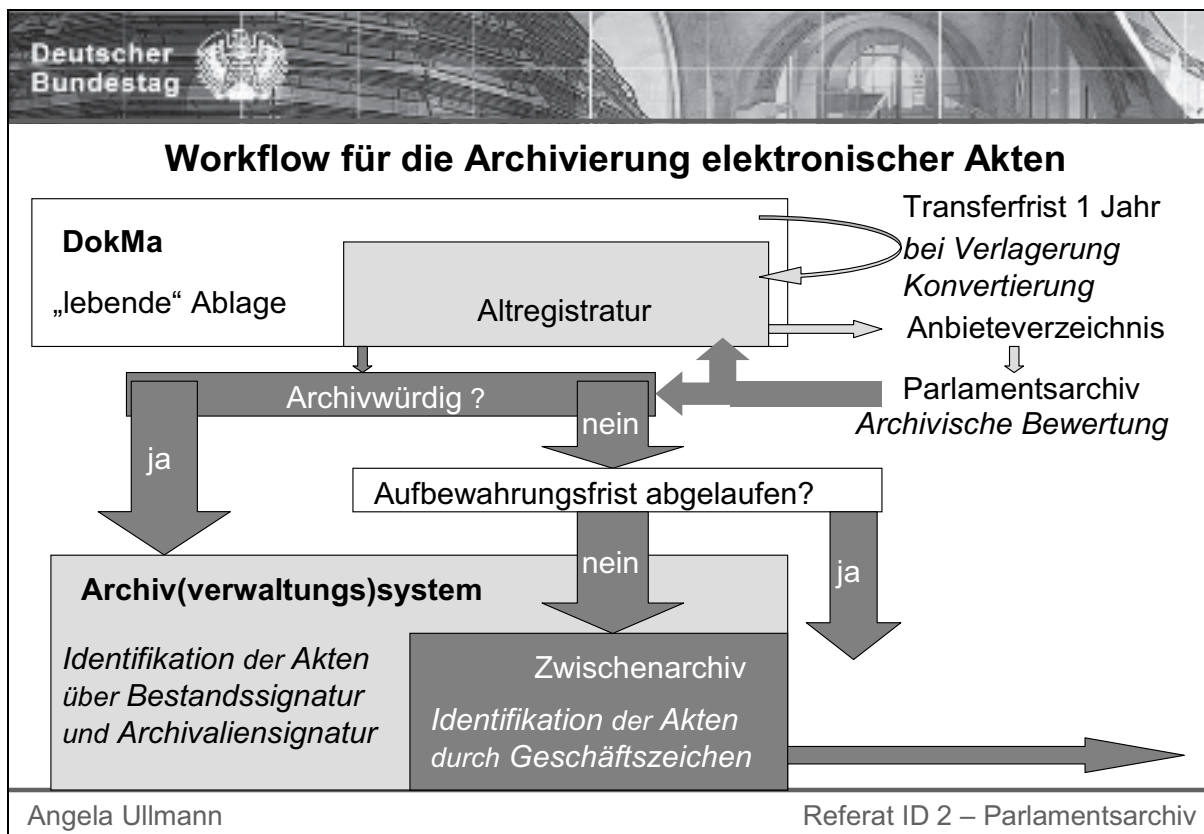
<sup>7</sup> Diese Zahl umfasst nicht die Sicherungskopien.

<sup>8</sup> Vgl. Angela Ullmann und Steven Rösler. Archivierung von Netzressourcen des Deutschen Bundestages. Veröffentlichungen aus dem Parlamentsarchiv des Deutschen Bundestages. Dezember 2005. URL [http://www.bundestag.de/wissen/archiv/oeffent/arch\\_netz\\_klein.pdf](http://www.bundestag.de/wissen/archiv/oeffent/arch_netz_klein.pdf); Angela Ullmann. Heute im Netz – morgen im Archiv. Die Archivierung des Internetangebotes des Deutschen Bundestages. In: Planungen, Projekte, Perspektiven. Zum Stand der Archivierung elektronischer Unterlagen. Hrsg. von Barbara Hoen. Veröffentlichungen des Landesarchivs Nordrhein-Westfalen 10. Düsseldorf 2006, S. 51 - 55.



Das noch nicht realisierte Modell für die Archivierung elektronischer Akten setzt bei der Verlagerung von Akten aus der lebenden Ablage in die Altregistratur ein. Der vorarchivische Bereich ist im DOMEA-Konzept und in den Festlegungen im Rahmen des Projektes „DokMa – Dokumentenmanagement in der Bundestagsverwaltung“<sup>9</sup> hinreichend behandelt und bleibt daher ausgespart. Als Frist für einen Transfer aus der Registratur in die Altregistratur ist ein Jahr nach der zdA-Verfügung vorgesehen. An die Verlagerung gekoppelt ist die Konvertierung der Dateien in ein archivfähiges Format. Nach einer noch nicht definierten Frist wird das Anbieterverzeichnis über die auszusondernden und nicht prospektiv als kassabel bewerteten Akten erzeugt und dem Parlamentsarchiv übermittelt. Das Archiv erhält einen webbasierten Zugriff auf die Altregistratur und trifft dann die endgültige Bewertungsentscheidung. Archivwürdige Akten gelangen direkt in das Archivverwaltungssystem (also in das Endarchiv), unabhängig davon, ob noch gesetzliche Aufbewahrungsbestimmungen laufen oder nicht. Akten, für die eine Archivwürdigkeit verneint wird, die aber noch gesetzlichen Aufbewahrungsfristen unterliegen, werden im Zwischenarchiv des Archivverwaltungssystem aufbewahrt und nach deren Ablauf gelöscht.

<sup>9</sup> Vgl. Angela Ullmann. Aussonderung und Archivierung elektronischer Akten in Verwaltung und Parlamentsarchiv des Deutschen Bundestages. Tagungsbericht zur Informationsveranstaltung des Bundesarchivs am 9. Juni 2005 (CD-ROM). URL <http://www.bundestag.de/wissen/archiv/oeffent/VortragArchivierungelekAkten.pdf>; Thomas Müller und Angela Ullmann. Dann klappts auch im Archiv. Neue Regelungen und Kooperationen auf dem Gebiet der Schriftgutverwaltung und -aussonderung in der Bundestagsverwaltung. In: Der Archivar 59 (2006), H. 3, S. 267 – 269.



Zur digitalen Quellensicherung des Parlamentsarchivs gehören darüber hinaus Videoaufzeichnungen. Denn auch diese sind digitale Quellen: audiovisuelle Aufzeichnungen von Plenardebatten, Sonderveranstaltungen, Öffentlichen Anhörungen und Ausschusssitzungen bzw. Sitzungen anderer Gremien – gespeichert seit September 2000 auf digitale Medien. Videoaufzeichnungen bilden den mit Abstand größten digitalen Archivbestand. Ende des Jahres 2006 verwahrte das Parlamentsarchiv insgesamt 2.400<sup>10</sup> DigitalBetacam-Kassetten<sup>11</sup>. Dabei handelte es zu 90% um Kassetten mit einer Länge von 124 Minuten. Hochgerechnet macht das einen Gesamtumfang von ca. 4.500 Stunden aus. Momentan übernehmen wir ca. 1.000 Stunden Aufzeichnungen im Jahr, darunter allein 520 Stunden Plenardebatten. Die Tendenz ist deutlich steigend, da immer mehr Ausschüsse öffentlich tagen oder Öffentliche Anhörungen veranstalten.

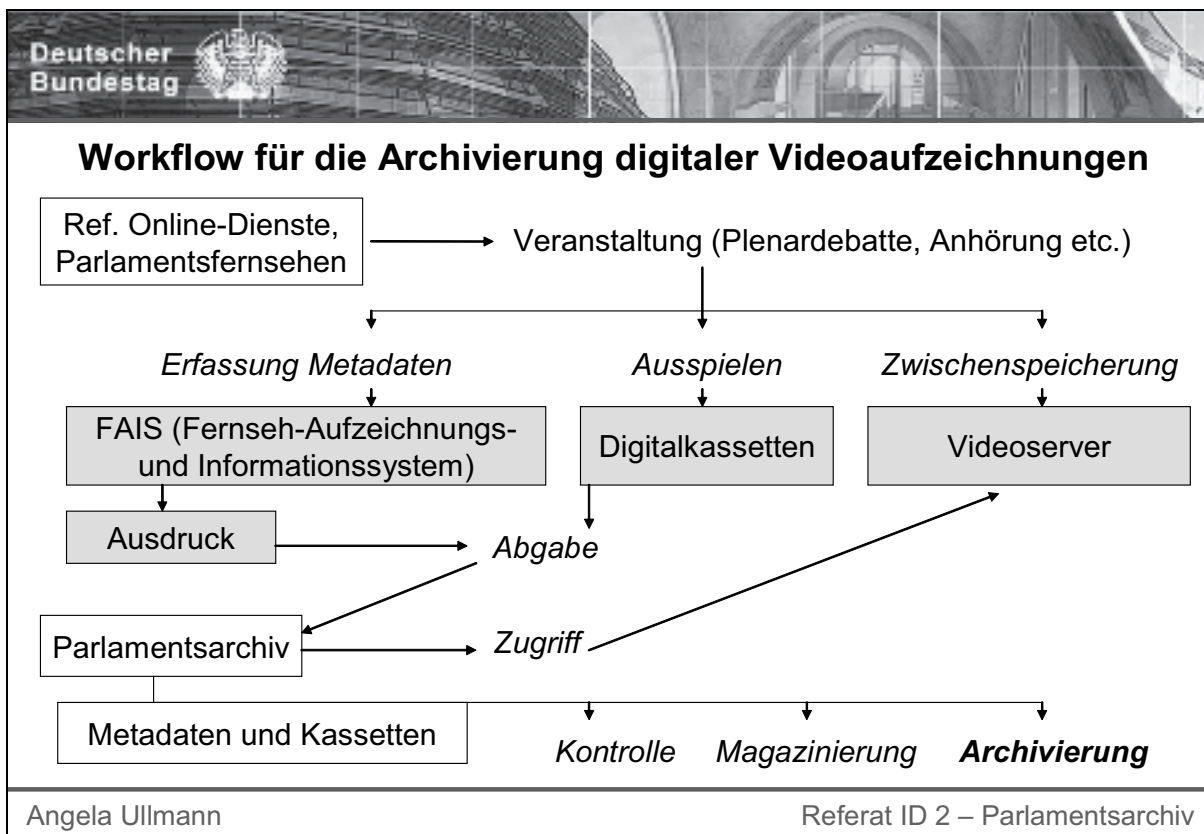
Für den Bereich der Übernahme und Archivierung von Videoaufzeichnungen ist der Workflow noch nicht optimiert worden:

Das Referat „Online-Dienste, Parlamentsfernsehen“ zeichnet die Veranstaltung auf, erfasst die Metadaten im so genannten FAIS (Fernseh-Aufzeichnungs- und Informationssystem)<sup>12</sup>, spielt die Aufzeichnung auf Digitalkassetten aus und speichert sie eine Zeitlang auf dem Videoserver. Das Parlamentsarchiv erhält dann die Digitalkassetten und die dazugehörigen Metadaten in Form von Papiausdrucken, kontrolliert beides anhand der „Stenographischen Berichte über die Verhandlungen des Deutschen Bundestages“, der Protokolle der Ausschusssitzungen o. ä., magaziniert die Kassetten mit den Aufzeichnungen und archiviert sie.

<sup>10</sup> Diese Zahl umfasst nur die „Erstaufzeichnungen“ (Originale) und nicht die Sicherungskopien.

<sup>11</sup> „Digital Betacam (auch Digibeta genannt) ist das meistgenutzte Speichermedium in der heutigen Fernsehtechnik.“ URL [http://de.wikipedia.org/wiki/Digital\\_Betacam](http://de.wikipedia.org/wiki/Digital_Betacam).

<sup>12</sup> Online-Recherche-Version unter URL <http://archiv.bundestag.t-bn.de/Archiv/servlets/Rede/Search>.



In absehbarer Zeit werden hier jedoch Veränderungen erfolgen; wir beschäftigen uns bereits mit den konzeptionellen Vorarbeiten. Denkbar wäre die direkte Kontrolle und ggf. Berichtigung der Metadaten im FAIS und die Ergänzung der Archivsignatur oder die digitale Übernahme der Metadaten über eine Schnittstelle in ein Archivverwaltungssystem. Das Parlamentsarchiv könnte künftig eventuell das Ausspielen vom Videoservert auf Digitalkassetten vornehmen. Aber alles dies muss vor dem Hintergrund der gesamten Rahmenbedingungen noch sorgfältig geprüft werden. Für die auch in der Bundestagsverwaltung unterhaltenen Datenbanken existieren noch keine Konzepte. Dieser Bereich bleibt bislang unberücksichtigt. In dem aus archivischer Sicht wichtigsten System „DIP – Dokumentations- und Informationssystem für parlamentarische Vorgänge“<sup>13</sup> erfolgen immerhin keine Datenlösungen.

<sup>13</sup> DIP wird vom Referat „Sach- und Sprechregister“ der Bundestagsverwaltung betreut. Es „wertet die öffentlich zugänglichen Materialien des Deutschen Bundestages und des Bundesrates (Plenardebatten, Gesetze, Anträge, Anfragen, Berichte) nach formalen und inhaltlichen Kriterien aus.“ URL [http://www.bundestag.de/bic/sach\\_sprech/index.html](http://www.bundestag.de/bic/sach_sprech/index.html).

– Übernahme/Workflows –

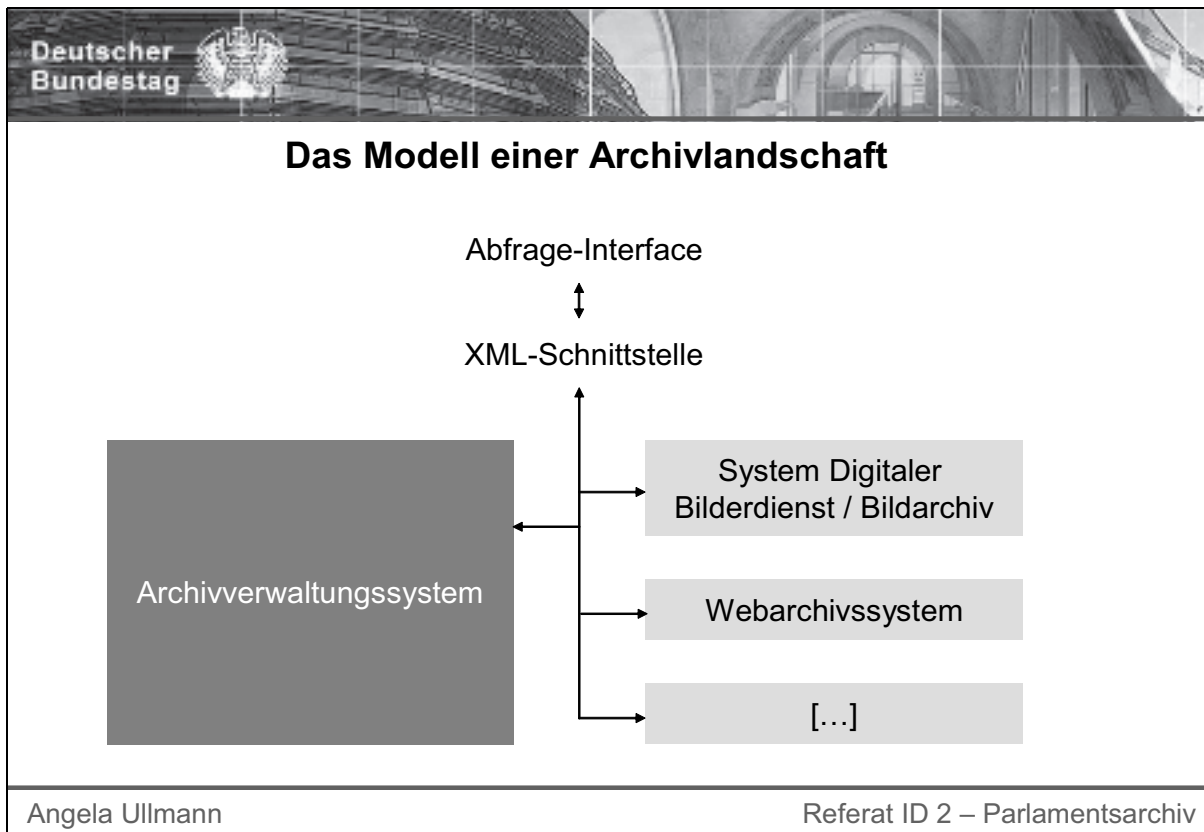
Nach der Archivierung befinden sich alle vier Archivaliengattungen physisch und logisch in völlig unterschiedlichen Bereichen:

	Bilder	Akten (in Planung)	Netz- ressourcen	Video- aufzeichnungen
Umfang per 31.12.2006	25 GB	0	120 GB	4.500 Stunden
Logischer Verbleib (Metadaten)	„Bildarchiv“ im System „Digitaler Bilderdienst / Bildarchiv“	Archivver- waltungs- system	Webarchiv- system	analoge Findmittel; geplant: Archivverwaltungs- system
Datenträger	Server			extern (Digital Betacam)
Standort / Verbleib Datenträger	Rechenzentrum DBT (Referat IT 1)		Online- Dienste / Parlaments- fernsehen (Referat PuK 4)	Parlamentsarchiv (Referat ID 2)
Backup-/ Sicherungs- medium	??		DLT	LTO
Verbleib Backup-/ Sicherungs- medium	Rechenzentrum DBT (Referat IT 1)		Parlamentsarchiv (Referat ID 2)	

Somit sind die Quellen zwar archiviert; auch eine Sicherungskopie existiert in jedem Fall. Wie aber kann ich jetzt eine thematische Recherche in allen Archivaliengattungen ermöglichen?

Diese Frage war der Ausgangspunkt für die Entwicklung des Modells einer Archivlandschaft, die alle einschlägigen Systeme miteinander verbindet. Ein Abfrage-Interface soll dabei die Anfrage über eine XML-Schnittstelle an alle einschlägigen Systeme weitergeben und wiederum die Treffer „einsammeln“ und ausgeben.<sup>14</sup>

<sup>14</sup> Mit einem ähnlichen Modell arbeiten im Übrigen die ZDF-Archive, die ebenfalls eine Fülle verschiedener Medien zu verwalten haben. Auch hier können alle einschlägigen Datenbanken über ein Archivportal angesteuert werden. Vgl. Veit Scheller. Die ZDF-Archive – Nur Dienstleister für interne Nutzer? In: Archiv und Wirtschaft 39 (2006), H. 4, S. 175 – 183.



Herzstück dieser Archivlandschaft ist ein offenes, modular aufgebautes und zukunftsfähiges Archivverwaltungssystem, das die Metadaten zu einzelnen digitalen und zu sämtlichen analogen Archivalien mit Ausnahme der Photographien<sup>15</sup> und perspektivisch auch digitale Archivaliengattungen wie elektronische Drucksachen oder Akten aufnimmt. Über eine Archivdatenbank verfügt das Parlamentsarchiv zurzeit noch nicht. Ende des letzten Jahres hat aber ein IT-Projekt begonnen, das die Beschaffung eines Archivverwaltungssystems zum Ziel hat. Als letzte Ausbaustufe strebt das Archiv die Realisierung der Archivlandschaft an.

Zuvor müssen jedoch noch erhebliche Vorarbeiten geleistet werden. Allein die Harmonisierung der unterschiedlichen Erschließungsstandards für die einzelnen Archivaliengattungen, die wiederum in den Systemen hinterlegt sind, birgt große Herausforderungen in sich.

Kommt auf diesem Wege zusammen, was zusammen gehört? Ich vermag es nicht zu sagen und kann auch ein Scheitern in dieser Frage nicht ausschließen. An dem Versuch aber führt kein Weg vorbei. Denn wir Archivare weisen mit Recht darauf hin, dass die überlieferten Quellen miteinander korrespondieren, einen Kontext bilden und wir diesen im Archiv sichern.

Kontext ist aber auch die Beziehung, in der die analoge und digitale Überlieferung zueinander stehen. Und dass ein Zusammenwachsen dessen, was zueinander gehört, eine ungeheure Herausforderung ist, wissen wir mittlerweile alle.

<sup>15</sup> Diese werden im System „Digitaler Bilderdienst / Bildarchiv“ erfasst.