

Der nestor-Leitfaden zur Digitalen Bestandserhaltung und seine Folgen für die Archive¹

Von CHRISTIAN KEITEL

Längst hat die Literaturproduktion zur digitalen Archivierung schwindelerregende Höhen erreicht, es vergeht kaum ein Jahr, in dem nicht ein neuer, noch wichtigerer Standard vorgestellt und der Aufmerksamkeit der Fachöffentlichkeit anempfohlen wird. Es ist daher nur legitim zu fragen:

Bedarf es wirklich eines neuen Standards zur digitalen Bestandserhaltung? Warum sollten sich Archivarinnen und Archivare mit dem Leitfaden zur digitalen Bestandserhaltung beschäftigen?

Die maßgeblichen Konzepte für die digitale Archivierung wurden bislang zum überwiegenden Teil nicht von den klassischen Archiven entwickelt.² OAIS, PREMIS und die Ergebnisse der Projekte CEDARS (Significant Properties) und PLANETS wurden maßgeblich von Mitarbeitern anderer Einrichtungen, den neuen digitalen Archiven verfasst. Das Performance Model wurde zwar vom australischen Nationalarchiv veröffentlicht, aber bislang nur wenig von den klassischen Archiven rezipiert.

Die vornehme Zurückhaltung der klassischen Archive bei der Entwicklung neuer Konzeptionen zur Bewältigung der digitalen Archivierung hat natürlich Folgen. Ein Archiv, das sich heute aufmacht, um digitale Unterlagen zu übernehmen, ist gezwungen, die an anderer Stelle entwickelten Konzepte zu verwenden. Dass sich dabei Bruchstellen zu den hergebrachten Konzepten auftun, ist geradezu zwangsläufig. Wie vermitteln wir zwischen der in PREMIS abgebildeten Trennung von Informationsobjekt (intellectual entity) und Repräsentation einerseits und der noch in ISAD(G) kodifizierten unverbrüchlichen physischen und logischen Einheit des Archivals? Wie können wir die Stufenerschließung nach ISAD(G) vereinbaren mit dem für OAIS grundsätzlichen Konzept des Archivierungspakets (AIP)? Und daran anschließend: Werden wir in einigen Jahren gezwungen sein, auf eine adäquate Abbildung von Tektonik und Klassifikation in unseren Archivierungssystemen verzichten zu müssen, da es nur noch Systeme auf dem Markt gibt, die den bisherigen konzeptionellen Vorgaben der neuen digitalen Archive entsprechen und für den viel größeren Markt der Bibliotheken entwickelt wurden?

Auf dem zentralen Gebiet der digitalen Bestandserhaltung hat die internationale Fachgemeinschaft derer, die sich mit digitaler Archivierung im weiteren Sinne beschäftigen, bislang wenigstens fünf fundamentale Ansätze entwickelt:

¹ In Erinnerung an Ernst *Fuchs*, Professor der Theologie, Lehrer der Hermeneutik und Großvater des Verfassers.

² Christian *Keitel*, Archivwissenschaft zwischen Marginalisierung und Neubeginn. In: *Archivar* 64 (2011), S. 33–37.

- OAIIS sieht im Preservation Planning einen von sechs zentralen Funktionsbereichen und ordnet diesem vier wesentliche Aufgaben zu: Erhaltungsstrategien und Standards entwickeln, Paketmodelle und Migrationspläne entwickeln, vorgesehene Zielgruppe beobachten und Technologie beobachten.³
- Ebenfalls von OAIIS stammt das Konzept der Designated Community. Digitale Archive sollten ihr gesamtes Handeln an den anzunehmenden Interessen ihrer künftigen Nutzer ausrichten. So unbestritten dieses Konzept seit der ersten Veröffentlichung im Jahr 2002 auch ist, es gibt doch recht wenige Überlegungen, wie dies in die Praxis umgesetzt werden könnte.
- 1999 schon stellte das CEDARS Projekt seine Überlegungen zu den Significant Properties vor.⁴ Durch den von Margret Hedstrom und Christopher Cal Lee 2002 auf dem DLM-Forum in Barcelona gehaltenen Vortrag fand das Konzept breite Resonanz in einschlägigen Fachkreisen und 2008 im Inspect Projekt eine erste konzeptionelle Umsetzung.⁵
- Dass im Kern die von Menschen wahrnehmbare Information erhalten werden sollte, hat bereits OAIIS festgehalten. Deutlicher formuliert wird es im ebenfalls 2002 erschienenen Performance Model der National Archives of Australia.⁶
- Schließlich zählt auch der 2005 veröffentlichte PREMIS-Standard mit Begriffen wie *Intellectual Entity* und *Representation* zur Gruppe der Konzepte, die für die Erhaltung digitaler Objekte einschlägig sind.⁷

Die fünf aufgeführten Konzepte wurden mit unterschiedlicher Intensität von der internationalen Fachdiskussion diskutiert. Dennoch konnte bislang kein übergreifendes Modell vorgelegt werden, das diese unterschiedlichen Ansätze unter einem Dach vereinigt hätte.⁸ Man bediente sich vielmehr aus dieser Fundgrube unterschiedlichster Vorstellungen manchmal hier und manchmal da und hoffte darauf, dass sich auf die Dauer schon ein einheitliches Vorgehen bei der Erhaltung digitaler Objekte entwickeln würde.

Digitale Archivierung ist derweilen dabei, von der zwar Angst machenden, aber doch in fernerer Zukunft liegenden Bedrohung zu einer unmittelbar anstehenden und ganz konkreten Aufgabe zu werden. Nun müssen die staatlichen Archive die digitalen Unterlagen bewerten und

³ Referenzmodell für ein Offenes Archiv-Informationssystem, Deutsche Übersetzung der nestor-Arbeitsgruppe OAIIS-Übersetzung / Terminologie, nestor-materialien 16, Frankfurt/Main 2012

⁴ Vgl. Kelly Russell: Digital Preservation and the Cedars Project Experience, <http://www.worldcat.org/arcviewer/1/OCC/2007/09/28/0000073852/viewer/file507.html>. Alle Abrufe wurden am 17.8.2012 überprüft.

⁵ Margaret Hedstrom und Christopher Cal Lee: Significant properties of digital objects: definitions, applications, implications. Proceedings of the DLM-Forum 2002, INSAR Supplement VII, S. 218–227. Zum Inspect-Projekt s. <http://www.significantproperties.org.uk/>.

⁶ Helen Heslop, Simon Davis und Andrew Wilson: An approach to the preservation of digital records. Canberra 2002, http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm16-47161.pdf.

⁷ Premis [Preservation Metadata: Implementation Strategies]: Version 2.1. (Ohne Ort) 2011, <http://www.loc.gov/standards/premis/version-2-1-announcement.html>.

⁸ Auch das PLANETS Projekt hat seinen Schwerpunkt darauf gelegt, ein Verfahren zur Bestimmung der zum gegenwärtigen Zeitpunkt am besten geeigneten Formate zu entwickeln, wodurch die Kohärenz der heutigen mit den künftigen Erhaltungsbemühungen aber nicht gesichert wird.

übernehmen, für die sie gemäß den Archivgesetzen bereits seit über zwanzig Jahren zuständig sind. Für die kommunalen und kirchlichen Archive, die Archive der freien Wirtschaft, des Rundfunks und aus anderen Bereichen kann ähnliches gesagt werden. Die Aufgabe der digitalen Archivierung steht allorten auf der Tagesordnung. Und mit dem Einstieg in die digitale Archivierung stellt sich erneut die Frage, welche Konzepte dabei aufzugreifen sind.

Gerade auf dem Gebiet der digitalen Bestandserhaltung ergeben diese Konzepte ein vielstimmiges und schwer auflösbares Bild. Wie kann beispielsweise zwischen der am Objekt haftenden Performance und den unabhängig vom Objekt formulierten Wünschen künftiger Nutzer vermittelt werden? Die einfachsten Lösungen sind nun, einfach zu beginnen und darauf zu hoffen, dass sich die offenen Fragen mit wachsender Erfahrung quasi von selbst klären. Das mit diesem Weg verbundene Risiko liegt zunächst darin, dass die klassischen Archive gezwungen sein könnten, Konzepte einzusetzen, die von anderen Einrichtungen für ganz andere Kontexte entwickelt wurden. Es gibt aber auch ganz praktische Schwierigkeiten, die es nahelegen, diesen Weg nicht zu favorisieren. Da ist zunächst die Frage eines in sich stimmigen kohärenten Vorgehens. Bei der digitalen Bestandserhaltung müssen die Archive immer wieder Entscheidungen treffen: Welches Dateiformat, welches Zeichenformat soll für die nächste Migrationsstufe gewählt werden? Auf was kann verzichtet werden, wenn nicht alles erhalten werden kann? Wie kann ein Archiv gewährleisten, dass die Nutzungsmöglichkeiten, die ein Archivar bei der archivischen Bewertung gesehen hat, auch nach Jahrhunderten noch bestehen? Wie gewährleisten wir also über viele Jahrzehnte einen einheitlichen, möglichst widerspruchsfreien Prozess bei der Bestandserhaltung? Dabei unterscheidet sich die digitale Bestandserhaltung durch die zahlreichen, regelmäßig wiederkehrenden Entscheidungen sehr deutlich von der Bestandserhaltung von Papier und Pergament. An die Stelle der passiven Bestandserhaltung tritt eine aktive Bestandserhaltung.

Die Kette der Entscheidungen beginnt spätestens mit der Bewertung der digitalen Unterlagen durch den Archivar.⁹ Sie beginnt also bereits vor der Übernahme. Damit ist den digitalen Archiven die Frage der Kohärenz in der digitalen Bestandserhaltung von Beginn an eingeschrieben. Nun ist es vielleicht denkbar, dass sich die Archivare eines digitalen Archivs bestimmte Gedanken zur Erhaltung ihrer digitalen Archivalien machen, die sie dann schriftlich oder (wahrscheinlicher) mündlich an ihre Nachfolger tradieren. Dies mag mit einigen wenigen Objekten noch funktionieren. Es wird aber nicht mehr funktionieren, wenn Millionen digitaler Archivalien jeweils adäquat und kohärent erhalten werden sollen. Für eine solche Aufgabe sollte dann doch ein einheitliches Konzept vorliegen, an dem sich die digitale Bestandserhaltung orientieren kann. Darüber hinaus wird es Menschen nicht mehr möglich sein, die anstehenden Aufgaben alle selbstständig umzusetzen. Die Umsetzung der getroffenen Entscheidungen muss daher an Maschinen (Computer) übertragen werden. Computer können dies nur vornehmen, wenn sie eindeutige und damit programmierbare Anweisungen bekommen, was ebenfalls von der Existenz eines Konzepts für die digitale Bestandserhaltung abhängig ist.

⁹ Denkbar ist es auch, diese Kette bereits mit der Einführung der digitalen Systeme beginnen zu lassen, in denen die digitalen Unterlagen dann später entstehen sollen. Da Archivare hierauf aber in aller Regel und entgegen verschiedentlich geäußerten Hoffnungen nur einen sehr beschränkten Einfluss haben, soll dieser Aspekt hier nicht weiter verfolgt werden.

Neben der schieren Menge ist es auch noch aus einem weiteren Grund ratsam, sich bereits heute auf die Suche nach einem einheitlichen Konzept zur digitalen Bestandserhaltung zu machen. Jede Entscheidung in der digitalen Bestandserhaltung birgt auch die Gefahr in sich, die Inhalte des Archivals irreversibel zu verändern. Auch besteht die Möglichkeit, die aus Gründen der Bestandserhaltung notwendigen Veränderungen nicht ausreichend zu dokumentieren, so dass zu einem späteren Zeitpunkt nicht mehr gesagt werden kann, ob noch ein authentisches Archival vorliegt. Da also die Authentizität der Archivalien bereits ganz zu Beginn der digitalen Archivierung durch die Erhaltungsmaßnahmen gefährdet werden kann, bedarf es klarer Vorstellungen, an was sich die Authentizität der digitalen Archivalien bemessen und wie diese auch erhalten werden kann.

Zusammenfassend lässt sich daher feststellen, dass es für digitale Archive gute Gründe gibt, bereits jetzt Ausschau nach einem einheitlichen Konzept zur digitalen Bestandserhaltung zu halten. Im Oktober 2011 hat die nestor-AG Digitale Bestandserhaltung nun einen *Leitfaden zur digitalen Bestandserhaltung* zur öffentlichen Kommentierung veröffentlicht, der erstmals ein Modell zur Integration der oben kurz angesprochenen fünf Konzepte anbietet. Der Leitfaden wurde von Archivaren, Bibliothekaren und Mitarbeitern anderer Einrichtungen verfasst. Durch die interdisziplinäre Erarbeitung stellt er einen Kompromiss zwischen den Ansichten seiner Mitglieder dar. Als Rahmenstandard ist er zugleich von seinem Anspruch her für verschiedene Umsetzungen offen. Damit stellt sich die Frage, wie er in den klassischen Archiven umgesetzt werden kann. Eine Antwort hierauf kann nur eine breitere Diskussion unter den klassischen Archiven geben. Auch besteht für diese noch die Möglichkeit, ihre Vorstellungen bereits frühzeitig in die Entwicklung des Konzepts einzubringen und ebenso rechtzeitig die Auswirkungen des Konzepts auf die bisherigen Standards und Prozesse zu prüfen.

Darüber hinaus gibt es zwei weitere Gründe für die klassischen Archive, sich jetzt ganz konkret mit dem Leitfaden zur digitalen Bestandserhaltung zu beschäftigen. Zum einen hilft die Beschäftigung mit einem übergreifenden Modell wie dem Leitfaden zur digitalen Bestandserhaltung jedem einzelnen Archiv, die eigenen Anforderungen zu konkretisieren. Zweitens sind viele Archive gerade erst dabei, tastend ihren Weg in die digitale Archivierung zu finden. Die hierfür notwendigen Aufwände können beträchtlich reduziert werden, wenn sich die klassischen Archive wenigstens in Grundzügen auf ein einheitliches Modell verständigen, wie sie die digitale Bestandserhaltung verstehen und umsetzen möchten. Die daraufhin entwickelten Vorstellungen und Softwaretools können dann auch von anderen Archiven eingesetzt werden. Es spart also auch Geld, wenn man sich rechtzeitig mit diesen konzeptionellen Fragen beschäftigt.

Der Leitfaden zur digitalen Bestandserhaltung beschreibt die Prozesse, die bei der Verarbeitung der gespeicherten Daten anstehen und schlägt zweitens vor, wie diese klassifiziert und auf eine systematische Art und Weise erhalten werden können. Es geht also um Verarbeitungs- und Erhaltungsprozesse.

Verarbeitungsprozesse

Zwischen den nackten unverarbeiteten Dateien, die auf einem Datenträger gespeichert sind und der Feststellung eines Archivars, diese oder jene Informationen sollten archiviert werden, liegen verschiedene Verarbeitungsprozesse, die im Leitfaden erstmals in einem Zusammenhang dargestellt werden. Dabei erfolgt die Verarbeitung zunächst durch eine Maschine (den Computer), dann durch den Menschen selbst. Diese Verarbeitung beginnt als Wahrnehmungsprozess. Mit der Darstellung der Stationen dieses Prozesses wird deutlich, welche Eingriffsmöglichkeiten und -notwendigkeiten bestehen. Folgende Stationen lassen sich unterscheiden:

- Die physischen Daten: Sie müssen vom Archiv auf jeden Fall erhalten werden, um die nachfolgenden Verarbeitungsschritte erst möglich zu machen.
- Die verarbeitende Maschine, also Hard- und Software des Computers. Dieser Bereich entwickelt sich aus Sicht der Archive mit geradezu naturgesetzlicher Zwangsläufigkeit. Entsprechend gering sind die Einwirkungsmöglichkeiten der Archive. Stattdessen müssen sie die von den Herstellern angebotenen Systeme als gegeben hinnehmen.
- Die Performance, also die Ausgabe der verarbeiteten Daten auf einem Monitor, durch einen Lautsprecher oder ein Eingabegerät für Computerspiele. Nun erst werden die Inhalte für einen Menschen über seine Sinne wahrnehmbar. Digitale Archivierung versucht, die Fähigkeit zu erhalten, ähnliche Performances auch in Zukunft reproduzieren zu können.
- Die wahrgenommenen Sinneseindrücke werden vom Rezipienten verarbeitet. Dieser integriert die Inhalte in sein bereits bestehendes Vorverständnis¹⁰ und kann sie dadurch verstehen. Es entsteht ein Informationsobjekt.

Welche Vorteile ergeben sich durch eine derartige kleinteilige Darstellungsweise?

- Sie definiert die Aufgabe der Archive näher. Egal ob das Archiv nun eine Migrations- oder eine Emulationsstrategie verfolgt: Es strebt danach, seine Fähigkeit zu erhalten, eine der ersten Performance ähnliche aktuelle Performance herstellen zu können. Die erste Performance ist in diesem Zusammenhang diejenige, die ein Archivar bei der archivischen Bewertung vorfand.
- Die Objekte an den beiden Enden des Prozesses, also Repräsentationen und Informationsobjekte, können klar voneinander unterschieden und archivisch angesprochen (bearbeitet) werden.
- Durch die Unterscheidung zwischen Performance und Informationsobjekt kann der unvermeidliche subjektive Anteil des Archivars genauer bestimmt werden.
- Die Unterscheidung zwischen Performance und Informationsobjekt ermöglicht es, auch die nicht sinnlich erfassbaren Elemente der Informationsobjekte konzeptionell zu greifen. Von einer Datenbank lassen sich beispielsweise in der Performance nur sehr geringe Teile sinnlich wahrnehmen. Erst durch die intellektuelle Durchdringung der mit

¹⁰ Damit hält die Hermeneutik, die bislang eher in Literaturwissenschaft, Philosophie und Theologie diskutiert wurde, Einzug in die archivfachlichen Diskurse. Vgl. Hans-Georg *Gadamer*: *Wahrheit und Methode*. Tübingen 1960 und Ernst *Fuchs*: *Hermeneutik*, Stuttgart/Bad Cannstatt 1954.

der Datenbank einhergehenden Möglichkeiten bekommt der Archivar eine Vorstellung davon, welches Objekt zur archivischen Bewertung und mithin zur Archivierung ansteht.

- Die Performance wird zu dem Ort, an dem die erhaltenswerten Eigenschaften vor und nach einer Migration gemessen und verglichen werden können. Hier müssen die signifikanten Eigenschaften benannt werden. Damit wird auch klar, dass ein Großteil der Tools, die derzeit Eigenschaften von digitalen Objekten erheben, sich nur auf die Datenebene bezieht. Entsprechend deutlich wird der noch ausstehende Forschungs- und Entwicklungsbedarf.
- Migrations- und Emulationsstrategie können in einem Modell dargestellt und damit in ihren Leistungen verglichen werden.

Der Prozess grenzt aber nicht nur ab, er integriert auch. Denn jede Änderung der Daten im Zuge der Migrationsstrategie oder der Software im Zuge der Emulationsstrategie führt zu einer anderen Performance. Dem Wahrnehmungsprozess schließen sich Erhaltungsprozesse an, die letztlich wiederum zu einer Änderung der Daten und/oder der Software führen.

Erhaltungsprozesse

Einzelne digitale Archivalien können wahrscheinlich ohne größere konzeptionelle Vorstellungen eine Zeit lang erhalten werden. Probleme entstehen erst, wenn die verschiedenen Erhaltungsbestrebungen über einen sehr langen Zeitraum konsistent gehalten werden müssen, wenn also der Nachfolger des heutigen Archivars im Objekt immer noch das sehen sollte, das der zuerst mit der Erhaltung betraute Archivar in ihm sah. Dieser Nachfolger benötigt dafür Hinweise, wie denn die ursprüngliche Performance aussah. Probleme können aber auch früher auftreten. Wenn nämlich nicht nur eine Handvoll digitaler Objekte, sondern Hunderttausende und Millionen dieser Objekte erhalten werden sollen. Nun soll der Computer wissen, wie er den heutigen Archivar zu unterstützen hat. Beide, der Nachfolger und auch der als Assistent auftretende Computer, benötigen eindeutige Angaben, um ihre Arbeit vornehmen zu können.

Um effizient vorzugehen und eine von Anfang an einheitliche Bestandserhaltung zu gewährleisten, sollten die Hinweise bereits während der Bewertung erfasst werden. Dabei ist es sowohl zu ihrer leichteren Erfassung als auch für eine später möglichst effiziente Verarbeitung naheliegend, die Hinweise in strukturierter Form zu erfassen. Es bedarf also einer Klassifikation dieser Hinweise.

Von welchen Ausgangspunkten können die Hinweise nun abgeleitet werden? Da sind zunächst natürlich die Objektarten selbst. Mit ihren Eigenschaften bestimmen sie den Rahmen, in dem künftige Nutzer im Archiv Informationen finden und verarbeiten können. Diese Objekte müssen, wie oben ausgeführt, durch eine Kette aktiver Entscheidungen des Archivars erhalten werden. Zumeist gibt es verschiedene Möglichkeiten der Erhaltung. Es muss also eine Entscheidung getroffen werden, die sich nicht eindeutig aus den Eigenschaften des Objekts ableiten lässt. Folglich müssen noch andere, außerhalb der Objekte liegende Kriterien hinzugezogen werden, um die Hinweise zur Erhaltung definieren zu können. Limitierende Faktoren sind hier zunächst die finanziellen Möglichkeiten des Archivs oder der interessierten Fachgemeinde. Ebenso wichtig sind die Erwartungen, die die künftigen Nutzer voraussichtlich an die Archivalien richten

werden. Es gilt also, diese Hinweise durch Berücksichtigung der von den Archivalien mitgebrachten Eigenschaften, der finanziellen Möglichkeiten und der anzunehmenden Erwartungen künftiger Nutzer strukturiert zu erfassen.

Der Leitfaden zur digitalen Bestandserhaltung sieht zunächst eine grobe Klassifikation digitaler Unterlagen vor. In diesem ersten Schritt werden anhand der in der Performance aufscheinenden Eigenschaften erste Gruppen gebildet. Diese Gruppen können mit dem Leitfaden zur digitalen Bestandserhaltung als *Informationstypen* bezeichnet werden. Beispielsweise könnten Akten, schwach strukturierte Daten, Fachverfahren, Webseiten und Audio- bzw. Videoobjekte unterschieden werden. Denkbar wäre es aber auch, die Fachverfahren in einem weiteren Sinne in Statistikverfahren einerseits und Fachverfahren im engeren Sinne zu unterscheiden. Auch könnten aus letzteren noch die geographischen Informationen (landläufig Geodaten) ausgegliedert werden, die wiederum in *Geodaten der Vermessungsverwaltung* und *Geodaten aus Fachverwaltungen* oder in *Geobasisdaten* und *Geofachdaten* aufgeteilt werden könnten. Diese Aufteilungen geben eine grobe erste Einteilung vor. Es wäre wünschenswert, wenn hierüber eine breitere Diskussion zwischen den Archiven einsetzen könnte. Zugleich sind diese Einteilungen auch abhängig von den von den einzelnen Archiven verfolgten Zielen. Sie können daher per definitionem nicht ein für alle Mal für alle Archive verbindlich festgelegt werden.

Die Festlegung der Informationstypen macht das Feld der zu erhaltenden Objekte überschaubarer. Sie genügt aber nicht, um die notwendigen Hinweise geben zu können. Am deutlichsten wird dies bei den Eigenschaften. Archive können aus verschiedenen Gründen nicht alle Eigenschaften übernommener Objekte erhalten. Zunächst ist dies in aller Regel schon aus finanziellen Gründen nicht möglich. Hinzu kommt noch die fehlende Dokumentation heute verbreiteter Dateiformate, wie dies beispielsweise sehr schön bei den Microsoft Office Formaten beobachtet werden kann. Selbst wenn also unbegrenzte Mittel zur Verfügung stünden, könnte daher nicht alles erhalten werden, da nicht alle Eigenschaften der Dateiformate dokumentiert sind. Es muss also ausgewählt werden. Die Teilmenge der zu erhaltenden Eigenschaften sind die signifikanten Eigenschaften. Sie müssen in einem zweiten Schritt ermittelt werden.

Die signifikanten Eigenschaften können nur durch einen bewussten Auswahlprozess festgelegt werden. Jede Entscheidung für die Eigenschaften A, B und C ist häufig eine Entscheidung gegen die Eigenschaften D, E und F. Hier kommen nun im verstärkten Maße die Faktoren zum Tragen, die außerhalb der einzelnen Archivalien liegen. Was ist finanzierbar? Was ist ganz praktisch umsetzbar? Was ist vor allem im Interesse der künftigen Nutzer?

Gerade die Interessen der künftigen Nutzer lassen sich nicht ganz einfach benennen. Wer kann schon in die Zukunft sehen? Manche Annahmen sind nicht möglich: Wir wissen nicht, ob in Zukunft ein Familienforscher Vorfahren mit dem Namen Schmidt, Müller oder Mayer sucht. Andere Annahmen können dagegen mit größerer Wahrscheinlichkeit getroffen werden. Die nestor AG digitale Bestandserhaltung geht davon aus, dass alle Nutzungen digitaler Archivalien sich letztlich auf eine von vier Grundformen zurückführen lassen:

- Betrachtung oder Wahrnehmung des Gesamtobjekts (z.B. Lesen eines Romans oder Ansehen eines Films)
- Suche nach einer spezifischen Information (z.B. über die Volltextsuche)
- Einlesen des Archivals in eine externe Software zur Weiterverarbeitung
- Abspielen des Archivals (nur bei Software, z.B. bei Computerspielen)

Bei einem Archiv können nun sehr unterschiedliche Informationsobjekte vom selben Informationstyp ankommen. Zumeist dürften für die Informationsobjekte eines Informationstyps unterschiedliche Nutzungsszenarien vorstellbar sein. Für die Informationsobjekte des Typs Text ist es z.B. denkbar, dass manche vor allem als ganzes Objekt genutzt werden sollen (z.B. ein typographisch aufwändig gestaltetes Gedicht), während andere außerdem noch auf einzelne Zeichenfolgen hin durchsuchbar sein sollten. Es ist daher notwendig, die Informationstypen weiter zu unterteilen. Dabei bietet es sich an, alle Informationsobjekte mit denselben signifikanten Eigenschaften in eine gemeinsame Gruppe zu bringen, da sie aufgrund ihrer signifikanten Eigenschaften auf dieselbe Art und Weise erhalten werden können. Diese Gruppe nennt der Leitfaden zur digitalen Bestandserhaltung *Erhaltungsgruppe*.

Die signifikanten Eigenschaften können als Bedingungen verstanden werden, die von künftigen Dateiformaten erfüllt werden müssen. Um noch einmal das soeben genannte Beispiel des Informationstyps Text aufzugreifen, könnten die Objekte, die nur als Gesamtobjekt genutzt werden (das Gedicht), auch als Bild archiviert werden, während sehr umfangreiche Strafverfahrensakten vermutlich nur unter Zuhilfenahme der Volltextsuche sinnvoll genutzt werden können. Diese Objekte stellen daher an die künftigen Dateiformate andere Anforderungen als das digitale Gedicht.

	Reihenfolge der Zeichen	Verarbeitbarkeit	Auszeichnung	Adressierbarkeit	Textebenen	Mögliche Formate (2012)	Mögliche Formate (2100)	Mögliche Formate (2200)
Text lesen	x					TIFF
Information suchen	x	x				PDF/A
Text weiterverarbeiten	x	x	x	x	x	XML

Tabelle 1: Mögliche Erhaltungsgruppen für Text, Annahme: Eine Erhaltungsgruppe pro Nutzungsziel. Links die Nutzungsziele, oben die möglichen signifikanten Eigenschaften bzw. rechts davon die daraus ableitbaren Dateiformate.

	Reihenfolge der Zeichen	Verarbeitbarkeit	Auszeichnung	Adressierbarkeit	Textebenen	Mögliche Formate (2012)	Mögliche Formate (2100)	Mögliche Formate (2200)
Text lesen	x					XML
Information suchen	x	x						
Text weiterverarbeiten	x	x	x	x	x			

Tabelle 2: Mögliche Erhaltungsgruppen für Text, Annahme: Eine Erhaltungsgruppe für drei Nutzungsziele. Links die Nutzungsziele, oben die möglichen signifikanten Eigenschaften bzw. rechts davon die daraus ableitbaren Dateiformate.

Die signifikanten Eigenschaften werden mit der Bewertung festgelegt. Sie gelten für die gesamte Lebenszeit des Archivals. Wenn die Dokumente zu einem Strafverfahren in einer zeichenbasierten (und damit im Volltext durchsuchbaren) Form übernommen wurden, sollte diese Durchsuchbarkeit auch nach der zweiten, dritten und vierten Dateiformatsmigration noch gegeben sein. Ähnliches könnte von der Verteilung der Farben und Formen in einem Bild, der Zahl der Datensätze in einer Tabelle und der Länge einer Audiodatei gesagt werden. Nach jeder Migration müssen diese Werte neu erhoben und mit den übernommenen Werten verglichen werden. Wenn sie voneinander abweichen, ist das gewählte Dateiformat nicht geeignet, um die definierten signifikanten Eigenschaften zu unterstützen. Die signifikanten Eigenschaften stellen damit den Maßstab, vor dem sich alle künftigen Dateiformate bewähren müssen. Dabei ist es von Vorteil, dass dieser Maßstab nur einmal (ganz am Anfang) festgelegt werden muss.

Nach der Festlegung der signifikanten Eigenschaften sollte das Archiv noch bestimmen, welche dieser Eigenschaften absolut und welche nur zu bestimmten (definierten) Graden erfüllt werden müssen. Die Zahl der Datensätze einer Tabelle muss sicherlich vollständig erhalten werden. Dagegen sind im Farbraum eines Bildes durchaus Abweichungen denkbar.

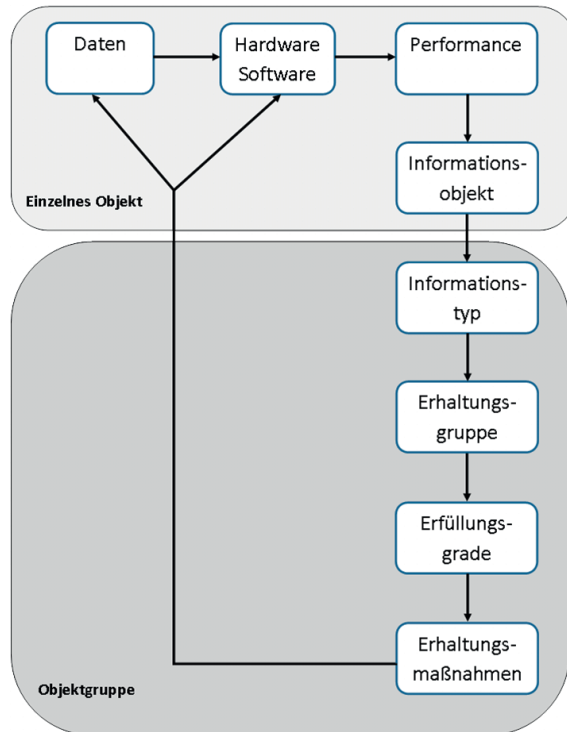


Abb. 1: Prozessfolge bei der digitalen Bestandserhaltung: Das einzelne Objekt wird wahrgenommen, für die Objektgruppe werden Festlegungen getroffen.

Neben den signifikanten Eigenschaften stellt der Leitfaden zur digitalen Bestandserhaltung noch weitere Bereiche vor, die bei der Erhaltung digitaler Unterlagen wichtig sind. Er beschreibt die von OAIIS erstmals definierte Representation Information und grenzt sie von den signifikanten Eigenschaften ab¹¹. Er beschreibt konkret, wie sich sowohl die Migrationsstrategie als auch die Emulationsstrategie mit dem definierten Instrumentarium umsetzen lässt. Und er enthält Hinweise zu den von OAIIS vorgesehenen Aufgaben zur Beobachtung der Nutzergruppen und der technischen Entwicklungen.

Authentizität

Die signifikanten Eigenschaften ermöglichen es, bei verschiedenen Prozessen der digitalen Archivierung Entscheidungen zu priorisieren. Dateiformate, die alle Eigenschaften des Objekts unterstützen, sind jenen vorzuziehen, die nur die signifikanten Eigenschaften unterstützen. Es ist wichtiger, die signifikanten Eigenschaften eines übernommenen Objekts zu sichern, als die übrigen Eigenschaften dieses Objekts. Solche Dateiformate, die alle signifikanten Eigenschaften unterstützen, sind jenen vorzuziehen, die nur manche unterstützen et cetera.

Die signifikanten Eigenschaften stellen also die Messlatte dar, an der sich die Erhaltungsmaßnahmen bewähren müssen. Es ist daher nur konsequent, an ihnen auch die Anforderungen im Bereich der Authentizität festzumachen. Die Werte, die bei der Überprüfung der signifikanten Eigenschaften eines Informationsobjekts erhoben wurden, werden vom digitalen Archiv zusammen mit den anderen Metadaten dieses Objekts gespeichert. Im Normalfall dürfte dies bei fast allen digitalen Archiven das Dateiformat XML sein. Diese in XML gespeicherten Informationen können relativ leicht in ein künftiges Nachfolgeformat übertragen werden. Wir wissen daher noch in fernerer Zukunft, dass auch in der dritten Repräsentation eines Audioobjekts dessen Länge 7 Minuten betrug. Wir wissen dies, obwohl die Primärdaten dieser dritten Repräsentation von den dann gängigen Computersystemen schon längst nicht mehr gelesen werden können. Damit geben die zu den signifikanten Eigenschaften erhobenen Daten einen lückenlosen Aufschluss darüber, wie dieses Objekt erhalten wurde. Abgesehen von dieser Nachweisfunktion definieren die signifikanten Eigenschaften aber auch den Rahmen dessen, was das digitale Archiv von dem Informationsobjekt auf jeden Fall erhalten wollte. Die signifikanten Eigenschaften stellen daher auch für den Nachweis der Authentizität die zentrale Messlatte.

¹¹ Im Umfeld des PLANETS Projekts wurde diskutiert, inwieweit sich diese beiden Metadaten Gruppen voneinander unterscheiden lassen. Representation Information umfasst die für eine konkrete Nutzung erforderlichen Metadaten. Signifikante Eigenschaften stellen dagegen die Norm, gegenüber der sich alle künftigen Performances bewähren müssen.

Resümee

Der Leitfaden zur digitalen Bestandserhaltung entwirft ein konsistentes Modell zur digitalen Archivierung. Die geschilderten Prozesse sollen eine Verarbeitung auch sehr großer Mengen digitaler Archivalien ermöglichen. Mit der Bewertung setzen sie bereits sehr früh im Lebenszyklus digitaler Objekte ein. Es liegt daher im Interesse der Archive, sich mit dem Leitfaden auseinanderzusetzen und eine eigenständige Position oder auch eine konkrete Umsetzung zu entwickeln.