

Das Repräsentationenmodell des Landesarchivs Baden-Württemberg

Christian Keitel

Um an der Informationsgesellschaft teilhaben zu können, müssen die Archive die von ihnen verwalteten Objekte digital nachweisen. Benutzer können die Archivalien dann schnell auffinden, Archivare können sie zielgerichtet erhalten. Die digitale Erfassung der Objekte bringt daher sowohl in der Benutzung als auch in der Bestandserhaltung erhebliche Vorteile. In dem Beitrag wird das vom Landesarchiv Baden-Württemberg entwickelte Repräsentationenmodell vorgestellt, das die Basis für ein solches einheitliches Nachweissystem bilden kann.¹

Ein einheitliches Nachweissystem

Mit der digitalen Erschließung und der Retrokonversion älterer Findmittel haben viele Archive bereits wesentliche Schritte hin zu dem gerade beschriebenen Ziel gemacht. Der digitale Nachweis konventioneller Archivalien kann daher als gut geplant und teilweise schon umgesetzt gelten. Wie aber sieht es mit den anderen Objekten aus, die die Archive verwalten? In den digitalen Findbüchern wird ein vorhandener Mikrofilm manchmal in den Bemerkungsfeldern des Bestandes oder der Archivalien vermerkt. Verwaltet werden Mikrofilme jedoch in aller Regel in einer eigenständigen Datenbank. Manchmal fehlen auch diese schriftlichen Aufzeichnungen und nur der Archivar selbst weiß von der Existenz des Mikrofilms. Digitalisate werden oft ähnlich wie Mikrofilme behandelt. In beiden Fällen müssen die Informationen zu den einzelnen Erscheinungsformen eines Archivals erst mühsam aus verschiedenen Quellen zusammengetragen werden. Es fehlt ein übergreifender Ansatz, der diese Metadaten gemeinsam verwaltet. Ähnlich sieht es bei den digitalen Archivalien aus, die zudem in verschiedene Objekte zerfallen. Auch zu ihnen haben die Archivarinnen und Archivare noch kein einheitlich anerkanntes Nachweisverfahren entwickeln können.

Sollen Informationen zu diesen unterschiedlichen Objekten überhaupt in einem einheitlichen System nachgewiesen werden? Zunächst ermöglicht ein solches System, alle Archivalien und ihre Erscheinungsformen im Zusammenhang ihrer Provenienz darzustellen. Zweitens werden in den nächsten Jahrzehnten immer mehr Übernahmen, Bestände und Archivalien hybrider Natur sein. Sie besitzen also zugleich einen

¹ Das Repräsentationenmodell des Landesarchivs Baden-Württemberg wurde zunächst im Projekt „Konzeption für ein digitales Landesarchiv“ auf der Basis von PREMIS (Preservation Metadata Maintenance Activity, <http://www.loc.gov/standards/premis/>, Abruf 8.4.2010) konzipiert und dann in einer Arbeitsgruppe detailliert ausgearbeitet. Beteiligt waren die Kollegen Dr. Thomas Fricke, Dr. Thomas Fritz, Udo Herkert, Rolf Lang, Dr. Kai Naumann, Dr. Clemens Rehm und der Verfasser.

Anteil in Papier und einen digital gespeicherten Anteil. Solche Einheiten sollten nicht zerrissen werden, sie können nur dann zusammen beschrieben werden, wenn ein einheitliches Nachweissystem eingerichtet wird. Drittens wäre es für Benutzer ein großer Vorteil, wenn sie bereits von zuhause einschätzen könnten, ob sie verhältnismäßig einfach eine Reproduktion auf Mikrofilm oder ein Digitalisat bekommen können. Dies gilt vor allem für solche potentielle Benutzer, für die ein Besuch des Archivs sehr aufwendig wäre (z.B. Historiker in den USA). Viertens wäre es auch aus Gründen der Bestandserhaltung sinnvoll, wenn die verschiedenen Erscheinungsformen eines Archivals auf einen Blick überschaubar wären. Beispielsweise würde dies die Verwaltung der verschiedenen Reproduktionsformen oder die Erstellung von Benutzerkopien erleichtern.

Ein einheitliches, materialunabhängiges Nachweissystem für alle Erscheinungsformen von konventionellen und digitalen Archivalien ist also aus verschiedenen Gründen erstrebenswert. Mit den Beständeübersichten und den Findbüchern liegt bereits ein zentrales System vor. Die Frage ist also, ob und wenn ja, wie dieses System entsprechend ausgebaut werden kann.

Entitäten digitaler Archivalien

Beginnen wir mit dem Nachweis digitaler Archivalien. Was verbirgt sich hinter diesem Terminus? Welche Rolle spielen dabei Dateien, Bitströme, Dateiformate, Zeichenformate, Informationen, Datenträger und Metadaten? Diese Einheiten und einige mehr haben Eingang in diverse Standards zur Archivierung digitaler Unterlagen gefunden. Sie alle beanspruchen somit ihren Platz in dem anzustrebenden Nachweissystem. Wie können diese Einheiten gruppiert werden? Nach welchen Kriterien sollten sie gruppiert werden?

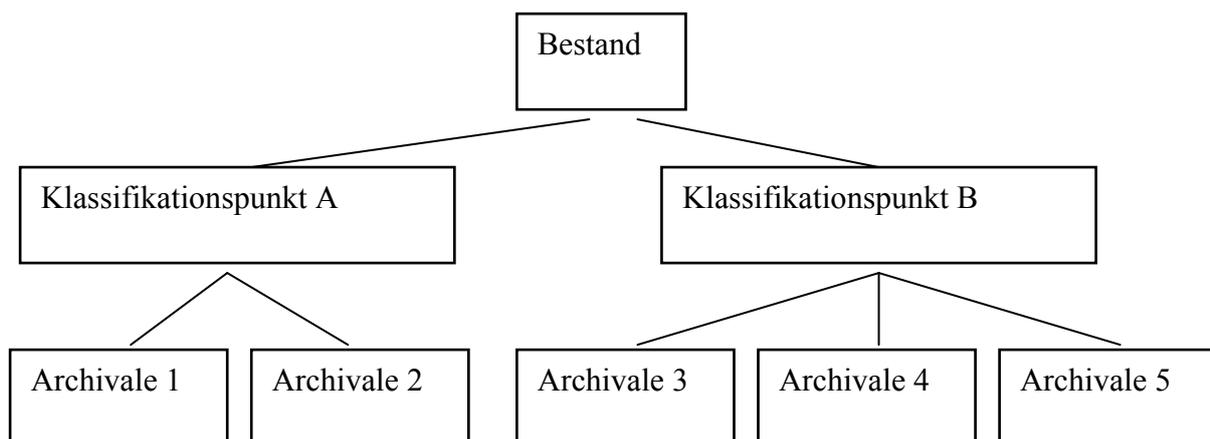
Digitale Archivierung besteht im Versuch, digital gespeicherte Informationen dauerhaft zu erhalten. Informationen sind das eigentliche Ziel der Archivierung; sie werden durch Ausgabegeräte wie Monitore oder Lautsprecher dargestellt und von Menschen interpretiert. Informationen selbst sind nicht greifbar, sie müssen in Form digitaler Daten gespeichert werden. Gerade die Daten sind nun eng an heutige Computersysteme gekoppelt. Erst durch die Mitwirkung von Hard- und Software werden sie schließlich so ausgegeben, dass ein Mensch sie als Informationen wahrnehmen kann.² Mit der kontinuierlichen Weiterentwicklung der Computersysteme müssen daher entweder die archivierten Daten verändert werden (Migration) oder künftige Computersysteme über Spezialprogramme (Emulatoren) in die Lage versetzt werden, die archivierten Daten noch verarbeiten und anzeigen zu können. Bei diesen

² Dieser Prozess ist einschlägig beschrieben im *Performance Model* des Australischen Nationalarchivs: Helen HESLOP, Simon DAVIS, Andrew WILSON, *An Approach to the Preservation of Digital Records*, 2002, http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf (Abruf 8.4.2010).

Prozessen ändern sich zwar die Daten, die ausgegebenen Informationen sollten aber unverändert bleiben. Die Unterscheidung zwischen dauerhaft zu erhaltenden Entitäten und solchen, die nur vorübergehend benötigt werden, ist daher fundamental für die weitere Gruppierung der einzelnen Einheiten. Dauerhaft sind demnach die Informationen selbst und ein Teil der Metadaten. Dagegen sind Dateien, deren Bitströme, Dateiformate, Zeichenformate, Datenträger und auch die für diese Elemente notwendigen Metadaten vor allem von kurzfristigem Interesse. Dennoch ist es notwendig, wenigstens die wichtigsten dieser Elemente dauerhaft nachzuweisen, um den Prozess der Bestandserhaltung nachvollziehbar zu halten.

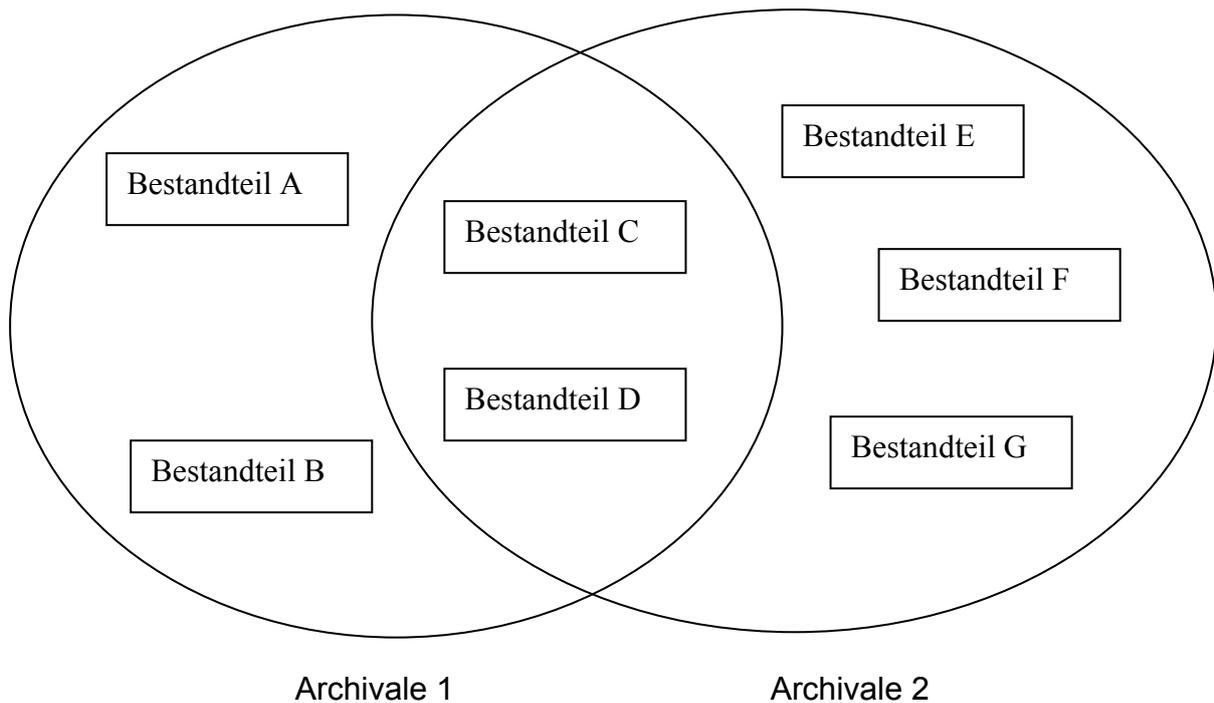
Hierarchische und relationale Nachweissysteme

Wie können die eher kurzlebigen Elemente nun in Beziehung gesetzt werden zu den sehr lange aufzubewahrenden Informationen und Metadaten? Wie also sollen diese Elemente in ein Nachweissystem eingebracht werden? Im Grunde gibt es zwei unterschiedliche Optionen, diese Frage zu lösen. Auf der einen Seite können die Angaben zu den kurzlebigen physischen Elementen in einem hierarchischen System den einzelnen Informationen, die durch sie repräsentiert werden, unmittelbar zugeordnet werden. Eine Information hat dann eine Vielzahl von datentechnischen Kindern. Archivische Beständeübersichten und Findbücher sind nach diesem Prinzip aufgebaut: Unterhalb eines Klassifikationspunkts befindet sich eine beliebige Menge zugehöriger Archivalien. Die Abbildungsregeln von ISAD(G) beschreiben ein solches (mono)hierarchisches System exemplarisch:



Hierarchisches System: Jedes Element kann nur einem übergeordneten Element zugeordnet werden.

Dagegen könnten die verschiedenen Einheiten auch in ein weitaus flexibleres relationales System eingeordnet werden. Die einzelnen Entitäten können dann in beliebige, immer wieder neu zusammenstellbare, nicht zwingende hierarchisch angeordnete, Zusammenhänge gestellt werden.



Relationales Modell: Es besteht keine feste Hierarchie, einzelne Elemente können in verschiedene Zusammenhänge gestellt werden; z.B. können sie von verschiedenen Archivalien nachgewiesen werden.

Welches der Modelle bietet sich nun für den Nachweis digitaler Archivalien am ehesten an? Auf der einen Seite steht die große Flexibilität des relationalen Modells. Jede Information wird nur einmal abgespeichert. Sie kann daher auch leicht zentral geändert werden. Auf der anderen Seite macht gerade die Beweglichkeit dieses Systems dessen langfristige Erhaltung besonders komplex und damit aufwendig. Digitale Archivalien sollen nicht nur für fünf oder zehn Jahre beschrieben und erhalten werden. Die Beschreibungen müssen über sehr lange Zeit in den unterschiedlichsten Computersystemen gut funktionieren. Der Übergang von einer technischen Plattform auf die andere darf also nicht zu viele Ressourcen verschlingen.

Bei diesen Überlegungen sind auch die Zusammenhänge mit den zur Verfügung stehenden Speicheroptionen zu berücksichtigen. Auch im Speicherbereich gibt es bei einer Speicherung auf Festplatten (RAID-Systemen) zwei Möglichkeiten, die wiederum die Alternative zwischen einer eher statischen und einer flexiblen Lösung aufzeigen. Statisch ist der Ansatz, die einzelnen Entitäten in einem hierarchisch organisierten Dateisystem zu speichern. Jede im Dateimanager angezeigte Ordner- bzw. Verzeichnis-Ebene hat genau ein übergeordnetes Verzeichnis und n untergeordnete Verzeichnisse oder Dateien. Flexibel wäre dagegen der Ansatz, die einzelnen Entitäten in einer relationalen Datenbank zu speichern. Diese Speicheroptionen erlauben nun die oben beschriebenen Nachweissystemen nur teilweise. Ein hierarchisches Nachweissystem kann in beiden Speicherformen abgebildet werden. Dagegen kann ein relationales Nachweissystem nicht angemessen in einem hierarchischen Speichersystem abgebildet werden. Jede kleine Änderung im Nachweissystem führt zu einer viel umfassenderen Änderung im Speichersystem. Die mit einem relationalen

Nachweissystem verbundene Stärke der Flexibilität wäre daher bei dieser Lösung von vornherein stark eingeschränkt.

Aufgrund der Komplexität und des damit verbundenen Aufwands erscheint es ratsam, auf ein zwar unflexibles, aber stabiles hierarchisches Nachweissystem zu setzen. Hierfür spricht auch die Nähe zu dem für die digitale Archivierung empfehlenswerten hierarchischen Speichersystem. Dieses ist erstens wegen seiner höheren Stabilität und zweitens wegen geringeren Abhängigkeiten (Datenbank³) zu bevorzugen. Nicht zuletzt wird dadurch die Möglichkeit erhalten, das System zum Nachweis digitaler Archivalien mit den bereits bestehenden Systemen zum Nachweis konventioneller Archivalien, also den Beständeübersichten und Findbüchern, zu vereinen, da diese ebenfalls hierarchisch strukturiert sind.

Komplexität erhöht aber auch in anderer Hinsicht den Aufwand. Dieser steigt mit der Zahl der zu verwaltenden Einheiten und der Zahl der verwendbaren Nachweisvarianten. Für den Nachweis digitaler Archivalien folgt daraus, dass so wenige Einheiten wie möglich Verwendung finden sollten. Die Nachweise sollten nach einem einheitlichen Muster erfolgen. Auch die für die digitalen Archivalien zu vergebenden Namen (Signaturen und *Identifier*) sollten von möglichst geringer Zahl sein.

Archivische Identifier (AID)

Nur wenn die Einzelteile eindeutig benennbar sind, sind sie aus Archivsicht auch handhabbar. Nur dann kann eine digitale Bestandserhaltung überhaupt durchgeführt werden. Diese Eindeutigkeit sollte schließlich für die gesamte Lebenszeit der digitalen Archivalien bestehen bleiben, d.h. die Namen sollen dauerhaft stets auf dieselben Bestandteile verweisen.

Wie könnte nun ein solcher Name für die zur digitalen Archivierung erforderlichen Entitäten gebildet werden? Klassische Signaturen eignen sich für diesen Zweck nur eingeschränkt, da Archivalien auch entgegen anderslautenden Absichten manchmal umsigniert werden. Auf der anderen Seite werden Archivalien seit langem von Menschen über Signaturen zitiert und es ist auch nicht abzusehen, dass dieses System sich in nächster Zeit ändern würde. Es bedarf daher eines zweiten Systems, das vor allem den bei der digitalen Archivierung eingesetzten Maschinen den Austausch eindeutig identifizierter Einheiten erlaubt. Bibliotheken sprechen von einem System der *Persistent Identifier*. Dieser bibliothekarische Ansatz ist aber schon deshalb nicht eins zu eins auf Archive zu übertragen, da er Bibliotheksobjekte in verschiedenen

³ Bei einer Migration auf einen anderen Datenträger müssen archivierte Daten aus der Datenbank exportiert werden. Dies ist in der Regel wesentlich langsamer als eine Migration von Dateien aus einem Dateisystem heraus. Zudem ist unklar, ob der Datenbankhersteller zum Zeitpunkt der Migration noch existiert. Andernfalls könnte es schon aufgrund fehlenden technischen Supports schwierig werden, archivierte Daten aus der Datenbank zu exportieren. Dieses Risiko bewegt die meisten digitalen Archive dazu, ihre Daten in hierarchisch organisierten Dateisystemen zu archivieren.

Einrichtungen gleichzeitig nachzuweisen sucht.⁴ Archivalien sind dagegen in aller Regel Unikate, die nur in einem Archiv liegen. Das Landesarchiv Baden-Württemberg hat daher ein System archivischer Identifikatoren, kurz AID, entwickelt. Bei der Einstellung eines neuen Datensatzes im Verzeichnungssystem vergibt dieses eine eindeutige Nummer, die in ein eigenes Metadatenfeld geschrieben wird. Die Nummer ist innerhalb des vergebenden Systems einmalig. Die Nummer wird ergänzt durch eine vorangestellte eindeutige Nummer, die für das vergebende System steht (z.B. 5-9676). Das System kann analog zur Vergabe der Telefonnummern begriffen werden. „5“ entspricht einer Ortsvorwahl, „9676“ ist innerhalb des Ortes nur einmal vergeben. Das System benötigt keine zentrale Stelle innerhalb der IT-Landschaft des Landesarchivs, welche die AIDs vergibt. Stattdessen ist es um beliebige Systeme erweiterbar. Wenn der Metadatenatz an ein anderes System (z.B. vom jetzigen Verzeichnungssystem an das Archivierungssystem oder später in das nachfolgende Verzeichnungssystem) übergeben wird, verändern sich AID und Signatur nicht, da sie ja innerhalb des Datensatzes gehalten werden. Primärdaten können über das System ebenfalls ausgezeichnet werden, indem die AID in den Dateinamen geschrieben wird. Dadurch ist keine Veränderung der Primärdatei notwendig. Diese Lösung ist auch bei jenen Dateiformaten möglich, die keinen Dateiheder vorsehen.⁵

Dateimodell und Repräsentationenmodell

Als Zwischenergebnis kann festgehalten werden: Der Zusammenhang zwischen den einzelnen, bei der digitalen Archivierung nachzuweisenden Einheiten soll ein hierarchischer sein, die Einheiten selbst lassen sich über AIDs eindeutig bezeichnen. Welche Einheiten sind aber überhaupt nachzuweisen? Für die langfristige Erhaltung digitaler Archivalien werden zwei Modelle diskutiert, die hier der Einfachheit halber als Dateimodell und Repräsentationenmodell bezeichnet werden. Beide Modelle haben drei Entitäten gemeinsam (Informationsobjekt, Datei, Bitstrom), das Repräsentationenmodell kennt außerdem noch die Repräsentation⁶ als vierte Entität.

Beide Modelle gehen zunächst davon aus, dass als Ziel der Archivierung Informationen in abgegrenzter Form (d.h. als Informationsobjekt) erhalten werden sollen. Im Dateimodell verweist diese Information nun unmittelbar auf eine beliebige Zahl von Dateien, die ihrerseits aus einem oder mehreren Bitströmen bestehen können.⁷

⁴ Vgl. Nikolaus BÜTIKOFER, Kriterienkatalog zur Überprüfung der Vertrauenswürdigkeit von PI-Systemen, <http://nbn-resolving.de/urn:nbn:de:0008-20080710140> (Abruf 8.4.2010).

⁵ Bei Dateiformaten mit Dateiheder wäre auch denkbar, dort die AID einzuschreiben.

⁶ Der Begriff der Repräsentation wurde vom Metadatenstandard PREMIS (vgl. Anm. 1) eingeführt.

⁷ Eine Datei mit mehreren Bitströmen wäre beispielsweise eine Multipage-Tiffdatei. Sie enthält dabei verschiedene Bilder, die jeweils durch einen eigenen Bitstrom repräsentiert wird. Die Zuordnung der Bitströme zu einer Datei ist bei beiden Modellen identisch, allerdings wird sie in Nachweissystemen oft nicht beschrieben.

Beim Repräsentationenmodell wird ein Informationsobjekt zunächst durch eine Repräsentation abgebildet. Diese verweist auf die Dateien, die wiederum auf die jeweiligen Bitströme zeigen. Die Repräsentation enthält physische Objekte (Dateien). Sie selbst ist jedoch ein gedachter (logischer) Container um diese Dateien.

Für die erstmalige Beschreibung eines digitalen Archivals eignen sich beide Modelle gleichermaßen. Die Unterschiede treten erst zu Tage, wenn wir die Beschreibung des Archivals über einen sehr langen Zeitraum betrachten. Was geschieht also, wenn einzelne Dateien von künftigen Computern aufgrund ihres aussterbenden Dateiformats kaum mehr gelesen werden können? Gemäß der Migrationsstrategie werden sie nun in ein aktuelles Dateiformat migriert, d.h. es wird eine neue Datei mit demselben Inhalt in einem anderen Dateiformat erstellt. Wie gehen nun die beiden Modelle mit dieser Situation um?

Beim Dateimodell wird mit der Einstellung ins Archivierungssystem ein Archivierungspaket gebildet.⁸ Dieses besteht aus den Dateien mit ihren Bitströmen sowie aus den zugehörigen Metadaten. Wenn nun die Datei A in eine Datei A' migriert wird, kommt die Datei A' zum bestehenden Archivierungspaket hinzu, dieses wird also erweitert. Beim Repräsentationenmodell sind die Repräsentation und das Archivierungspaket gleichgesetzt. Die Repräsentation als solche ist nicht erweiterbar. Bei einer anstehenden Dateimigration muss daher eine neue, zweite Repräsentation gebildet werden. Sie unterscheidet sich von der ersten Repräsentation dadurch, dass sie anstelle der Datei A die Datei A' enthält.

PREMIS erlaubt als maßgeblicher Standard zur Beschreibung digitaler Archivalien beide Modelle. Eine Datei kann also in manchen Fällen unmittelbar einem Informationsobjekt zugeordnet sein, während sie in anderen Fällen zunächst einer Repräsentation und diese ihrerseits einem Informationsobjekt zugeordnet ist. Diese Flexibilität kann auf die lange Sicht von einer Einrichtung kaum durchgehalten werden. Zu groß ist die Gefahr von Verwechslungen und daraus resultierenden Datenverlusten. Wie oben ausgeführt empfiehlt es sich daher aus Gründen der langfristigen Stabilität, sich für eines der beiden Modelle zu entscheiden.

Auf den ersten Blick erscheint das Dateimodell flexibler. Es muss nicht für jede Dateimigration eine neue Repräsentation gebildet werden. Dennoch gilt es, einige Fragen zu klären:

- Wie können Primärdaten-Dateien aufgespalten oder zusammengefasst werden? PREMIS nennt das Beispiel eines eingescannten Buches. 189 Seiten werden zunächst durch 189 Scans dargestellt. Mit dem Dateimodell muss stets eine Datei in eine Nachfolgedatei migriert werden. Die Zahl der Dateien

⁸ Zum Archivierungspaket vgl. den Standard Open Archival Information System (OAIS), <http://public.ccsds.org/publications/archive/650x0b1.pdf> (Abruf 8.4.2010). Im engeren Sinne handelt es sich um eine *Archival Information Unit*, also die unterste Ebene eines Archivierungspakets. Eine *Archival Information Unit* kann Bestandteil anderer Archivierungspakete sein. Die folgenden Ausführungen bewegen sich zunächst der Einfachheit halber nur auf der konzeptionellen Ebene. Auswirkungen für die Speichermöglichkeiten werden weiter unten diskutiert.

bleibt daher stets dieselbe. Es ist aber nicht möglich, nach einer Migration von 189 Ausgangsdateien in eine Nachfolgedatei (z.B. im Format PDF/A) beide Erscheinungsformen modellkonform abzubilden.⁹ Ein zweites Beispiel: 2002 übernahm das Landesarchiv Baden-Württemberg die Volkszählung 1970 in vier Dateien (eine pro Regierungsbezirk). Die Dateien enthielten Datensätze in zwei unterschiedlichen Datensatztypen. Zur Erleichterung der späteren Benutzung wurde jede Datei jeweils in zwei Dateien migriert (je eine pro Datensatztyp). Zugleich fand eine Migration vom Festbreitenformat ins CSV-Format statt. Übernommen wurden also vier Dateien, archiviert wurden acht Dateien (Metadaten nicht eingerechnet). Eine solche Migration ist mit dem Dateimodell schlechterdings nicht abbildbar. Dagegen ist es für das Repräsentationenmodell kein Problem, die Zahl der enthaltenen Dateien zu verändern.¹⁰

- Wie kann eindeutig zitiert werden? Es genügt nicht, nur das Informationsobjekt zu benennen, denn dieses besteht zu verschiedenen Zeitpunkten aus unterschiedlichen Dateien. Eine genaue Quellenangabe müsste beim Dateimodell etwa wie folgt lauten: „EL 413/6 DO 1 mit den Dateien A“, B, C“, D“, E“ und den Metadaten vom Stand 24.5.2035“. Eine derartige Benennung ist zu schwerfällig, wenn sie aber gekürzt wird, ist sie nicht mehr eindeutig.
- Wie kann das Archivierungspaket erweitert werden? Beim Dateimodell erscheinen drei Möglichkeiten denkbar. Zunächst kann das Archivierungspaket als logische Einheit verstanden werden, deren Bestandteile (Dateien) an ganz unterschiedlichen Orten liegen. Eine neue Datei kann sehr leicht in die logische Beschreibung dieses Pakets aufgenommen werden. Diese Lösung ist einerseits unproblematisch. Auf der anderen Seite erscheint den meisten Archiven aus Gründen der Stabilität sinnvoll, wenigstens den Kern ihrer Archivierungspakete doch als physische Einheit zu verstehen. In diesem Fall wird das Archivierungspaket in der Datei eines Containerformats (z.B. Zip, Tar) gespeichert. Ein solches Paket kann nun zwar erweitert werden, allerdings besteht dann die Gefahr der Instabilität. Wird dagegen das gesamte Paket neu abgespeichert, dann fragt es sich, warum dieses nicht als neue Repräsentation angesprochen und vor allem nachgewiesen werden soll.
- Wie sind die Metadaten bei der Migration einer Datei anzupassen? Gewöhnlich wird eine Primärdaten-Datei durch Metadaten beschrieben. Wenn nun die Datei der Primärdaten migriert wird, muss auch für die neue Datei eine eigene Metadatendatei erstellt werden. Solange nicht alte Dateien gelöscht werden sollen, wächst somit die Zahl der Dateien, die für ein digitales Archivale ste-

⁹ Von dieser Grundannahme ging noch das bahnbrechende, PREMIS vorausgehende Metadatenmodell der neuseeländischen Nationalbibliothek aus: National Library of New Zealand, Preservation Metadata. Metadata Standards Framework – Metadata Implementation Schema, 2003, <http://www.natlib.govt.nz/downloads/metaschema-revised.pdf> (Abruf 8.4.2010).

¹⁰ Gemeinsam ist den Modellen, dass das Informationsobjekt stets durch dieselbe Zahl von Einheiten (Dateien bzw. Repräsentationen) repräsentiert wird.

hen, kontinuierlich an. Tendenziell erhöht sich dadurch die Komplexität des Archivals, es wird immer weniger überschaubar.

Eindeutigkeit der Beschreibung, Erweiterbarkeit des Archivierungspakets, Zitierbarkeit und die Handhabung der Metadaten sprechen bei der Archivierung digitaler Unterlagen gegen das Dateimodell. Unterstellt ist dabei die Verwendung der Migrationsstrategie und das Ziel, die Unterlagen über sehr lange Zeiträume hindurch erhalten zu wollen. Bei einer reinen Emulationsstrategie wäre das Repräsentationenmodell vielleicht nicht erforderlich. Diese wird jedoch von den Archiven, die derzeit bereits digitale Archivierung betreiben, nicht verfolgt, auch zeichnet sich kein Strategiewechsel ab.

Digitale Objekte und ihre Repräsentationen im Landesarchiv Baden-Württemberg

Aus den genannten Gründen hat sich das Landesarchiv Baden-Württemberg 2006 für das Repräsentationenmodell entschieden. Das Dateimodell soll nicht alternativ verwendet werden. Das Informationsobjekt wird als „Digitales Objekt“ bezeichnet. Jedes Digitale Objekt hat mindestens eine Repräsentation, die ihrerseits aus mindestens einer Datei besteht. Bitströme werden nicht nachgewiesen. Das System ist einfach, klar und eindeutig, zur Abfolge Datei – Repräsentation – Digitales Objekt gibt es keine alternative Möglichkeit der Verzeichnung.

Ein besonderes Problem war es, den Umfang bzw. die Ebene des Digitalen Objekts zu bestimmen. Die Volkszählung 1970 besitzt beispielsweise 9,8 Millionen Datensätze von 9,1 Millionen Befragten (manche Einwohner mussten zwei Fragebögen ausfüllen), sie wurde in vier Dateien (je eine pro Regierungsbezirk) übernommen. Die Zahl der denkbaren Digitalen Objekte liegt daher bei diesem Beispiel zwischen eins und 9,8 Millionen. Alle Optionen erscheinen gerechtfertigt, jede Entscheidung trägt daher den Charakter einer Setzung. Zunächst wurde angestrebt, dass Digitale Objekte – soweit möglich – eine vergleichbare Informationsmenge wie die bereits im System befindlichen konventionellen Archivalieneinheiten umfassen sollten. Nachgewiesen wurde die Volkszählung 1970 schließlich als ein Digitales Objekt. Für die digitale Bestandserhaltung ist es einfacher, nur ein Objekt verwalten zu müssen. Auch bietet diese Option als einzige die Möglichkeit, beliebige Teilmengen als Benutzungspakete¹¹ auszugeben.

Grundsätzlich werden wie in diesem Beispiel auch die anderen vom Landesarchiv archivierten Digitalen Objekte mit einem Archivalie respektive einer Verzeichnungseinheit gleichgesetzt. Digitale Objekte sind daher die Einheit, die bestellt werden

¹¹ Benutzungspakete wurden im OAIS-Standard als *Dissemination Information Packages* (DIP) definiert. Je nach Interesse eines Benutzers könnte ein solches Paket nach Ablauf der nicht verkürzbaren Sperrfristen (Statistik) nur die Datensätze eines Ortes, die Datensätze aller Jugendlichen oder aber alle Datensätze enthalten.

kann und die vom Archiv als Einheit erhalten werden soll. Die Vielgestaltigkeit der zu Bewertung und Übernahme anstehenden digitalen Unterlagen und ihre prinzipiell mögliche Unabgeschlossenheit bereiten nun immer wieder Probleme beim Versuch, aus diesen Unterlagen erhaltbare, d.h. abgeschlossene Informationsobjekte zu bilden. Im Metadatenschema des Landesarchivs Baden-Württemberg ist daher die Option vorgesehen, dass ein Digitales Objekt andere Digitale Objekte enthalten kann. Ein Beispiel: Zur Volkszählung 1970 wurden ein Unterobjekt mit den Erhebungsdaten selbst und ein weiteres mit Scans der originalen Erhebungsbögen gebildet. Bislang wurden höchstens zwei Stufen (Ober- und Unterobjekt) benötigt.

Hybridobjekte

Hybridobjekte können als eine logische Einheit verstanden werden (z.B. die Steuerakte zu einer Person), die sowohl einen papiernen als auch einen digitalen Bestandteil besitzt. Künftige Benutzer haben aller Voraussicht nach zumeist ein Interesse an der ganzen Steuerakte, sie muss daher als eine Archivalieneinheit nachgewiesen werden. Die beiden Einzelteile dieses Archivals sind aber auf sehr unterschiedlichen Wegen zu erhalten. Hybridobjekte können im Landesarchiv Baden-Württemberg ebenso wie Digitale Objekte Unterobjekte besitzen. Die Unterobjekte bestehen bei ihnen aus einem Digitalen Objekt und der Papierakte. Beide Einheiten können separat erhalten und zusammen bestellt werden. Unter dem Aspekt des Benutzerinteresses unwahrscheinlich aber möglich wäre auch eine separate Bestellung der beiden Einzelteile.

Repräsentationen

Digitale Objekte werden im Nachweissystem des Landesarchivs auf einer Ebene zusammen mit den Papierakten, Urkunden und Amtsbüchern abgebildet. Unterhalb des Eintrags eines Digitalen Objekts erscheinen dessen Repräsentationen, beide sind Bestandteile in der Baumstruktur von Findbüchern und Beständeübersichten des Landesarchivs. Wodurch ist nun eine Repräsentation charakterisiert?

Nach ISAD(G) ist ein Verzeichnungssystem hierarchisch aufzubauen. Die Einheiten einer tieferliegenden Ebene unterscheiden sich durch ihre Inhalte, deren Gemeinsamkeiten auf dem übergeordneten Knoten des Systems beschrieben werden. Eine Akte zerfällt so in mehrere Vorgänge, ein Vorgang wiederum in mehrere Dokumente (vgl. auch das DOMEA-Konzept). Auf den ersten Blick gleicht diese Darstellung derjenigen im Repräsentationenmodell des Landesarchivs. Ein Digitales Objekt enthält mehrere Repräsentationen, eine Repräsentation mehrere Dateien. Die ähnliche graphische Darstellung hat hier aber eine andere Bedeutung. Die Repräsentationen enthalten eben keine unterschiedlichen Inhalte, wie dies bei den Vorgängen der Fall

ist. Stattdessen sollen alle Repräsentationen eines Digitalen Objekts im Kern dieselbe Information enthalten.

Was ist nun das Gemeinsame aller Repräsentationen eines Informationsobjekts bzw. eines Digitalen Objekts und durch was unterscheiden sie sich? Schließlich müssen sich Repräsentationen zueinander unterscheiden, um überhaupt als verschiedene Entitäten sinnvoll beschrieben zu werden. Welche Unterschiede sind nun „erlaubt“? Ausgangspunkt dieser Überlegungen muss das Informationsobjekt, also die von Menschen definierte abgegrenzte Menge an Informationen sein. Diese Informationen sollen über eine sehr lange Zeit transportiert werden. Was bedeutet es nun vor dem Hintergrund des *Performance Model*¹², dass sich die Information in immer wieder neu erstellten Repräsentationen ausprägt? Da nicht alles erhalten werden kann, müssen bei den zur Bewertung und Übernahme anstehenden digitalen Unterlagen diejenigen Eigenschaften definiert werden, die auf jeden Fall erhalten werden sollen. Diese Eigenschaften bilden eine Teilmenge aller Eigenschaften, die Fachwelt spricht von *Significant Properties* bzw. signifikanten Eigenschaften. Ein einfaches Beispiel: Die Reihenfolge der Zeichen in einem Text oder die Verteilung der Farben in einem Bild können solche signifikante Eigenschaften darstellen. Welche Eigenschaften nun als signifikant eingeschätzt werden, hängt davon ab, welche künftigen Benutzer ein Archiv erwartet.¹³

Die Repräsentationen eines Informationsobjekts sollten also alle dieselben signifikanten Eigenschaften enthalten. Andere Eigenschaften können von Repräsentation zu Repräsentation variieren (z.B. die Dateiformate der enthaltenen Dateien).¹⁴

Unterschieden werden sollte schließlich der Begriff der Repräsentation auch von dem im OAI-Standard geprägten Begriff der *Representation Information*. *Representation Information* ermöglicht das Verständnis eines Datenobjekts, es übersetzt dieses Objekt in sinnvollere Konzepte. Eine Dateiformatbeschreibung wäre zum Beispiel eine solche *Representation Information*. Diese qualifiziert also Daten oder Informationen für einen bestimmten Zweck. Dagegen ist eine Repräsentation in dem hier verwendeten und von PREMIS definierten Sinne ein abgegrenztes Objekt zur Wiedergabe des Informationsobjekts.

¹² wie Anm. 2.

¹³ Weitere Angaben zu diesen Zusammenhängen in Christian KEITEL, Benutzerinteressen annehmen und signifikante Eigenschaften festlegen. Einige neue Aufgaben für Archivare. In: Archive im digitalen Zeitalter. Überlieferung – Erschließung – Präsentation. 79. Deutscher Archivtag in Regensburg (Tagungsdokumentation zum Deutschen Archivtag 14), Neustadt a.d. Aisch 2010, S. 29–42. Die Nestor-AG Digitale Bestandserhaltung bereitet einen Leitfaden zu der Thematik vor, weitere Informationen unter <http://www.langzeitarchivierung.de>.

¹⁴ Eine Repräsentation kann unterschiedliche Dateiformate enthalten.

Archivierungspakete

Der OAIS-Standard begreift Archivierungspakete (AIPs/*Archival Information Packages*) als logische Einheiten, die nicht zwingend physisch zusammengehörig gespeichert werden müssen. Die Stabilität eines Archivierungspakets wird durch dessen Speicherung als physische Einheit einerseits erhöht (die Einzelteile können nicht verlustig gehen). Auf der anderen Seite erschwert ein solches Vorgehen jede Veränderung im Laufe des Archivierungsprozesses. Diese Anforderungen sind daher in ein ausgeglichenes Verhältnis zu bringen. Das Landesarchiv Baden-Württemberg macht sich dabei die von OAIS ausdrücklich genannte Option einer Schachtelung einzelner Archivierungspakete zunutze. Im Kern steht die Repräsentation. Sie wird als ein physisches Archivierungspaket, d.h. als aus diversen anderen Dateien bestehende Containerdatei (Informatiker reden von „Archivdatei“) gespeichert. Typische Formate dafür sind TAR und ZIP. In Findbuch und Beständeübersicht ist diese Repräsentation in eine eindeutige hierarchische Struktur eingebettet. Die übergeordneten Dateien bilden so zusammen mit der Repräsentation ein zweites, nur gedachtes logisches Archivierungspaket. Sie verorten die Repräsentation im Archiv. Außerdem kann es notwendig sein, von der Repräsentation aus auf andere Entitäten im Archiv zu verweisen. Diese Verweise können entweder nur von Menschen lesbar sein (z.B. „Formatbeschreibungen stehen im Bereich Dokumentation“) oder über eine AID auch maschinenlesbar auf eine spezifische Entität verweisen (z.B. auf einen Aktenplan). Zusammen mit diesen Entitäten wird somit ein drittes, wiederum logisches Archivierungspaket gebildet. Alle drei Pakete können wie die Schalen einer Zwiebel verstanden werden. Die äußerste Zwiebelschale schließlich bilden Verweise aus dem Archiv hinaus. Als Beispiel kann hier ein Verweis auf die Formatdatenbank PRONOM der National Archives angeführt werden. Für die jeweiligen Informationen wird in diesem Modell die maximal mögliche Stabilität erreicht. Zugleich nimmt die Stabilität der einzelnen Pakete von der Repräsentation als dem Zwiebelinneren bis zur äußersten Schale stetig ab. Die Repräsentation umfasst die Primärdaten. Die äußeren Schalen enthalten dagegen nur Metadaten. Diese bestehen zu großen Teilen aus *Representation Information* im Sinne von OAIS. *Representation Information* wird benötigt, um die Primärdaten verstehen zu können. Sie ist abhängig von den Vorkenntnissen der Benutzer. Anders ausgedrückt: Alles, was die Benutzer nicht an Vorkenntnissen mitbringen, muss als *Representation Information* beschrieben werden. Weiß der Benutzer gar nichts, muss die ganze Welt, angefangen mit Buchstaben, Zahlen, Wörtern und Grammatikregeln der deutschen Sprache beschrieben werden. Natürlich ist so ein System schon logisch nicht realisierbar, da auch die jeweiligen Beschreibungsmittel ad infinitum beschrieben werden müssten. Das Zwiebelschalenmodell der Archivierungspakete ermöglicht es aber, von der vom Archiv selbst für das Archivale vorgehaltenen Information über die ohnehin im Archiv vorhandene Information und die an einem spezifischen Ort der Welt vorhandene Information überzuleiten zu der

Information, die in der Welt ohnehin zur Verfügung steht. So gesehen wäre die Welt als solche das äußerste Archivierungspaket.

Repräsentationen digitaler und konventioneller Archivalien

Die Repräsentationen erlauben es, digitale Unterlagen angemessen in das bestehende Nachweissystem für Archivalien einzubauen. Dieses System ist über seine hierarchische Abbildung stabil, es ist einheitlich (keine zweite Tektonik wird angelegt), es bietet keine verwirrenden alternativen Verzeichnungsvarianten und es ist durch die Reduzierung auf wenige Beschreibungselemente einfach. Nicht zuletzt ist es eindeutig, da im Gegensatz zum Dateimodell eine Signatur immer dasselbe bezeichnet.

Das Phänomen der mehrfachen Ausgabe derselben Information ist nicht nur bei den digitalen, sondern auch bei den konventionellen Unterlagen aufzufinden. Was sind Mikrofilme oder vom Archiv angefertigte Digitalisate anderes als weitere Repräsentationen der Papierakte oder der Pergamenturkunde? Sicherlich sind hier die Unterschiede zwischen dem Original und der weiteren Erscheinungsform größer, diese kann bei konventionellen Archivalien nicht alle signifikanten Eigenschaften des Originals enthalten, da deren Materialität fehlt. Dennoch eignet sich das Repräsentationenmodell auch zum Nachweis aller Mikrofilme und Digitalisate der konventionellen Unterlagen. Es behebt dadurch das Manko, dass diese zumeist in vom Verzeichnissystem gesonderten Datenbanken verwaltet werden. Das Landesarchiv Baden-Württemberg plant daher alle Archivalien über das Repräsentationenmodell nachzuweisen. Dieses Modell zeichnet sich durch folgende Eigenschaften aus:

- Digitale und konventionelle Archivalien werden in einem System nachgewiesen.
- Bei allen Archivalien wird zwischen Informationen und Repräsentationen unterschieden.
- Repräsentationen werden stets unterhalb der untersten Ebene der Verzeichnungseinheiten nachgewiesen (d.h. in der Regel unterhalb der Archivalieneinheit).
- Der Repräsentationsdatensatz beschreibt
 - bei konventionellen Unterlagen deren Ausprägung selbst,
 - bei digitalen Unterlagen den Container, in dem sich die Dateien befinden.
- Repräsentation 1...
 - ist bei konventionellen Unterlagen das Original,
 - umfasst bei digitalen Unterlagen die übernommenen Dateien.

Jedes Modell hat Grenzen. Beim Repräsentationenmodell gibt es aus heutiger Sicht vor allem zwei problematische Fälle. Zum einen könnte das Archiv gewillt sein, nur

einen Teil eines Archivaes (z.B. nur das Siegel einer Urkunde) reproduzieren zu wollen. Das Digitalisat des Siegels kann nicht eine zweite Repräsentation der Urkunde darstellen, da dabei nur ein Teil der ursprünglichen Information (der Pergamenturkunde) transportiert wird. Möglich wäre, eine logische Ableitung der Urkunde vorzunehmen, die dann wie auch die Urkunde über eigene Repräsentationen verfügt. Diese Lösung würde das System zwar nicht sprengen. Der elegantere Weg wäre jedoch, auch die anderen Teile zu digitalisieren, um das ganze Archivale Benutzern in seiner neuen Erscheinungsform anbieten zu können.

Zweitens könnte ein Bestand schon mikroverfilmt sein, ohne dass ein für die Darstellung der Repräsentationen geeignetes digitales Findbuch vorliegt. Denkbar ist es, entweder unmittelbar unterhalb des Bestands zwei Repräsentationen oder ein Archivale mit Stellvertreterfunktion anzulegen, das dann seinerseits diese Repräsentationen enthält. Soweit ist das Modell nicht gestört. Es ist aber unschön, dass diese Einträge dann gelöscht werden sollten, wenn die Archivalien des Bestandes auch digital nachgewiesen sind. Angesichts der eingangs festgestellten Tendenz zur umfassenden digitalen Erfassung aller Archivalien, scheint es sich hierbei aber um ein vorübergehendes Problem zu handeln. Diskutiert werden kann auch, ob ein Repräsentationsdatensatz stets unmittelbar auf die konkrete, abgegrenzte Erscheinungsform verweisen soll. Bei Beständen, zu denen mehrere Mikrofilme erstellt wurden, würde dies die Zuordnung von Film und Einzelbild und damit sehr viel Aufwand erfordern. Nach ausgiebiger Diskussion kam die Arbeitsgruppe zum Ergebnis, dass dies nicht unbedingt notwendig ist. Es genügt, alle Archivalien eines Bestandes mit der Angabe „Mikrofilme A bis E“ zu versehen. Wichtig erscheint vor allem der Hinweis auf eine zusätzlich vorliegende Erscheinungsform. Exakte Angaben sind zwar wünschenswert, aber nicht unbedingt erforderlich.

Wie bereits erwähnt hat jedes Modell seine Grenzen. Dennoch gibt es genügend Argumente, die für eine Umsetzung des Repräsentationenmodells sprechen. Nach seiner Einführung sind für jedes Archivale alle Erscheinungsformen auf einen Blick erkennbar. Damit ist nicht nur für die Benutzung ein wesentlicher Schritt erreicht. Auch für Fragen der Bestandserhaltung sind dann alle relevanten Informationen von einer Stelle aus abrufbar.