

Dienstleistung via Dokumentenmanagement am Beispiel der Mieterakten der GBG - Mannheimer Wohnungsbau mbH: konzeptionelle Überlegungen zur Frage der Langzeitarchivierung

Christoph Popp

Im Anschluß an die Präsentation der Schriftgutverwaltung der GBG darf ich Ihnen einen Ausblick auf ein Projekt geben, mit dem das Stadtarchiv Mannheim in Zusammenarbeit mit der GBG eine Strategie ausarbeiten will, wie die in der Vergangenheit und Gegenwart entstehenden elektronischen Unterlagen dauerhaft archiviert und nutzbar gemacht werden können. Dieses Vorhaben ist ein Teil des Projekts ELAN "Elektronische Archivierung und Nutzung".¹

Archivtheoretische Überlegungen haben bislang vier Lösungsstrategien herausgearbeitet: das Museumsmodell, die Emulation, die Migration und die Konversion. Der Lösungsansatz Museumsmodell, die jeweilige Hard- und Software gemeinsam mit den zu archivierenden Daten aufzubewahren, wird hier nicht weiter verfolgt: er würde die Archive zu technischen Museen mit unabwägbareren Folgekosten werden lassen. Ähnliches gilt für die im Generationswechsel anstehende Migration von Daten, Programmen und Programmumgebungen.

Ähnlich unpraktikabel erscheint derzeit auch der dritte theoretische Lösungsansatz: die Emulation.² Es dürfte nur mit gewaltigem Aufwand möglich sein, sämtliche Elemente der Hard- und Software, die eine heutige Systemumgebung ausmachen, zu emulieren. Allein die Unzahl von Prozessorvarianten, Bus-Systeme, Schnittstellen, Maustreiber, Graphik- und anderen Karten, verbunden mit den Spezifikationen der Software in ihren verschiedenen Versionen, die auch in unserem eigenen Computeralltag schon für genug Kopfzerbrechen sorgen, lässt eine komplette Emulation auf das immer neuste System als wenig realistisch erscheinen. Wem von uns bereitet nicht schon der Austausch einer Graphikkarte oder die Adaption eines Druckertreibers Probleme - und dies sind im Vergleich zur Emulation geradezu Fingerübungen. Die Rekonstruktion einer vergangenen Systemumgebung erfordert heute bereits Spezialkenntnisse, die von einer/einem ArchivarIn nicht erwartet werden können - betriebswirtschaftlich wird die Emulation weder bei der Übernahme noch bei der Benutzung vertretbar sein.³ Zudem greift Emulation immer auf proprietäre Softwarebestandteile oder Formate zurück; das Archiv müsste also bei jedem Programmbestandteil abklären, ob eine Benutzungslizenz erworben wurde und zu welcher Nutzung sie berechtigt.

Ein Motiv für Entwicklung eines Emulationskonzepts scheint die stille Hoffnung zu sein, auf diesem Wege mit dem sich in und durch die Daten ausdrückenden Verwaltungshandeln auch dessen Evidenz unverfälschbar zu erhalten. Dies scheint mir jedoch eher trügerisch Evidenz,

¹ Der Text ist die leicht überarbeitete Fassung des Vortrages, der auf der Tagung aus zeitlichen Gründen nur in geraffter Form vorgetragen wurde. Zu ELAN vgl. den Beitrag von Andreas Engel in diesem Band.

² Siehe dazu Frank M. Bischoff: Emulation - das Archivierungskonzept der Zukunft? In: Michael Wettengel (Hrsg.): Digitale Herausforderungen für Archive (Materialien aus dem Bundesarchiv, Heft 7), Koblenz 1999, S. 15-23.

³ Als Beispiel verweise ich auf den Aufwand, den es bedarf, um Tetris, ein Computerspiel der ersten Generation in einer C 64 - Emulation zum Laufen zu bringen.

die sich bei papiergestützten Vorgängen beispielsweise in Paraphen und Vermerken ausdrückt, verschwindet, wenn die Zuordnung des Feldes "Paraphe", des Feldes "Eingangsdatum" oder des Feldes "Vermerk" zu einem bestimmten Feld "Schreiben XY" nur noch über eine flüchtige und manipulierbare Verknüpfung innerhalb eines Datensatzes herstellbar ist. Der zwingende Beweis, den etwa eine blaue Paraphe auf einem eingegangenen Schreiben darstellt, wandelt sich bei der Verbindung zweier Datenfelder zu einer bloßen Vermutung.

Der technologisch und organisatorisch gewaltige Aufwand der elektronischen Signatur sollte auf Archive wie ein Menetekel wirken. Wenn der Beweis der Authentizität eines Dokuments durch bloßen Zeitablauf oder den einkalkulierbaren Erfolg eines Hackers zunichte gemacht werden kann,⁴ dürfen sich Archive nicht auf die analoge Lösungsstrategie verlassen, Evidenz sei durch Emulation zu sichern.

Die Evidenz erfährt bei der Ablösung der papiergestützten Aktenführung durch eine elektronisch konfigurierte Vorgangsbearbeitung eine entscheidende Verwandlung: sie sitzt nicht mehr "zwischen den Zeilen", in Symbolen und "Anordnungen von Elementen auf dem Papier",⁵ sondern in der Verbindung von genau definierten Feldern eines Datensatzes mit dem Ablaufschema der real gelaufenen Anwendung. Die Definition der Feldinhalte ist nicht Bestandteil des Datensatzes selbst - dort sind nur formale Angaben gespeichert⁶ - sondern der Codelisten, der jeweiligen Anwendung, der Dienstanweisungen und des Programmumfeldes. Deshalb kann die Evidenz nur mittels der Dokumentation dieser Angaben erhalten werden.

Es bleibt als realistische Strategie zur Sicherung von Evidenz- und Informationswert die Konversion der zu archivierenden Daten in ein softwareunabhängiges, möglichst stabiles Format, das durch die Massenhaftigkeit seiner Anwendung die Sicherheit der Aufwärtskompatibilität bietet. Die standardisierte Dokumentation des Programm- und Verwaltungsumfeldes wird zum unverzichtbaren Bestandteil der Überlieferung selbst; die Sicherung des Evidenzwertes geschieht durch die Normung des Überlieferungsverfahrens.

Die Diskussion innerhalb der Archivlandschaft der Bundesrepublik ist ambivalent: zwar haben sich verschiedene regionale und überregionale Archivtage mit dem Problem beschäftigt, andererseits geht das Engagement der meisten Archive kaum über das Zur-Kennntnis-Nehmen hinaus. Die Schere zwischen der Dringlichkeit der Aufgabe und der Zurückhaltung bei der praktischen Beschäftigung mit elektronischer Vorgangsbearbeitung öffnet sich immer weiter. Auf der Ebene der archivfachlichen Diskussion sind kaum neue Impulse zu erwarten, wenn nicht konkrete Erfahrungen gemacht und reflektiert werden.

⁴ Siehe dazu: Frank Bischoff: Zur Archivfähigkeit digitaler Signaturen in elektronischen Registern, in: in: Udo Schäfer; Nicole Bickhoff (Hrsgg.): Archivierung elektronischer Unterlagen (Werkhefte der staatlichen Archivverwaltung Baden-Württemberg A 13), Stuttgart 1999, S. 183-198 und Udo Schäfer: Authentizität. Vom Siegel zur digitalen Signatur, in: ebda, S. 165 - 182.

⁵ Angelika Menne-Haritz: Das Provenienzprinzip - ein Bewertungssurrogat?, in: Der Archivar, Heft 2, 1994, Sp. 229-252, hier Sp. 248f.

⁶ Zum Beispiel als zweistelliges Ziffernfeld; ob die Ziffernfolge 03 nun den Familienstand oder die Zahl der Vorstrafen angibt, geht aus dem Datensatz selbst nicht hervor. Zu ihrer Interpretation bedarf es der Überlieferung der Metadaten, d.h. von Daten über die logische Struktur, die inhaltliche Struktur, die Arbeitsumgebung und den institutionellen Zusammenhang.

Die Glaubwürdigkeit und Durchsetzungsfähigkeit archivischer Forderungen an Verwaltung, Gesetzgeber und Industrie kann nur gewinnen, wenn wir auch auf der Ebene der konkreten Umsetzung präsent sind.⁷

Die Beteiligung der Stadt Mannheim an einem Modellvorhaben zur elektronischen Archivierung und Nutzung ist die logische Konsequenz aus der Vorreiterrolle der Stadt bei der Anwendung der Informationstechnologie im Rahmen der Verwaltungsmodernisierung. Die Beteiligung eines Unternehmens wurde vom Stadtarchiv von Anfang an gewünscht. Zum einen sollte ein bereits mit größeren Datenmengen laufendes System in die Untersuchung einbezogen werden, zum anderen eröffnet sich dadurch ein weiteres und interessantes Feld für Dienstleistungsangebote der Archive.

Die Beteiligung der GBG-Mannheimer Wohnungsbaugesellschaft erwuchs aus einer klassisch-traditionellen Aufgabe des Stadtarchivs. Im Jahr 2001 wird die GBG ihr 75. Jubiläum feiern; das Unternehmen bat im Vorfeld das Stadtarchiv um Mithilfe bei der Erstellung einer Festschrift. Da eine solche ohne den Aufbau eines Unternehmensarchivs nicht zu realisieren wäre, wurde der Referent damit betraut.

Aus der Konfrontation mit den - für die GBG selbst unbefriedigenden - Ergebnissen der bisherigen Schriftgutverwaltung⁸ und den Erfordernissen eines unternehmensgerechten Informationsmanagements war die GBG offen für das Angebot des Archivs, bei der Schriftgutverwaltung beratend tätig zu werden. Das archivische Prinzip des Interventionismus traf hier bei Geschäftsleitung und Allgemeiner Verwaltung auf offene Ohren.⁹ Die Unterstützung der GBG bei der Umstellung der Schriftgutverwaltung betrifft sowohl die klassische Schriftgutverwaltung (Aktenplan, Umgang mit Schriftgut, Informationsmanagement) als auch die Strukturen der Langzeitspeicherung der Daten.

Gerade letzteres stellt sich im Moment als Katalog von Fragen dar, die in der konkreten Umsetzung im Rahmen des Projekts einer Klärung nähergebracht werden sollen.

a) Formate und Schnittstellen

Aus der Vielzahl von möglichen Formaten, in denen Texte, Graphiken und Datenbanken gespeichert werden können, hat sich durch theoretische Überlegungen bereits eine strenge Auswahl ergeben.¹⁰ Die Langzeitarchivierung von bewegten Bildern, Tondokumenten und HTML-Dokumenten wird dabei ausgeklammert: Zum einen hat sich hier die Entwicklung

⁷ Einen Appell an die Industrie zu richten, "leicht anwendbare und kostengünstige Lösungen für die kurz- und langfristige Aufbewahrung und Zugänglichkeit digitaler Unterlagen anzubieten", wie dies das DLM - Forum 1999 in Brüssel tat – vgl. Der Archivar, Heft 1, Jg. 2000, S. 61-63 - mutet eher hilflos und resignativ an.

⁸ Informationsverluste durch wilde Kassation, durch mangelhafte Aktenführung und Ablage sowie Verlust der Übersicht über den Stand der Sachbearbeitung, der Zuständigkeit, der Aktivitäten anderer Stellen des Hauses.

⁹ Dazu auch Udo Schäfer: Büroautomation in der Landesverwaltung Baden-Württemberg. Archivische Strategien, in: Frank Bischoff (Hg): Archivierung von Unterlagen aus digitalen Systemen, Münster 1997, S. 31-47.

¹⁰ Zusammenfassend siehe Raphael Ostermann: Potentielle Dateiformate zur Langzeitarchivierung von Dokumenten unter Berücksichtigung von Primär- und Metainformation, in: Michael Wettengel (Hg): Digitale Herausforderungen für Archive, Koblenz 1999, S. 25 - 35 und: Insar, Beilage III, 1997: Leitlinien für den Umgang mit elektronischen Informationen, Maschinenlesbare Daten und elektronische Dokumente.

noch nicht so weit stabilisiert, dass mit einem zukunftsfähigen Format gearbeitet werden könnte, zum anderen fallen solche Dateien bei der GBG in archivwürdiger Form nicht an.

Vorab stellt sich die Frage, ob eine Trennung von reinen Textdateien und Text/Graphik-Dateien sinnvoll ist. Dem Vorteil einer relativ stabilen und zukunftssicheren Speicherung von strukturierten CI-Daten steht der Nachteil einer Aufspaltung der Dateiformate gegenüber, wenn die Dokumente der abzuspeichernden Anwendung auch Elemente enthalten, die sich nur im Image-Format wiedergeben lassen.

Eine generelle Abspeicherung im Image-Format, wie sie das Bundesarchiv auch für Dateien aus Textverarbeitungssystemen bevorzugt, die routinemäßig in TIFF Dateien umgewandelt werden, erkaufte den Vorteil der Einheitlichkeit mit dem Verzicht auf direkte Recherchemöglichkeit in den Dokumenten.¹¹ Da Archive aber generell Wert auf Recherche nach Aktentitel bzw. Klassifikation legen sollten, scheint letzteres verschmerzbar. Voraussetzung dabei ist allerdings eine klare, systematisch geordnete Ablagestruktur der Dateien in der Verwaltung bzw. in dem Unternehmen selbst. Zu prüfen wäre auch ein Dateiformat für die Speicherung von Vektorgraphiken der Planungsabteilung, da sich bislang kein Standard herausgebildet hat.¹²

Für die Übernahme von Datenbanken gilt als Arbeitsprinzip, dass das auszuwählende Programm ein SQL 2-fähiges Standardprogramm sein muss.¹³ Dies gilt auch für das im Stadtarchiv bislang gewählte ACCESS, dessen Schnittstelle ODBC SQL-fähig ist.

Die Speicherung von Dateien als flat file¹⁴ erkaufte den Vorteil der Systemunabhängigkeit mit dem Nachteil, nicht direkt und effizient aufruf- und recherchierbar zu sein. Unstrukturierte sequentielle Dateien müssen vor einer Nutzung erst wieder in eine Datenbankanwendung eingelesen werden. Hierbei wäre zu prüfen sein, wie das Verhältnis von Zugriffsdauer zu Speicherkapazität und Kosten sich gestaltet.

Zur Lösung des Problems, Daten und Dateistrukturen sowie Metadaten gleichzeitig und eindeutig definiert abspeichern zu können, empfiehlt sich die im Februar 1999 verabschiedete Metasprache XML (eXtensible Markup Language). XML speichert in der definierten Struktur der einzelnen Daten; innerhalb eines Textes beispielsweise kann präzise definiert werden, welche Zeichenfolge die Überschrift ist, welche der Text und welche z. B. eine Fußnote. Analoges gilt für Datenbanken, für eingearbeitete Graphiken und Verweise. Damit sind gezielte Recherchen genauso möglich wie eine eindeutige Ausgabe auf Schirm oder Drucker.¹⁵ Das GBG - Projekt präferiert demzufolge, XML als Langzeitspeicherformat zu testen.

¹¹ Michael Wettengel: Aussonderung und Archivierung elektronischer Akten, in: Ders. (Hg.), Digitale Herausforderungen (wie Anm. 10), S. 89-96.

¹² Immerhin scheint CGM = Computer Graphics Metafile solider als das bislang noch häufiger anzutreffende CAD - Format zu sein.

¹³ SQL = Structured Query Language ist eine Abfragesprache, die auf verschiedene Datenbanken, Datenbankstrukturen und Tabellenformate gleichermaßen zugreifen kann.

¹⁴ Flat files als Speicherungsart ist die Speicherung von reinen, softwareunabhängigen Datenstrings, bei denen nur die Stammdaten und die Struktur abgespeichert werden. Bei Textdokumenten sind dies eindimensionale Dateien, bei Datenbanken unstrukturierte sequentielle Dateien.

¹⁵ Eine Kurzinformation zur Struktur von XML: <http://www.w3.org/xml>.

b) Übernahme und Medium

Die Übernahme elektronisch erzeugter Unterlagen setzt eine positive Bewertungsentscheidung voraus. Dies ist durchaus keine Selbstverständlichkeit, da vor allem die erste Generation von IT-Anwendungen häufig in Bereichen mit massenhaft gleichförmigem Schriftgut eingesetzt wurde, mithin in einem Bereich, dessen Unterlagen häufig genug keine Archivwürdigkeit zugemessen werden konnte.

Eine verantwortliche Bewertungsentscheidung muss auch definieren, welche Datensätze und/oder welche Felder dieser Datensätze archivwürdig sind. Dies ist in der EDV-Welt weniger ein Platz- und auch kein Speicherproblem, sondern eine archivisch zu erbringende Leistung für Erschließung und Auswertung. Übernommen wird also nicht die Originaldatenbank, sondern eine neu und ausdrücklich zum Zweck der dauernden Aufbewahrung gerichtete.¹⁶ Dies ist eine Abwandlung der Archivierungsstrategie bei papiergestützten Unterlagen, die im Normalfall ja ohne archivische Veränderung übernommen werden. Damit ergibt sich zwingend, dass diese Veränderung von Datenstruktur, Umfang und Inhalt dokumentiert werden muss, um bei künftiger Auswertung die real gelaufene Funktionalität berücksichtigen zu können. Auch daraus ergibt sich die Wanderung der Evidenz aus den Unterlagen in die Dokumentation.

Technisch gesehen sollte die Speicherung sicherheitshalber in zwei getrennten Speichermedien vorgenommen werden.¹⁷

c) Dokumentation

Die Erfahrungen, die das Bundesarchiv bei der Übernahme von Unterlagen der DDR machen musste, lassen die mangelhafte Dokumentation als das größte Problem erscheinen.¹⁸ Deshalb wird sich das Projekt mit besonderer Sorgfalt der Frage widmen, welche Unterlagen in welcher Form als Dokumentation aufzubewahren sind und wie diese Dokumentation zu erschließen ist. Als Ziel steht dabei ein Standard-Dokumentations-Katalog in Analogie zum dem Dokumentationskatalog des Bundesarchivs bei der Übernahme von Originaldateien.¹⁹

Dokumentiert werden muss:

- das Verhältnis der Archivversion zur ursprünglichen Struktur und Inhalt,
- die logische Struktur der Aufzeichnung,
- der Kontext der technischen Metadaten der Primäranwendung (Hard-/Software, Dateistruktur, Datenbeschreibung, Verknüpfungen),

¹⁶ Archivierung ist die Neuordnung, Auswahl oder Aggregation von Daten zu einem kompakteren Datenbündel, vgl. INSAR Leitlinien, S. 21.

¹⁷ Vgl. Schäfer: Authentizität (wie Anm. 4), S. 179 und den Verweis auf das Projekt "The Preservation of the Integrity of Electronic Records".

¹⁸ Michael Wettengel: Archivierung digitaler Datenbestände aus der DDR nach der Wiedervereinigung, in: Schäfer; Bickhoff (Hrsg.), Archivierung (wie Anm. 4), S. 223 - 239.

¹⁹ Vgl. den Katalog von Wettengel, ebd., S. 230 f.

- der verwaltungstechnische Kontext (Herkunft, Historie und Verwendungszusammenhang, abgebende Stelle und Organisationseinheit, Rechtsgrundlage und Ziel der DV-Anwendung),
- die Dokumentation der real praktizierten Anwendung, da neue Datenbankanwendungen z.B. Auswertungen erlauben, die im Original struktur- oder rechengeschwindigkeitsbedingt nicht möglich waren. Auch wenn "nur" die höhere Geschwindigkeit der Abfrage einen Sprung von der Quantität zur Qualität bedeutet, muss dies auf jeden Fall dokumentiert werden.

d) Erschließung und Verzeichnung

Die Erschließung und Verzeichnung maschinenlesbarer Unterlagen weicht zwangsläufig von der traditionellen Form der Erschließung von Unterlagen ab, dürfte aber im Rahmen des Provenienzprinzips keine besondere Schwierigkeit darstellen.

e) Rechtsform

Die Daten werden genauso als Depositum übernommen wie die bisherigen Übernahmen traditioneller Ablieferungen. Für Datensicherung und Pflege werden analog kostendeckende Gebühren erhoben. Es wird zu prüfen sein, ob die Daten auf stadteigenen Anlagen gespeichert und benutzt werden können; im Falle einer Speicherung auf Rechnern anderer Rechtsträger wäre eine geeignete Form der Vertragsgestaltung zu suchen.

f) Benutzung

Gedacht ist die Gestattung von Benutzung nach den archivrechtlich üblichen Kriterien. Eine Benutzung geschieht offline auf einem dritten Speichermedium, das auf Antrag hergestellt wird, möglicherweise auf einer CD - ROM.²⁰ Eine sog. "Abfragekopie" ist eine datenschutzrechtlich unbedenkliche Version, bei der die Anonymisierung durch feldbezogene Auswahl vorgenommen wurde, in einem benutzerfreundlichen bzw. vom Benutzer gewünschten Format. Damit wäre sogar ein rechtlich unbedenklicher Zugriff via Internet möglich. Ein genereller online-Zugriff dürfte aus datenschutzrechtlichen Gründen wie auch aus Kostengründen nicht sinnvoll sein.

Nach Abschluß der Benutzung sollte die Abfragekopie vernichtet werden, um Weitertradiere von bewußten oder unbewußten Manipulationen zu vermeiden. Alternativ dazu könnten die aufbereiteten Daten werden für künftige Benutzer aufbewahrt werden, wenn eine entsprechende Vereinbarung mit dem Erstnutzer getroffen werden kann.. Dabei wäre aber unmissverständlich deutlich zu machen, in welchem Verhältnis die Abfragekopie zur Originaldatenbank steht.

²⁰ Vgl die Praxis des Bundesarchiv: Anweisung für die archivarische Tätigkeit Nr. 52, siehe Wettengel: Archivierung digitaler Datenbestände (wie Anm. 18), S. 234

Analog zu der aus der aus Gründen der Bestandserhaltung vorgenommen doppelten Überlieferung für die beiden Aufgabenbereich dauerhafte Aufbewahrung (Sicherungskonzept) und einfache Zugänglichkeit (Benutzungskonzept) könnten auch bei digitalen Unterlagen Benutzerdateien in leicht zugänglicher Datenstruktur vorgehalten werden. Die entsprechenden Überlegung des Bundesarchivs können hier Vorbild sein.