

Was sollen wir erhalten? Anmerkungen zu den Eigenschaften digitaler Objekte

1. Einführung

Die Archivierung elektronischer Dokumente stellt neue Anforderungen an die Archivarinnen und Archivare (im Folgenden wird in allen Pluralformen die männliche Form als grammatischer Plural verwendet). Die nationale und internationale Diskussion der vergangenen Jahre drehte sich um Fragen im Zusammenhang mit der Gewährleistung von Lesbarkeit, Zugänglichkeit und Verkehrsfähigkeit der Daten sowie um Fragen der Speicherung. Hinzu kommen organisatorisch-konzeptionelle Aspekte, die insbesondere die Datensicherheit, Authentizität und Integrität betreffen. Nur ein abgestimmtes Vorgehen unter Berücksichtigung aller genannten Faktoren erlaubt den langfristigen Erhalt der elektronischen Dokumente.

Viele Kollegen betraten archivarisches Neuland und setzten sich intensiv mit dem Feld der Informationstechnologie (IT) auseinander. Inzwischen wird niemand mehr ernsthaft anzweifeln, dass Archivare künftig auch auf diesem Gebiet ein profundes Grundwissen benötigen, um die Anforderungen bei der Übernahme und Archivierung elektronischer Daten zu erfüllen. Auf der Basis der für die Archivierung von Unterlagen geltenden Prinzipien und unter Einbeziehung ihrer besonderen Perspektive stellen Archivare in diesem Zusammenhang andere Fragen und legen andere Schwerpunkte, als die Kollegen der IT.

Eine Arbeitsgruppe des VdW-Arbeitskreises "Elektronische Archivierung" hat sich mit dem elektronischen Objekt auseinandersetzt und gefragt, was davon erhaltenswert ist und wie man für die Erhaltung sorgen kann. Ausgangspunkte waren insbesondere folgende Fragestellungen: Was eigentlich ist das digitale Objekt, das ins Archiv übernommen werden soll? Aus welchen Teilen besteht es? Welche Eigenschaften besitzt es? Welche seiner Eigenschaften müssen unbedingt erhalten werden? Und schließlich: Wie können diese als bedeutsam definierten Eigenschaften des digitalen Objekts dauerhaft gesichert werden?

Damit stellt sich die Bewertungsfrage bei elektronischen Objekten in einer neuen, bislang erst in Ansätzen diskutierten Form: Soll ein zur Übernahme angebotenes digitales Objekt überhaupt archiviert werden? Und wenn ja, was von diesem Objekt soll eigentlich aufbewahrt werden? Gerade weil diese Fragen den unmittelbaren Kompetenzbereich der Archivare betreffen, verlangt ihre Beantwortung deren besondere Aufmerksamkeit.

Dieser Text fasst die Ergebnisse der in der Arbeitsgruppe angestellten Überlegungen zusammen. Er will Anregungen für eine archivfachliche Auseinandersetzung mit dieser komplexen Thematik geben und für die neuen Fragen und Herausforderungen, die sich im Zusammenhang mit der Archivierung elektronischer Dokumente stellen, sensibilisieren. Dadurch, so die Intention der Mitglieder der Arbeitsgruppe, verbessert sich auch die Kommunikationsfähigkeit von Archivaren in der Diskussion mit den IT-Kollegen.

Neben der Beschreibung des digitalen Archivguts und einer dabei entstehenden Skizze der Eigenschaften elektronischer Objekte sollten auch Perspektiven für ein rationelles Vorgehen in den Archiven aufgezeigt werden. Hier bietet sich durch die Klassifikation der elektronischen Objekte eine Möglichkeit, Gleiches gleich zu behandeln. Die Interessen der Nutzergemeinschaft, der so genannten Designated Community, dienen als Orientierung.

Im Verlaufe der Diskussionen in der Arbeitsgruppe wurde immer deutlicher, dass es eines längerfristigen Lernprozesses bedarf, um die unterschiedlichen Sichtweisen und Anforderungen in ein stimmiges Modell zu integrieren.

2. Das digitale Objekt

Die Archivierung digital gespeicherter Informationen ist nur durch Anpassungen möglich. Die derzeit allgemein favorisierte Strategie, die Migration, führt dazu, dass die Informationen in immer wieder neuer und anderer physischer Form (Datenobjekte) abgelegt wird, weil es keine dauerhaft haltbaren Datenträger, Hard- und Software gibt. Es entstehen verschiedene Ausprägungen der Informationsobjekte; diese Ausprägungen werden auch Repräsentationen genannt.

Bislang waren die Archivare daran gewöhnt, Informationen und Informationsträger als unauflösbare Einheit zu verstehen, die als "Original" bezeichnet wurde. Dennoch zeigen verschiedene Beispiele, dass auch analoge Informationsobjekte in verschiedenen Repräsentationsformen vorliegen können: Der Abdruck eines ägyptischen Hieroglyphensteins, Abschriften von Texten, Abzüge von Fotografien oder Repliken historischer Gegenstände – stets stellt sich die Frage nach dem Verhältnis der Repräsentation zum Original und nach der Vollständigkeit der aus dem Original übertragenen Informationen.

Elektronisch gespeicherte Informationen können nur über Datenobjekte erhalten werden, die selbst nicht dauerhaft überliefert werden können, sondern immer wieder ersetzt werden müssen. Es entsteht so eine Zuordnung von einem Informationsobjekt zu vielen Ausprägungen.

Informationsobjekt (z.B. Text)

-----|-----
1 2 3 4 5 (Datenobjekte)
doc Tiff pdf pdf/A ...

Grundsätzlich sollten Archive Mehrfachüberlieferung vermeiden. Da wir uns gegenwärtig noch in einem Lernprozess hinsichtlich der Archivierung digitaler Objekte befinden, sollten jedoch möglichst viele Optionen für den künftigen Umgang mit digitalen Objekten offen gehalten werden. Daher erscheint es sinnvoll, zunächst auch das "Original" (z.B. die übernommene doc-Datei) oder die übernommene Form aufzubewahren. Es sollte zusätzlich mindestens eine archivierungsfähige Form als Ausgangspunkt für künftige Migrationen erzeugt werden, sofern eine solche nicht übernommen wurde.

Abgesehen von diesen eher pragmatischen Überlegungen ändern das Auseinanderfallen von Informations- und Datenobjekten auch die Bewertungsmaßnahmen, da nun zusätzlich zu der Frage, ob ein elektronisches Dokument als archiwwürdig eingestuft wird, auch geklärt werden muss, welche Eigenschaften sich in allen Ausprägungen wiederfinden sollen. Die Entscheidung impliziert umgekehrt, auf welche Eigenschaften man im Konfliktfall verzichten kann, wenn bei einer zukünftigen Migration nicht alle abgebildet werden können.

Eine Grundlage für die Entscheidung bieten die antizipierten Bedürfnisse künftiger Nutzer. Abgeleitet davon können schließlich die signifikanten Eigenschaften der Informationsobjekte benannt werden, also die Eigenschaften, die für das Objekt im Hinblick auf seine spätere Nutzung konstitutiv sind und die darum in allen Ausprägungen erhalten werden müssen. Diese bilden eine Teilmenge aller möglichen Eigenschaften eines Informationsobjekts. Neben den Vorgaben, die von den Benutzerbedürfnissen abgeleitet werden können, muss auch gefragt werden, zu welcher Objektklasse, (etwa Text, Bild, Audio usw.) die Unterlagen gehören.

Das methodische Vorgehen in der Arbeitsgruppe sah folgendermaßen aus: Zunächst erfolgte eine Klärung darüber, welche Objektklassen existieren und in welcher Beziehung einzelne Objektklassen zueinander stehen. Die Bestimmung von Objektklassen ermöglicht es, bei einer anstehenden Migration gleichartige Objekte als solches zu erkennen und dadurch automatisiert verarbeiten zu können. Nach einer Bestimmung der diesen Objektklassen zugehörigen Eigenschaften sollten daraus die jeweils signifikanten Eigenschaften herausgefiltert werden.

3. Objektklassen und ihre Eigenschaften

Die folgende Auflistung von Objektklassen umfasst die verschiedenen elektronischen Dokumente, die Archiven zur Übernahme und Archivierung angeboten werden. Jeder Objektklasse sind Eigenschaften zugeordnet. Es werden einfache oder Elementar-Objekte, z.B. Text- oder Bildobjekte, von komplexen oder Container-Objekten, etwa Websites oder e-Mails, unterschieden. Alle genannten Objekttypen lassen sich als digitales Objekt verstehen und können als solches auf elementare Objekte zurückgeführt werden. Nicht betrachtet wurden Programme.

3.1 Digitales Objekt

- enthält keine anderen Dateien, kein Container
- Bit-Stream
- Metadatenfähigkeit
- definierter Anfang und definiertes Ende
- Decodiervorschrift (... "Lies mich als"...)
- Kompositionsfähigkeit (Elementarobjekt kann andere Elemente einbetten und ist in andere einbettbar)
- plus: strukturierende Information

3.2. Elementarobjekte

Die zur Klasse der Elementarobjekte gehörenden digitalen Objekte sind Einzelobjekte. Auch Audio und Videoobjekte lassen sich bis zu einem bestimmten Grad auf ein elementares digitales Objekt zurückführen.

Textdokumente

- kodierte Information
- alphanumerische sprachliche und graphische Zeichenmenge in festgelegter Reihenfolge
- logische Struktur (Kapitel, Absatz, Listen)
- Adressierbarkeit (einzelne Strukturelemente sind ansteuerbar)
- Auszeichnung/Formatierung —> zeigt Struktur
- Verweisfähigkeit
- Hyperlink als technischer Verweis
- Sichtbarkeit versteckter Textebenen

Bilddokumente Bitmap-Datei – Rasterbild

- Zweidimensional
- festgelegte Außengrenze (i.d.R. rechteckig, äußere Form begrenzt das Bild)
- Punkt bezogene Farbe und Farbsättigung im definierten Farbraum
- Auflösung durch Anzahl der Pixel
- Pixelform
- Ebenenfähigkeit
- Ausblendbarkeit bestimmter Ebenen

Grafik/Zeichnung Vektordatei, z.B. Konstruktionszeichnung, auch 3-D-Objekt

- Koordinierungssystem muss vorhanden sein
- Algorithmus muss vorhanden sein
- Texturen müssen einzubetten sein (Oberflächenstruktur)
- Vektoren
- Objekteigenschaften sind erweiterbar

Bewegte Bilder

- lineare Abfolge von Bildern
- Bildrate
- kann Audiospur enthalten
- Mehrkanalfähigkeit
- Synchronität von Bild und Ton (festgelegte Reihenfolge der Bilder und Töne und festgelegte Gleichzeitigkeit von Bild und Ton)
- Komposition aus Elementarobjekten

Audiodateien

- lineare Abfolge von Tonsignalen
- Abtastfrequenz im definierten Klangraum
- Mehrkanalfähigkeit
- Timecode
- Bit-Tiefe (mit welcher Bitzahl wird die abgetastete Frequenz genommen, entspricht Farbtiefe und Bildschärfe bei Bildern)
- Komposition aus Elementarobjekten

3.3 Container-Objekte

Zu dieser Objektklasse gehören Objekte, die aus verschiedenen Bestandteilen (Elementarobjekten) zusammengesetzt sind, z.B. e-Mails, Websites und Datenbanken. Bei den Container-Objekten sind außer den Einzelobjekten selbst auch deren Beziehungen zueinander zu beschreiben und bei der Betrachtung der Eigenschaften zu berücksichtigen. Außerdem können auch Metadaten wie z.B. Sender und Empfänger bei e-Mails zu den Eigenschaften gezählt werden. Bei Websites ist das Verhalten (Behaviour) von Bedeutung.

Signifikante Eigenschaften

Als signifikante Eigenschaften werden die Eigenschaften bezeichnet, die für ein digitales Objekt unbedingt erhalten werden müssen, da sie als wesentlich angesehen werden müssen.

Die im angelsächsischen Raum bereits seit 1999 breit diskutierte Frage nach den signifikanten Eigenschaften (http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf Zugriff am 24.11.2009) hat mittlerweile auch den deutschsprachigen Raum erreicht:

http://www.archivtag.de/at2009/abstract_keitel.pdf

[http://nestor.sub.uni-](http://nestor.sub.uni-goettingen.de/seminar_bundesarchiv/docs/20080410_PREMIS_Objekt.pdf)

[goettingen.de/seminar_bundesarchiv/docs/20080410_PREMIS_Objekt.pdf](http://nestor.sub.uni-goettingen.de/seminar_bundesarchiv/docs/20080410_PREMIS_Objekt.pdf)

http://ikeep.com/da_problemsgoals.pc

(alle Zugriffe am 4.12.2009)

Grundsätzlich stellt sich die Frage, ob diese Eigenschaften dem Objekt generisch innewohnen oder ob sie ihm von außen zugesprochen werden. Falls Letzteres zutrifft, schließt sich die Frage an, wer die signifikanten Eigenschaften für ein digitales Objekt definiert. Für die Bewertungsdiskussion bedeutet dies, dass Archivare einerseits bewerten müssen, ob sie ein bestimmtes elektronisches Objekt überhaupt überliefern wollen und wenn ja, welche seiner Eigenschaften unbedingt erhalten werden müssen.

Der Verlust von Eigenschaften bei digitalen Objekten bedeutet immer auch den Verlust von Information. Daher sollten Archivare danach streben, möglichst alle Eigenschaften der digitalen Objekte unverändert zu erhalten. Das Ziel-Dateiformat bei der Konversion sollte beispielsweise danach gewählt werden, ob es neben einer möglichst weit reichenden Alterungsbeständigkeit möglichst alle Eigenschaften aufweist, die auch das Ursprungsformat besaß. Diese Anforderung an das Zielformat besitzt eine hohe Priorität. Nur dann, wenn kein Zielformat verfügbar ist, das alle Merkmale des Quellformats aufweist, muss über die "signifikanten Eigenschaften" entschieden werden, die auf jeden Fall erhalten werden müssen. Die signifikanten Eigenschaften dienen also als Qualitätskriterien für die Wahl von Zielformaten bzw. von alternativen Abbildungsvorschriften bei der Migration. Aus allen Eigenschaften einer Archivalienklasse definieren Archivare die Eigenschaften, die für eine künftige Nutzung unbedingt erhalten werden müssen. Dies sind die signifikanten Eigenschaften. Sie sind Teilmenge der Summe aller Eigenschaften von Archivalien einer Objektklasse und werden unterschiedlichen und sich wandelnden Benutzerinteressen gerecht. Bei künftigen Migrationen müssen sie unbedingt erhalten werden. Dabei sind signifikante Eigenschaften nicht an einzelne Formate gekoppelt, sondern werden für Objektklassen formatunabhängig definiert, z.B. Objekt ist digital, hat eine definierte Abfolge von Zeichen, hat mehrere Ebenen.

4. Signifikante Eigenschaften und ihre Anwendung im Archiv

Matrix

Die Benennung einer Eigenschaft als signifikant allein genügt als Handlungsanweisung für die Erhaltung digitaler Archivalien noch nicht. Außerdem muss festgestellt werden, in welchem Grad die Eigenschaft von künftigen Datenobjekten bzw. Ausprägungen erfüllt werden muss. Bei notwendig werdenden Veränderungen der elektronischen Daten, etwa im Zuge einer Migration, muss darum auch die Frage beantwortet werden: Was will ich mit den Daten später tun können? Welche Eigenschaften – respektive signifikanten Eigenschaften – der Daten müssen erhalten werden, um sicherzustellen, dass das gewünschte Nutzungsziel auch in Zukunft erreichbar bleibt. Die Definierung und Bewertung der signifikanten Eigenschaften verhilft nicht dazu, möglichst viel zu erhalten, sondern das Wichtige und Richtige zu erhalten. Nur so kann das Nutzungsziel dauerhaft erreicht werden. Von der Antwort auf diese Frage hängt es ab, in welcher Qualität Daten erhalten werden. Die Entscheidung des Archivs darüber, ob Daten in maximaler, mittlerer oder einfacher Qualität erhalten werden, ist von hoher Tragweite: wenn man sich für die einfache Qualität entschieden hat, ist es später nicht mehr möglich, die Qualität zu verbessern.

Am Beispiel einer Bilddatei kann ein mögliches Vorgehen nachvollzogen werden: Zunächst werden die mit dem Objekt verbundenen Eigenschaften notiert. Für jede Eigenschaft ist nun zu entscheiden, ob sie in maximaler Qualität, also der abgegebenen Datei entsprechend, oder in Abweichung von dieser in mittlerer oder einfacher Qualität überliefert werden muss und ob dies überhaupt geleistet werden kann. Einige Eigenschaften sind konstitutiv und müssen überliefert werden. Die Überlieferung in Abweichung vom Original impliziert, dass es zukünftig nur noch diese Form geben wird. Daher muss bei der Entscheidung über die Qualität der künftige Nutzungszweck unbedingt Berücksichtigung finden. Auch die Wahl eines Archivierungsformats ergibt sich wiederum aus der Entscheidung über die zu überliefernde Qualität des elektronischen Dokuments. Verringert man beispielsweise bei einer Bilddatei die Auflösung, so wird im Archiv nur diese geringere Auflösung überliefert. Der Nachdruck im Großformat ist damit künftig nicht mehr möglich. Allerdings kann mit der Datei die Information verbunden bleiben, dass das aus der Bilddatei hervorgehende Bild in der Vergangenheit als Poster

oder Plakat verwendet wurde. Die größtmögliche Varianz für die künftige Nutzung bietet derzeit die Speicherung in TIFF, andere Formate ziehen möglicherweise Qualitätsverluste der Überlieferung nach sich.

Die oben beschriebenen Eigenschaften sind nicht abschließend beschrieben. Daher kann es notwendig werden Ergänzungen vorzunehmen. In einer Matrix wie der folgenden erfolgt dann die Bewertung der Eigenschaften unter Berücksichtigung der jeweiligen Nutzungshorizonte. Ergebnis ist ein Hinweis für die Praxis, welche Eigenschaften in welcher Weise erhalten werden müssen, wenn man ein digitales Objekt in Zukunft in einer bestimmten Weise nutzen möchte.

Matrix für Bilddokument

Bilddaten – Eigenschaften		Maximale Qualität	Mittlere Qualität	Einfache Qualität
Zweidimensionale feste Außengrenze	muss			
Farbsättigung				
Exakt so erhalten				
Farbraum unbestimmt farbig				
bitonal				
Anzahl der ursprünglichen Pixel				
Wie in Ausgangsdatei				
Kann verringert werden				
Pixelform	muss			
Auflösung				
Wie in Ausgangsdatei				
Kann verkleinert werden				
Ebenenfähigkeit				
Als Einzelbitmap				
Ebenen inklusive Pfade				
Ursprüngliche Pixelinfo				

Wie in Ausgangsdatei				
Kann verändert werden				
FORMAT		TIFF		

5. Fazit

kehrt man nun zurück zur Ausgangsfrage: "Was sollen wir erhalten?", so wird deutlich, dass die Bewertung elektronischer Dokumente Aspekte berücksichtigen muss, die Archivarinnen und Archivaren ungewohnt erscheinen. Die Antizipation eines künftigen Nutzungszwecks gehört ebenso in diesen Bereich wie die Benennung und Bewertung der Eigenschaften von elektronischen Archivaldokumenten und die Benennung der als signifikant bewerteten Eigenschaften. Die in der Regel geisteswissenschaftlich ausgebildeten Beschäftigten in den Archiven stehen angesichts der technischen Aspekte der elektronischen Archivierung zweifellos vor neuen Herausforderungen. Es erscheint allerdings unabdingbar, sich gerade mit diesen Fragen zu beschäftigen. Nur dann, wenn von Seiten der Archive Klarheit darüber besteht, was erhalten werden soll, kann man Anforderungen an die IT formulieren und mit den dortigen Mitarbeitern gemeinsam nach Lösungen suchen, die sowohl die technischen wie auch die archivischen Aspekte berücksichtigen. Bewertungsfragen nehmen darum künftig auch auf technische Aspekte Bezug. Selbst wenn der notwendige Speicherplatz der Archivierung keine Grenzen mehr setzen sollte, so wird durch die begrenzte Haltbarkeit der Datenträger und die bei Konvertierungen/Migration zwangsweise entstehenden Veränderungen der Daten die Frage danach, was erhalten werden soll und in welcher Qualität eine zentrale Rolle spielen. Steigende Ausgaben für IT-Dienstleistungen, eine wachsende Komplexität der elektronischen Überlieferung bei gleichzeitig einsetzenden Sparmaßnahmen in den Archiven machen ein pragmatisches und rationelles Vorgehen notwendig, das dem Kostendruck begegnet und gleichzeitig die Interessen der Archive berücksichtigt.

An der Projektarbeit wirkten mit:

Ulrich Kamp, Stefan Przigoda, Bettina Hasselbring, Dr. Christian Keitel, Dr. Annerose Knoch, Thomas Pereira Antunes, Annette Handrich, Heinz Achatz, Kirsten Teipel, Dr. Hans-Jürgen Sproß, Kerstin Früh, Angela Ullmann, Dieter Gross, Wolfgang Richter, Dr. Christian Hillen, Dr. Britta Leise, Tomislav Novoselac, Kai Franke, Romy Meyer, Dr. Christoph Popp, Dr. Martin Burkhard, Dr. Ulrike Gutzmann