

Elektronische Archivierung im Bundesarchiv – Die Suche nach einem geeigneten Metadatenschema

Karsten Huth

Einleitung

Die Suche nach einem geeigneten Metadatenschema begann am Bundesarchiv im August 2005. Zu diesem Zeitpunkt gab es schon mehrere internationale Initiativen und Arbeitsgruppen, die bereits auf dem Weg waren. Dieser Aufsatz versucht, sowohl die Suche des Bundesarchivs als auch die der internationalen Initiativen zusammenzufassen. Jede Suche beginnt mit einem ersten Schritt in unbekanntes Neuland. Dass es eher die grundlegenden und elementaren Fragen der Erhaltung und Bewahrung sind, auf die man bei der Suche nach den geeigneten Metadaten für die digitale Archivierung stößt, mag überraschen. Die Jahrhunderte alte Vorherrschaft des Papiers schien alle grundlegenden Fragen der Bestandserhaltung geklärt zu haben. Aufbewahrt werden Akten, Bücher und Kunstwerke. Ziel der Archivierung ist der Erhalt eines archivwürdigen Objekts und dieses archivwürdige Objekt hat einen intrinsischen Wert, in den meisten Fällen das auf Papier gedruckte Wissen des Verfassers. Gegebenenfalls muss ein Objekt durch geeignete restauratorische Maßnahmen stabilisiert werden, aber das Objekt steht im Mittelpunkt der Bemühungen. Einzig bei der Mikroverfilmung wird dieses Prinzip durchbrochen. Das Objekt steht aber nur scheinbar selbsterklärend wie ein Monolith in der Landschaft. Ein Artefakt wie

der über 2000 Jahre alte Computer von Antikythera¹ zeigt, dass ein Objekt ohne seinen Kontext kaum Antworten gibt, sondern tausend neue Fragen aufwirft, die ohne weitere Funde nicht beantwortet werden können. Der Computer von Antikythera wirkt wie ein aus der Zeit gefallenes, unpassendes Puzzleteil. Gut erhalten gibt es uns doch keine Antworten, weil die Kultur, die es einst geformt hat, von der Bildfläche verschwunden ist. Die Frage nach dem Ziel all unserer Bemühungen war somit auch schon vor der digitalen Revolution gegeben. Bewahren wir Objekte aus der Vergangenheit oder das Wissen über die Vergangenheit? Aus welchem Stoff formen wir unser gemeinsames kulturelles Gedächtnis heute und in der Zukunft? Im allgemeinen Getöse der digitalen Revolution ist es schwer, das Alte im scheinbar Neuen noch zu erkennen. Archive, Bibliotheken und Museen haben immer Daten über ihre Objekte erschlossen, wie hätten sie sonst arbeiten können? Neu sind nur der Begriff der Metadaten und die Arten von Informationen, die zusätzlich bei der digitalen Archivierung zur Erschließung anfallen. Um diese neuen, zusätzlichen Informationen und ihre Erschließung soll es in diesem Aufsatz gehen.

Das digitale Objekt und sein Kontext

Das Bundesarchiv hat Erfahrungen mit der Archivierung von digitalen Objekten, die bereits von der technischen Entwicklung überholt wurden, gesammelt. Die Archivierung der Datenbanken aus Behörden und Einrichtungen der DDR haben gezeigt, dass außer der Speicherung der Primärdaten und der Informationen zur archivischen Beschreibung und Erschließung von Beständen und Objekten in Findbüchern auch noch zusätzliche Angaben notwendig sind, um später auf das digitale Objekt zugreifen zu können.² Diese Metadaten müssen nicht nur die digitalen Objekte hinreichend beschreiben, sie müssen auch das Umfeld, in welchem die Objekte geschaffen, bearbeitet und genutzt wurden, berücksichtigen. Dieses Umfeld, der Kontext, muss näher analysiert und spezifiziert werden.

Die besondere Eigenschaft eines digitalen Objektes ist seine simultane Existenz auf drei Ebenen. Jedes digitale Objekt kann gleichzeitig aus unterschiedlichen Perspektiven als ein physisches, logisches und konzeptuelles Objekt betrachtet werden.³

- physisches Objekt: Zeichen, die auf einem Informationsträger eingeschrieben sind. Die Art und Weise der physischen Beschaffenheit dieser Zeichen kann aufgrund der unterschiedlichen Beschaffenheit des Trägers sehr verschieden sein. Auf einer CD-ROM sind es die sogenannten „pits“ und „lands“ auf der Trägoberfläche, bei magnetischen Datenträgern sind es Übergänge zwischen magnetisierten und nicht magnetisierten Teilchen (die so genannten Flusswechsel). Auf der physischen Ebene haben die Bits keinerlei Bedeutung, das heißt, Bits, die zu einem Text gehören, unterscheiden sich in keiner Weise von Bits, die Teil eines Computerprogramms sind.

¹ Schirawski, Nikolai: 7 Dinge, die es nicht geben dürfte – oder?, in: P.M. Magazin, 07, 2003, S. 29-30, http://www.pm-magazin.de/de/heftartikel/artikel_id712.htm.

² Rathje, Ulf; Wettengel, Michael, Digitale Datenbestände von Behörden und Einrichtungen der DDR im Bundesarchiv, in: Historical Social Research, Vol. 24, 1999 Nr. 4, S. 70-101, hier S. 73.

³ Thibodeau, Kenneth, Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, in: The State of Digital Preservation. An International Perspective, Washington D.C. 2002, S. 4-31, hier S. 6-10, <http://www.clir.org/pubs/abstract/pub107abst.html> (Mai 2006).

- logisches Objekt: Bits, die vom Informationsträger gelesen und nun von einer Software interpretiert werden. In dieser Ebene existieren die Objekte als Bitsequenz, die als Dateiformat erkannt wird.
- konzeptuelles Objekt: Zeichen und Objekte, die vom Menschen wahrgenommen und sinnvoll interpretiert werden können. Das konzeptuelle Objekt ist die eigentliche, für den Betrachter bedeutungsvolle Einheit (z.B. ein Text, Musikstück, Film, elektronische Akte).

Ausgehend von dieser Betrachtung ist klar, dass die alleinige Beschreibung eines digitalen Objekts auf der konzeptuellen Ebene nicht ausreicht, da ein digitales Objekt untrennbar mit seiner technischen Umgebung verwoben ist. Um ein Objekt für den Menschen interpretierbar auf einem Monitor darzustellen, müssen viele technische Komponenten (Darstellungsprogramm, Betriebssystem, Hardware) aufeinander abgestimmt arbeiten. Wird dieses vornehmlich technische Wissen nicht gemeinsam mit der inhaltlichen und formalen Beschreibung gespeichert, so ist das digitale Objekt bei einer Veränderung der technischen Umgebung im schlimmsten Fall für den Nutzer nicht mehr lesbar darzustellen, was einem Totalverlust des Archivobjekts gleichkommt.

Wo findet man das nötige Wissen über die Darstellung eines digitalen Objekts? Es beinhaltet Kenntnisse über das Ablesen des Bitstroms von der Oberfläche eines physischen Speichers, Kenntnisse über die technische Interpretation eines Dateiformates durch spezielle Programme sowie Kenntnisse zur Darstellung der interpretierbaren Information über die entsprechenden Hardwarekomponenten (z.B. Monitor, Drucker, Lautsprecher usw.). Oft befindet sich das Wissen implizit in den ausführbaren Programmen.⁴ Das Wissen über die notwendigen Interaktionen zwischen der Maschine und dem Nutzer der Informationen befindet sich dagegen oft in den Köpfen der Entwickler und denen, die mit einer technischen Plattform regelmäßig arbeiten. Wer regelmäßig mit Windows-Betriebssystemen zu tun hat, der weiß, dass er mit einem schnellen wiederholten Druck auf die linke Maustaste bei einer Positionierung des Cursors auf das entsprechende Symbol eine Datei öffnen kann. Dieses Alltagswissen wird jedoch mit der Zeit verloren gehen, wenn Windows-Betriebssysteme nicht mehr flächendeckend verwendet werden. Das notwendige Wissen über die Darstellung eines digitalen Objekts ist demnach weit verstreut. Es befindet sich in den Köpfen der Entwickler und Anwender, in den Algorithmen der Software sowie in den technischen Beschreibungen und Handbüchern. Die digitale Archivierung macht es notwendig, dieses Wissen wieder zusammenzutragen und zu erschließen.

Anforderungen an ein Metadatenchema

Die maschinenlesbaren Bestände des Bundesarchivs wurden bislang in so genannten Dossiers auf Papier verzeichnet. Der große Vorteil eines Dossiers ist seine unkomplizierte Lesbarkeit durch das menschliche Auge. In den Dossiers wurden bereits Informationen zu allen drei Ebenen eines digitalen Archivobjekts erschlossen:

⁴ DFG Projekt Datenbankgestützte „Langzeitarchivierung Digitaler Objekte – Schlussbericht“, Universität der Bundeswehr, München, 2005, hier S. 11 http://www2-data.informatik.unibw-muenchen.de/LZA/publications/Schlussbericht_Langzeitarchivierung.pdf (Mai 2006).

Struktur eines Dossiers

Dossier für B 119 MD/1

Ordner anlegen mit sieben Untergliederungen

1. Vorbemerkung (konzeptuell, logisch, physisch)
2. Beschreibung der benutzbaren Version (logisch)
3. Beschreibung der Originalversion (logisch, physisch)
4. Codelisten (konzeptuell, logisch)
5. Dokumentation (konzeptuell, logisch)
6. Protokolle der technischen Bearbeitung: Kopierung, Konvertierung, Kassation (physisch, logisch)
7. Literaturhinweise (konzeptuell)

Der große Nachteil des Dossiers ist seine schlechte maschinelle Verwertbarkeit innerhalb eines Archivs. Die Metadaten liegen unstrukturiert vor und lassen sich nur schwer in geläufige, relationale Datenbanksysteme exportieren. Zudem ist die Anfertigung eines Dossiers sehr aufwendig und kostenintensiv, ein technisch gestütztes Information Retrieval ist nicht möglich, und es gibt keine Anbindung zu einem automatischen Datenaustausch, z.B. zwischen der abgebenden Institution und dem Archiv. Die Abkehr von den Dossiers hin zu maschinell verwertbaren Metadatenstrukturen ist ein notwendiger Schritt.

Die Anforderungen an ein Metadatenschema sind sehr hoch, weil die gespeicherten Metadaten später die Grundlage für einen geregelten Betrieb des Archivs bilden. Nur wenn alle Daten korrekt verzeichnet sind, wird man in Zukunft auf die digitalen Archivobjekte zugreifen können. Durch geordnete Metadaten wird das notwendige Wissen über die digitalen Objekte strukturiert gespeichert, so dass der Archivar in Zukunft die richtigen Fragen an sein Archivsystem stellen kann und zuverlässige Antworten erhält. Folgende Fragen sind in Bezug auf eine dauerhafte Bewahrung digitaler Objekte grundlegend:

Wie kann man ein Objekt x mit den heutigen technischen Mitteln darstellen und nutzen?

Diese elementare Frage beinhaltet die Angabe einer kompletten und aktuell lauffähigen Hardware-/Softwareplattform, die ausführliche Beschreibung des Dateiformats, Hinweise auf Passwörter und Verschlüsselungen sowie Gebrauchsanweisungen für die Nutzung.

Wer hat wann in welcher Art und Weise das Objekt verändert?

Die meisten digitalen Archive werden die Migration als Maßnahme zur Erhaltung der digitalen Archivobjekte wählen. Bei dieser Strategie entstehen mit jedem Migrationsschritt neue Objekte, die sich auf der logischen Ebene der Bit-Sequenzen vollkommen von ihren Ausgangsobjekten unterscheiden. Dabei stellt sich automatisch die Frage nach der Authentizität des neu entstandenen Objekts. Ein vertrauenswürdige Archiv muss die Veränderungen dokumentieren und festhalten, inwieweit auch die für den Nutzer wesentliche konzeptuelle Ebene eines Objekts betroffen ist. Es ist wichtig, dass alle Maßnahmen und die beteiligten Personen und Werkzeuge (z.B. Konvertierungsprogramme) zu jedem Zeitpunkt zurückverfolgt werden können. Dadurch kann die Authentizität eines Archivobjekts beurteilt werden.

Andere wichtige Fragen kommen hinzu:

Wie lange sind die physischen Datenträger noch haltbar?

Gibt es technische Entwicklungen, die eine Migration bestimmter Dateiformate erforderlich machen?

In welcher technischen Umgebung ist ein digitales Objekt x entstanden?

In welchem Zustand wurde es im Archiv abgeliefert?

In welcher Beziehung steht ein Objekt x zu anderen digitalen Archivobjekten?

Hier kann zunächst nur eine Auswahl möglicher Fragen angeführt werden. Mit den Erfahrungen, die durch die Entwicklung von Test- und Pilotsystemen gewonnen werden, kommen bestimmt neue Fragen auf. Nebenbei sollte ein Metadatenchema auch die folgenden Eigenschaften aufweisen, damit es in einem Archiv verwendet werden kann:

Es sollte die konzeptuellen, logischen und physischen Metadaten erfassen und ökonomisch strukturieren.

Es sollte eine große Anwendergemeinde haben, wenn möglich als Standard vorliegen.

Es sollte in XML umsetzbar sein und es sollten möglichst schon fertige XML-Schemadateien vorliegen.

Es sollte auch für den Betrieb in relationalen Datenbanken geeignet sein.

Die Verwendung eines Standards oder eines weit verbreiteten Schemas ist für das Archiv von Vorteil, weil es von den Erfahrungen anderer Nutzer profitiert und sich das Archiv die kosten- und zeitintensive Eigenentwicklung eines Metadatenchemas sparen kann. Außerdem erleichtert die gemeinsame Verwendung eines Standards die Zusammenarbeit zwischen den Archiven und ermöglicht einen reibungslosen Datentransfer zwischen der abgebenden Institution, den Archiven und den Archivnutzern. Die Umsetzung des Schemas in das XML-Format ist notwendig für die dauerhafte Speicherung der strukturierten Metadaten, die als digitales Objekt den gleichen Gefahren unterworfen sind, wie die primären Archivobjekte. Die Verwendung eines Dateiformats, das strategisch günstig für eine nachhaltige Speicherung ist, ist daher dringend anzuraten. XML-Dateien können auf der Basis von ASCII- oder Unicode-Zeichensätzen erstellt werden. Beide Zeichensätze sind plattformunabhängig, der ASCII-Code ist seit beinahe 40 Jahren im Einsatz. XML-Dateien können zur Not mit einfachen Texteditoren dargestellt werden. Gleichzeitig sind XML-Dateien aber auch maschinell verwertbar. Sie können in gängige Datenbanksysteme importiert werden. Eine Doppelstrategie bietet sich an. Langfristig werden die Metadaten im XML-Format gesichert. Für den täglichen Betrieb werden die Daten in ein aktuelles Datenbanksystem importiert. Von dieser Datenbank aus haben Archivmitarbeiter und Archivnutzer den Zugriff auf die Metadaten und können sie mit modernen Retrieval-techniken durchsuchen.⁵

⁵ Heuscher, Stephan [u.a.], Providing Authentic Long-term Archival Access to Complex Relational Data, Schweizerisches Bundesarchiv, 2004, hier S. 11, http://arxiv.org/PS_cache/cs/pdf/0408/0408054.pdf (Mai 2006).

Internationale Entwicklungen und Forschung am Bundesarchiv

Seit Längerem gibt es eine internationale Diskussion darüber, welche Daten für die langfristige Verfügbarkeit von digitalen Objekten benötigt werden. Die daraus resultierenden Arbeitsgruppen haben seit Kurzem verwertbare Ergebnisse in Form von neuen Veröffentlichungen und Metadatenschemata vorgelegt. Man ist sich nun stärker darüber im Klaren, welche speziellen Informationen ein Informationspaket aus Metadaten zur Langzeitarchivierung enthalten muss. Maßgebend für die Ergebnisse der Arbeitsgruppen ist das „Open Archival Information System“, ein Rahmenwerk zum Aufbau und zur Pflege eines digitalen Archivs, das den Status eines ISO-Standards erreicht hat. Innerhalb des Rahmenwerks wird ein Bauplan zur Bildung eines „Archival Information Package“ (AIP) beschrieben, das sowohl aus dem digitalen Archivobjekt wie auch aus den notwendigen Metadaten besteht.⁶ Dieses AIP ist das Rückgrat eines digitalen Archivs und die Informationen, die in dem Packet gespeichert werden, sind entscheidend für den Erfolg oder den Misserfolg der digitalen Überlieferungsbildung.

Ein AIP besteht aus den folgenden Komponenten:

Content Information:

Data Object: das digitale Archivobjekt.

Representation Information: die notwendigen Informationen, die benötigt werden, um das Objekt darzustellen.

Preservation Description Information:

- Reference Information: Formale Beschreibung des konzeptuellen Objekts (z.B. Titel, Verfasser, Identifizier).
- Provenience Information: Information, die die Herkunft und Geschichte des Objekts dokumentiert. Dies schließt auch die technischen Maßnahmen wie Migrationen ein.
- Fixity Information: Informationen darüber, ob das Data Object in ungewollter Art und Weise verändert oder beschädigt wurde. Beinhaltet die Angabe von digitalen Signaturen, Hash-Werten und Passwörtern.
- Context Information: Beschreibt die Beziehung des Archivobjekts zu seiner Umwelt. Beinhaltet die Beziehungen zu anderen Objekten.

Der Aufbau eines AIPs diente vielen Initiativen, die sich mit der Entwicklung eines Metadatenschemas befassten, als Vorlage für ihre eigenen Vorschläge. Die im Jahr 2000 gegründete „OCLC/RLG Working Group on Preservation Metadata“ veröffentlichte im Juni 2002 ihr Rahmenwerk „A Metadata Framework to Support the Preservation of Digital Objects“.⁷ Die Vorschläge der Arbeitsgruppe hielten sich sehr eng an den OAIS-Bauplan eines AIPs. 2003 wurde erneut eine Arbeitsgruppe mit dem Namen PREMIS (Preservation Metadata Implementation Strategies) von OCLC und der RLG ins Leben gerufen. Hauptaufgabe war diesmal die Entwicklung eines implementierbaren Kernsatzes von Metadaten zur Archivierung von digitalen Objekten.⁸ Zunächst

⁶ Reference Model for an Open Archival Information System (OAIS) Blue Book, NASA, Washington DC, 2002, <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> (Mai 2006).

⁷ Preservation Metadata and the OAIS Information Model – A Metadata Framework to Support the Preservation of Digital Objects. OCLC/RLG Working Group on Preservation Metadata, Dublin, Ohio, 2002, http://www.oclc.org/research/projects/pmwg/pm_framework.pdf (Mai 2006).

⁸ Caplan, Priscilla; Guenther, Rebecca, Practical Preservation: The PREMIS Experience, in: Library Trends, Vol. 54, No. 1, Summer 2005, S. 111-124, http://www.loc.gov/standards/premis/caplan_guenther-librarytrends.pdf (Mai 2006).

wurde eine Umfrage über die Archivierungspraktiken in verschiedenen Institutionen durchgeführt.⁹ Dann wurde auf der Grundlage der Umfrageergebnisse und den Ergebnissen der vorhergehenden Arbeitsgruppe der Metadaten-Kernsatz erstellt. Das Ergebnis wurde im Mai 2005 als Data Dictionary veröffentlicht.¹⁰ In diesem Dictionary werden der Aufbau und jedes einzelne Element des PREMIS Metadatenschemas beschrieben.

Die umfassende Evaluierung bestehender Verfahren und Kompetenzen auf dem Gebiet der digitalen Archivierung ist ein Vorzug des PREMIS Metadatenschemas gegenüber anderen verfügbaren Schemata. Zurzeit durchläuft PREMIS eine 18-monatige Evaluierungsphase, in der die Erfahrungen bei der Verwendung des Schemas innerhalb einer offenen Diskussionsgruppe (PREMIS Implementors Group Forum)¹¹ ausgetauscht werden. Die praktischen Erfahrungen aus der Gruppe sollen in eine überarbeitete Version des Data Dictionary einfließen. Neben der Diskussionsgruppe wird es eine Managing Agency und ein Editorial Committee zur Koordinierung neuer Entwicklungen des Standards und der Produkte (XML-Schemata, Data Dictionary, Berichte) geben.¹²

Die Möglichkeiten von PREMIS

Mit den Mitteln aus dem BMBF-geförderten Projekt nestor¹³ wurde eine AG „Archivische Metadaten“ gebildet. Zunächst war die Entwicklung einer deutschen Eigenlösung auf der Grundlage internationaler Vorgänger geplant. Nach einer Einarbeitungsphase wurde am Bundesarchiv der Entschluss gefasst, die bereits vorliegenden Metadatenschemata der Nationalbibliothek Neuseelands¹⁴ und das PREMIS Metadatenschema zunächst als Ganzes, mit digitalem Archivgut aus dem Bestand des Bundesarchivs zu testen. Dabei traten die Vorteile von PREMIS offen zutage. Schnell wurde deutlich, dass sich im Hinblick auf die Ausgefeiltheit der internationalen Vorlagen die Eigenentwicklung eines Metadatenschemas kaum lohnen würde. Deshalb wurde PREMIS als Metadatenschema zur Erfassung von technischen, strukturellen Metadaten sowie zur Dokumentation der vom Archiv vorgenommenen Maßnahmen verwendet.

Mit Hilfe der in PREMIS erschlossenen Daten können wesentliche Fragen, die bei der digitalen Archivierung auftreten, beantwortet werden.

Frage 1:

Wie kann man ein Objekt x mit den heutigen technischen Mitteln darstellen und nutzen?

⁹ Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community. Report by the joint OCLC/RLG Working Group Preservation Metadata: Implementation Strategies (PREMIS). Dublin, Ohio, 2004, <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf> (Mai 2006).

¹⁰ Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group, Dublin, Ohio, 2005, <http://www.oclc.org/research/projects/pmwg/premis-final.pdf> (Mai 2006).

¹¹ <http://www.loc.gov/standards/premis/pig.html> (Mai 2006).

¹² <http://www.loc.gov/standards/premis/ma.html> (Mai 2006).

¹³ <http://www.langzeitarchivierung.de/> (Mai 2006).

¹⁴ National Library of New Zealand – Preservation Metadata, 2003, http://www.natlib.govt.nz/files/nlnz_data_model.pdf (Mai 2006).

```

<environment>
  <environmentCharacteristic>known to work</environmentCharacteristic>
  <environmentPurpose>erstellen</environmentPurpose>
  <environmentPurpose>migrieren</environmentPurpose>
  <environmentPurpose>editieren</environmentPurpose>
  <environmentNote>aktuelle Zugriffsplattform - Stand 1/2006</environmentNote>
  <software>
    <swName>Microsoft Word</swName>
    <swVersion>97 SR-2</swVersion>
    <swType>Textverarbeitung</swType>
    <swOtherInformation>Mit diesem Programm wurde die Datei erstellt.</swOtherInformation>
  </software>
  <software>
    <swName>Windows XP</swName>
    <swVersion>Version 2002 Service Pack 2</swVersion>
    <swType>Betriebssystem</swType>
    <swOtherInformation>Betriebssystem auf dem das Programm Microsoft Word läuft.</swOtherInformation>
  </software>
  <hardware>
    <hwName>Intel Celeron 2.00 GHz</hwName>
    <hwType>Prozessor</hwType>
    <hwOtherInformation>AMD und Intel Prozessoren mit vergleichbarer oder höherer Leistung sind ebenso geeignet.</hwOtherInformation>
  </hardware>
  <hardware>
    <hwName>504 MB RAM</hwName>
    <hwType>Arbeitsspeicher</hwType>
    <hwOtherInformation>Auch höhere RAM-Werte sind möglich.</hwOtherInformation>
  </hardware>

```

Abbildung 1: Environment

Abbildung 1 beschreibt die Hardware-/Software-Umgebung (Environment) für eine Datei im Windows-Word-97-Format. Das Element Environment enthält mehrere Unterelemente. Die Angabe „known to work“ des Elements „Environment Characteristics“ bedeutet, dass die entsprechende Datei mit der folgenden Hardware-/Software-Plattform darstellbar ist. Unter „Environment Purpose“ kann angegeben werden, welche Handlungen mit der Datei in dieser technischen Umgebung möglich sind. In dem Beispiel können wir die Datei nicht nur darstellen, wir können sie auch durch eine Dateneingabe editieren oder in ein anderes Dateiformat konvertieren. Die folgenden Elemente „Software“ und „Hardware“ beschreiben die notwendigen Programme bzw. die einzelnen Hardwarekomponenten. Das Beispiel fasst sich sehr kurz. Man kann die technische Umgebung der Datei noch deutlich detaillierter beschreiben. In unserem Beispiel haben wir nur Angaben zu einem Textverarbeitungsprogramm, das unser Windows-Word-97-Format darstellen kann, und zum passenden Betriebssystem, auf dem das Textprogramm aufsetzt. Unter „Hardware“ finden sich Angaben zum Prozessortyp und seiner Geschwindigkeit sowie zur Größe des Arbeitsspeichers. Alle notwendigen Informationen zur Beantwortung von Frage 1 liegen somit erschlossen in einer einheitlichen Struktur vor.

Frage 2:

Wer hat wann in welcher Art und Weise das Objekt verändert?

```

<eventType>migration</eventType>
<eventDateTime>2005-10-20T15:00:00+01:00</eventDateTime>
<eventDetail>EBDCIC Datei wird in zwei ASCII csv Dateien migriert. Dateiendungen aus technischen Gründen fed.</eventDetail>
<eventOutcomeInformation>
  <eventOutcome>OK</eventOutcome>
</eventOutcomeInformation>
<linkingAgentIdentifier>
  <linkingAgentIdentifierType>Barch_ID_B2_Agentnameschlüssel</linkingAgentIdentifierType>
  <linkingAgentIdentifierValue>ans_konverterEXE</linkingAgentIdentifierValue>
  <linkingAgentRole>Konvertierer</linkingAgentRole>
</linkingAgentIdentifier>
<linkingObjectIdentifier>
  <linkingObjectIdentifierType>Barch_ID_B2_filenameschlüssel</linkingObjectIdentifierType>
  <linkingObjectIdentifierValue>fns_0009_A9881.B7101.N11</linkingObjectIdentifierValue>
</linkingObjectIdentifier>
<linkingObjectIdentifier>
  <linkingObjectIdentifierType>Barch_ID_B2_filenameschlüssel</linkingObjectIdentifierType>
  <linkingObjectIdentifierValue>0009_A9881.B7101.N11_2_0.fed</linkingObjectIdentifierValue>
</linkingObjectIdentifier>
<linkingObjectIdentifier>
  <linkingObjectIdentifierType>Barch_ID_B2_filenameschlüssel</linkingObjectIdentifierType>
  <linkingObjectIdentifierValue>0009_A9881.B7101.N11_2_0.fed</linkingObjectIdentifierValue>
</linkingObjectIdentifier>

```

Abbildung 2: Event

Mit dem Element „Event“ und seinen Unterelementen können die vorgenommenen Maßnahmen eines Archivs und die damit einhergehenden Veränderungen an den digitalen Archivobjekten verzeichnet werden. Das Beispiel zeigt die Beschreibung einer Migration (Event Type). Eine Tabelle aus der DDR-Datenbank zur Volkszählung, die ursprünglich als eine Datei im EBCDIC-Format (ein Codeformat, das auf Großrechenanlagen verwendet wird) vorliegt, wird konvertiert, so dass zwei neue Dateien im ASCII-CSV-Format entstehen (Event Detail und Linking Object Identifier). Die Migration fand am 20.10.2005 um 15:00 Uhr statt (Event Date Time). Die Maßnahme ist wie gewünscht verlaufen (Event Outcome).

```
<agentIdentifier>
..... <agentIdentifierType>Barch_ID_B2_Agentnameschlüssel</agentIdentifierType>
..... <agentIdentifierValue>ans_konverterEXE</agentIdentifierValue>
</agentIdentifier>
<agentName>Konverter.exe</agentName>
<agentType>Programm</agentType>
```

Abbildung 3: Agent

Das Element „Agent“ gibt Auskunft über die an einer Maßnahme beteiligten Personen und die verwendeten technischen Hilfsmittel. Abbildung 3 zeigt die Angaben über das Konvertierungsprogramm, mit dem die Migration (s. Abbildung 2) durchgeführt wurde. In diesem Fall wurde die Migration mit einem Programm namens Konverter.exe durchgeführt. Somit stehen alle Informationen zur Beantwortung von Frage 2 zur Verfügung.

In der Praxis hat man es oft mit komplexen digitalen Objekten zu tun. Das Verhältnis zwischen dem konzeptuellen Objekt (z.B. eine Akte) und dem logischen Objekt (mehrere Dateien eines Formats oder mehrere Dateien in unterschiedlichen Formaten, die zu einer Akte gehören) ist selten 1 zu 1. Besonders die sogenannten „born digital“, Objekte, die es einzig in elektronischer Form gibt (z.B. Webseiten, Datenbanken), weisen komplizierte Beziehungsstrukturen zwischen den einzelnen Dateien auf. Es ist unbedingt notwendig, die Informationen über die Strukturen innerhalb der Metadaten nachzuweisen. PREMIS bietet die Möglichkeit, zusammengehörige Dateien (files) zu einem übergeordneten Objekt, einer Representation, zusammenzufassen. Eine Representation ist definiert als ein Satz von Dateien (files) und den strukturellen Metadaten, der für die komplette Darstellung einer intellektuellen Einheit (intellectual entity) notwendig ist. Unter einer intellectual entity ist ein konzeptuelles Objekt zu verstehen (z.B. eine elektronische Akte, ein Buch usw.).

Ein Beispiel aus dem Bestand des Bundesarchivs kann das Konzept näher erklären. Die Daten zur DDR-Volkszählung des Jahres 1971 wurden in einzelnen Tabellen gespeichert. Jede Tabelle besteht aus 52 Feldern mit Angaben zum Ort, der Gebäudeart, Anzahl und Nutzungsart der Wohnungen usw. Eine Tabelle enthält jeweils die Daten für eine Region, in unserem Beispiel für Berlin. Nach der Migration der Tabelle aus dem EBCDIC-Format in zwei Dateien des ASCII-CSV-Formats (s. Abbildung 2 u. Abbildung 3) liegen die Daten in einem aktuell darstellbaren Format vor. Bedauerlicherweise wurden die Daten während der Volkszählung nicht in Klarschrift, sondern speicherplatzsparend in codierter Form eingegeben (s. Abbildung 4).

```
"W71";13;001000;50;013400;50;01;1;001;1;01;1;1;09;4;1;2;0;0;1;0;0;0;1;0;001;001;1;1;04
"W71";13;001000;50;013400;50;01;1;001;1;02;1;1;09;5;2;2;0;0;1;0;0;0;1;0;002;001;1;1;03
"W71";13;001000;50;013400;50;01;1;001;02;1;1;09;5;2;2;0;0;1;0;0;0;1;0;002;002;2;1;01;
"W71";13;001000;50;013400;50;01;1;001;1;03;1;1;09;4;1;2;0;0;1;0;0;0;1;0;002;001;1;1;04
"W71";13;001000;50;013400;50;01;1;001;03;1;1;09;4;1;2;0;0;1;0;0;0;1;0;002;002;2;1;01;
"W71";13;001000;50;013400;50;01;1;001;1;04;1;1;01;4;1;3;0;0;1;0;0;0;0;1;002;001;1;1;03
```

Abbildung 4: Beispiel Tabelle

Das Beispiel aus Abbildung 4 macht deutlich, dass die Informationen zwar darstellbar sind, sie sind für das menschliche Auge jedoch nicht zu interpretieren. Die Tabelle wäre somit keine gültige Representation, da wir allein durch sie nicht die entsprechende intellektuelle Einheit, die Volkszählungsdaten aus dem Jahr 1971 des Berliner Raums, darstellen können.

Die Zahlen zwischen zwei Semikola sind der verschlüsselte Wert eines Feldes, jede Reihe ist ein Datensatz. Um die Tabelle lesbar darzustellen, benötigt man alle entsprechenden Codelisten, in denen für jeden verschlüsselten Wert die entsprechende Information verzeichnet ist.

1;Art der Toilette - WC in der Wohnung	
2;Art der Toilette - WC ausserhalb der Wohnung	aber im Gebaeude
3;Art der Toilette - WC ausserhalb des Gebaeudes	
4;Art der Toilette - Trockentoilette in der Wohnung	
5;Art der Toilette - Trockentoilette ausserhalb der Wohnung	aber im Gebaeude
6;Art der Toilette - Trockentoilette ausserhalb des Gebäudes	

Abbildung 5: Codeliste 37

Demnach besteht eine Representation der Volkszählungsdaten des Jahres 1971 über den Raum Berlin aus den zwei ASCII-CSV-Dateien, die die Daten der Tabelle enthalten, den 16 notwendigen Codelisten zur Entschlüsselung der Werte sowie einer genauen Beschreibung der Struktur der Datenbanktabelle (s. Abbildung 6).

DE-2-MD/27-Volkszählung¶

¶

Integrierte Datensatz- und feldbeschreibungen ASCII-csv -- Basissatz Personendaten 1971¶

PH= Privathaushalt¶

GE= Gemeinschaftseinrichtung¶

Lfd.-Num- mer	Erfa- sungs- Num- mer	Stellen- vor- z	Stellen- bis- z	Feldlänge	Feldbezeichnung	Codierung (s.-Datei)
1	1	1	3	3	Satzart	P-71
Regional- und Ordnungsangaben						
2	2	5	6	2	Gemeindegemeinschafts-Land	C001.csv
		7	13	6	Gemeindegemeinschaft	Regierungsbezirk/Kreis/Gemeinde laut amtlichem Gemeindeverzeichnis vom 31.12.1990 (Datei:0007_A9661.B8112.N01)
3	3	14	15	2	Stützpunkt/Ortsteil (neu)	In nach 1971 eingemeindeten Gemeinden neu nummeriert, sonst Nummerierung gleich 1971
4	4	17	22	6	Gemeindegemeinschaft - -> Bezirk - -> Kreis - -> Gemeinde	Amthches Gemeindeverzeichnis vom 31.12.1981
5	5	24	25	2	Stützpunkt/Ortsteil (alt)	Nummerierung laut Einteilung der Gemeinden zur Zählung 1971
6	6/7	27	29	3	Zählung - -> PH redefines GE - Lfd.-Nummer des Zählbereiches - Lfd.-Nummer des Zählabschnittes im Zählbereich ODER - -> GE - Lfd.-Nummer der GE	*
7	8	31	33	3	Wohnbezirk	Fallweise in Städten
8	9	35	36	2	Lfd.-Nummer des Gebäudes im Zähl-	1-99 oder Leerfeld

Abbildung 6: Definition der Tabellenstruktur

Nur über die Definition der Tabellenstruktur kann einem Tabellenfeld ein entsprechender Feldname zugeordnet werden, und nur die Feldnamen mit den Werten der Felder zusammen ergeben eine sinnvolle, vollständige Information für den Nutzer. In den Metadaten muss diese Beziehung der Dateien zueinander unbedingt verzeichnet sein. Es kann schließlich nicht das Ziel eines Archivs

sein, lose Dateien zu erhalten. Am Ende der digitalen Archivierung muss für den Archivnutzer eine erschließbare, authentische Information stehen. In dem dargestellten Beispiel ist dies nur möglich, wenn die Informationen, die in mehreren Dateien gespeichert sind, als ein geordnetes Ganzes abrufbar sind. Ein Metadatenchema muss demnach in der Lage sein, die Beziehungen zwischen den Dateien zu erfassen und darzustellen. PREMIS kann die zusammengehörigen Dateien als Representation zusammenfassen und die Beziehungen der Dateien näher beschreiben.

```

<relationshipType>strukturell</relationshipType>
<relationshipSubType>has part</relationshipSubType>
<relatedObjectIdentification>
  <relatedObjectIdentifierType>Barch_ID_B2_filenameschlüssel</relatedObjectIdentifierType>
  <relatedObjectIdentifierValue>fns_datensatzbeschreibungen_ascii_csv_demd27_3103</relatedObjectIdentifierValue>
  <relatedObjectSequence>1</relatedObjectSequence>
</relatedObjectIdentification>

<relationship>
  <relationshipType>strukturell</relationshipType>
  <relationshipSubType>has part</relationshipSubType>
  <relatedObjectIdentification>
    <relatedObjectIdentifierType>Barch_ID_B2_filenameschlüssel</relatedObjectIdentifierType>
    <relatedObjectIdentifierValue>fns_0009_A9881_B7101.N11-1-0</relatedObjectIdentifierValue>
    <relatedObjectSequence>2a</relatedObjectSequence>
  </relatedObjectIdentification>

<relationship>
  <relationshipType>strukturell</relationshipType>
  <relationshipSubType>has part</relationshipSubType>
  <relatedObjectIdentification>
    <relatedObjectIdentifierType>Barch_ID_B2_filenameschlüssel</relatedObjectIdentifierType>
    <relatedObjectIdentifierValue>fns_CI001</relatedObjectIdentifierValue>
    <relatedObjectSequence>3</relatedObjectSequence>
  </relatedObjectIdentification>

```

Abbildung 7: Relationship

Abbildung 7 zeigt die Verknüpfung der Dateien mit der übergeordneten Representation. Die Verknüpfung erfolgt über eindeutige Identifikatoren oder Schlüssel, die den einzelnen Objekten (Representation, File) zugewiesen werden. Das Beispiel zeigt die Verknüpfung von drei Dateien. Die Elemente „Relationship Type“ und „Relationship Subtype“ geben an, dass die Dateien ein Teil einer Representation sind. Die Dateien werden durch ihren Schlüssel im Element „Related Object Identifier Value“ identifiziert. Das Element „Related Object Sequence“ gibt die Reihenfolge an, in der ein Nutzer die Dateien öffnen sollte. An Position 1 steht die Beschreibung der Struktur der Datenbanktabelle. Danach folgt die eigentliche Tabelle, hier mit 2a angegeben, da es sich um zwei Dateien handelt. Dann werden alle Codelisten in der Reihenfolge der entsprechenden Felder verzeichnet.

Der Bestand der DDR-Datenbanken und seine Erschließung durch das Bundesarchiv machten deutlich, worauf in Zukunft bei der Archivierung von digitalen Objekten in Archiven geachtet werden muss. Die Bedingungen, unter denen dieser Bestand von den Mitarbeitern des Bundesarchivs wieder nutzbar gemacht wurde, können hier nicht weiter ausgeführt werden.¹⁵ Der Umstand, dass sich dieser Bestand, der von der technischen Entwicklung längst überholt wurde, mit einem modernen Metadatenchema erschließen ließ, macht Mut, PREMIS auch für alle anderen Arten von digitalem Archivgut zu verwenden.

Alle übrigen Anforderungen an ein Metadatenchema werden von PREMIS wie folgt erfüllt:

- Es sollte die konzeptuellen, logischen und physischen Metadaten erfassen und ökonomisch strukturieren.

¹⁵ Rathje, Ulf: Archivierung von DDR-Daten im Bundesarchiv – ein Rückblick auf zehn Jahre, in: Funktionseiten der DDR. Theoretische Kontroversen und empirische Befunde, Historische Sozialforschung Sonderheft 28, 2003, S. 57-72.

PREMIS erschließt vor allem die logischen und physischen Metadaten. Inhaltliche und formale Metadaten können je nach Bedarf des betreffenden Archivs hinzugefügt werden. Dazu können bereits etablierte Standards wie EAD, EAC und ISAD (G) mit den PREMIS-Metadaten verknüpft werden.

- Es sollte eine möglichst große Anwendergemeinde haben, wenn möglich als Standard vorliegen.

Zurzeit bildet sich eine internationale Anwendergemeinde. Neben der SUB Göttingen und der Deutschen Bibliothek beteiligt sich im Archivbereich das Schweizerische Bundesarchiv mit seiner Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen an der Einführung und Weiterentwicklung von PREMIS.

- Es sollte in XML umsetzbar sein und es sollten möglichst schon fertige XML-Schemadateien vorliegen.

XML-Schema Dateien stehen zur freien Verwendung im Netz bereit.

- Es sollte auch für den Betrieb in relationalen Datenbanken geeignet sein.

Das PREMIS Datenmodell ist bereits in Teilen normalisiert, d.h. es liegt schon als Entity-Relationship Model vor. Da alle gebräuchlichen Datenbanksysteme auf solchen Modellen aufgebaut sind, wird der Import der PREMIS-Metadaten in eine SQL-basierte Datenbank erleichtert.

Ergebnis

Zusammenfassend lässt sich sagen, dass mit der Entwicklung des PREMIS Metadatenschemas nun alle Bauteile vorliegen, um ein Archival Information Package gemäß des OAIS-Standards zu entwickeln. Ein AIP besteht demnach aus dem digitalen Archivobjekt (Data Object), einem kompletten Satz an PREMIS-Metadaten (Representation Information, Fixity Information, Provenance Information) und einem Satz an Metadaten zur archivischen, inhaltlichen Erschließung (Reference Information). Nur wenn diese Informationen mit einem eindeutigen Bezug auf das Archivobjekt erhalten und nutzbar bleiben, ist der langfristige Zugriff auf ein digitales Archivobjekt möglich. Die Erfahrungen des Bundesarchivs mit den DDR-Datenbanken machen dies deutlich. Die Mitarbeiter, die Statistiker und Techniker, die die Datenbanken erstellt haben, stehen nicht mehr zur Verfügung. Die technische Infrastruktur ist veraltet und längst entsorgt. Die Institutionen, die die Zählungen veranlasst haben, ja selbst der Staat, dem sie dienten, existieren nicht mehr. Dennoch ist es dem Bundesarchiv gelungen, die DDR-Daten durch das mühevoll gesammelte von technischen Metadaten wieder nutzbar zu machen. In Zukunft sollten diese Metadaten strukturell einheitlich und maschinell verwertbar gespeichert werden. Das PREMIS-Metadatenchema spiegelt den momentanen Stand des Wissens am besten wider. Seine Verwendung für die Archivierung von digitalem Archivgut bietet sich an.