

Die Spitze des virtuellen Eisbergs.

Ein Werkstattbericht zu ersten Begegnungen des Hauptstaatsarchivs Düsseldorf mit digitalen Unterlagen*

Axel Koppetsch / Matthias Meusch

Das Hauptstaatsarchiv Düsseldorf (HSA) hat zwar bisher keine digitalen Unterlagen übernommen, doch lassen diverse Vorboten darauf schließen, dass der Eintritt in die Ära virtueller Archivalien unmittelbar bevorsteht. So haben etwa gezielte Nachfragen bei Behörden und Einrichtungen im Zuständigkeitsbereich des Ministerialarchivs zu geplanten oder bereits im Einsatz befindlichen IT-Verfahren in einigen Fällen mehr oder weniger akuten Handlungsbedarf erzeugt und den Verdacht bestätigt, dass die im nordrhein-westfälischen Archivgesetz (§3 Abs. 4) verankerte Vorschrift, nach der sich Behörden vor Einführung von Datenverarbeitungssystemen mit dem zuständigen Archiv über archivfähige Datenformate verständigen müssen, kaum bekannt ist und entsprechend wenig Beachtung findet.

Vor diesem Hintergrund erscheint der Rückgriff auf das Bild des virtuellen Eisbergs in fünffacher Hinsicht angemessen, denn

- Eis ist, gerade in Zeiten des Klimawandels, flüchtig;
- Eisberge bewegen sich langsam, aber unaufhaltsam;
- der weitaus überwiegende Teil bleibt den Blicken verborgen, stellt aber die größere Herausforderung dar;
- ein Eisberg kommt selten allein;
- Eiswarnungen sollten auch von solchen Schiffen nicht ignoriert werden, die als unsinkbar gelten.

Der Beitrag versteht sich nicht nur insofern als Werkstattbericht, als noch keine greifbaren oder gar übertragbaren praktischen Erfahrungen vorgelegt werden können. Auch die theoretischen Überlegungen sind noch im Fluss und werfen z. T. noch mehr Fragen als Antworten auf, doch entspricht die Reihenfolge der vorgestellten Beispiele immerhin einem zunehmenden Konkretisierungsgrad hinsichtlich bisher erwogener Lösungsansätze.

Die ausgewählten Themen verweisen zugleich darauf, dass digitale Unterlagen auch eher unvermutet auf indirekten Wegen begegnen können: Zum einen stehen die Überlegungen sowohl zu den beim Rechenzentrum der Finanzverwaltung zentralisierten digitalen Steuerakten wie zu den Websites der Polizeibehörden in Zusammenhang mit landesweiten Bewertungsprojekten, die sich zunächst auf analoge Unterlagen beschränkten, bis im Laufe der Arbeit deutlich wurde, dass die Einbeziehung auch digitalen Materials unumgänglich ist. Zum anderen resultiert die Beschäftigung mit den digitalisierten Plänen der Deutschen Bahn AG aus den Anforderungen, die die Übernahme der analogen Exemplare im Hinblick auf Bestandserhaltung, Raumbedarf und Lagerungstechnik stellen würde.

1) Zentralisierung digitaler Steuerakten beim Rechenzentrum der Finanzverwaltung (RZF)

Das RZF geht zurück auf die 1954 gegründete Lochkartenzentrale und erfüllt heute mit ca. 600 Mitarbeitern (davon ca. 280 Programmierern) u. a. folgende Aufgaben: Berechnung von Steuern, Steuererstattungen und Besteuerungsgrundlagen, Buchführung über Einnahmen und Ausgaben, Übermittlung von Daten an Behörden außerhalb der Finanzverwaltung sowie Sicherstellung und Pflege der IT-Struktur in den 112 Festsetzungs- und Erhebungsfinanzämtern in Nordrhein-Westfalen. Als Landesoberbehörde fällt das RZF in die Zuständigkeit der Abteilung 3 des HSA (Ministerialarchiv), während die nachgeordneten Finanzämter von der Abteilung 2 bzw. den Staatsarchiven in Münster und Detmold betreut werden. Deren Übernahme von Steuerakten folgt

* Der Beitrag beschreibt den Stand zum Zeitpunkt der Tagung Mitte März 2003. Spätere Entwicklungen konnten an dieser Stelle keine Berücksichtigung finden.

bislang einem Archivierungsmodell, das – in starker Vereinfachung – die Vorgänge zu bedeutenden Persönlichkeiten des öffentlichen Lebens, zu Betrieben ab einer bestimmten Größe und zu einigen der für eine Wirtschaftsregion typischen Firmen berücksichtigt.

Im Rahmen der eingangs erwähnten Umfrage zu Planung und Einsatz von IT-Verfahren bei den Landesoberbehörden hat sich nun kürzlich herausgestellt, dass das RZF für den Bereich der Einkommensteuerpflichtigen bereits gegenwärtig und für den Bereich der Gewerbebetriebe in naher Zukunft als digitale Registratur der Finanzämter fungiert bzw. fungieren wird. Das dabei eingesetzte Verfahren betrifft einerseits die sog. Grunddaten wie Name, Geburtsdatum, Adresse etc., die in den Auskunftssystemen des RZF und der Finanzämter identisch vorgehalten und stets ohne Speicherung der Historie aktualisiert werden. Sie dienen zum einen der Aufgabenerledigung bei der Besteuerung, zum anderen als Grundlage für Datensammlungen, die das RZF im Auftrag des Finanzministeriums generiert. Davon unterschieden werden andererseits die Steuerbescheide, die ausschließlich im RZF auf WORM gespeichert und bei Bedarf von den Finanzämtern abgerufen werden können, die also nicht mehr über eigene Unterlagen verfügen.

Dieser Befund ist für die Archive in (mindestens) fünffacher Hinsicht problematisch: Zum einen ist damit zu rechnen, dass vom RZF für das Finanzministerium zusammengestellte Grunddaten dort in Entscheidungsprozesse eingehen und im Falle der Archivwürdigkeit zu Hybridüberlieferungen führen können. Zum zweiten stellt sich die Frage, ob es sich bei den im RZF gepflegten Grunddaten um Datenbestände handelt, die „aus verarbeitungstechnischen Gründen vorübergehend vorgehalten werden“ und als solche nach §3 Abs. 4 des Archivgesetzes nicht der Anbietungspflicht unterliegen. In Bezug auf die Steuerakten insgesamt ist drittens zunächst unter den Archiven zu klären, ob die weitere archivarische Betreuung einschließlich Bewertung und ggf. Übernahme durch das Ministerialarchiv in seiner Zuständigkeit für das RZF als Landesoberbehörde, durch die für die Finanzämter zuständigen Dezernate der drei Staatsarchive oder im Zusammenwirken aller Genannten erfolgen soll. Die Beantwortung dieser Frage ist viertens kaum zu trennen von der Entscheidung, wo ggf. als archivwürdig bewertete Unterlagen archiviert werden sollen, die entweder nach ihrer Abgabeprovenienz dem Ministerialarchiv oder nach ihrer Entstehungsprovenienz der Abteilung 2 des HSA bzw. den Archiven in Münster und Detmold zuzuschlagen wären. Fünftens schließlich ist zu bedenken, ob das bisherige Archivierungsmodell unter den veränderten technisch-organisatorischen Rahmenbedingungen noch unverändert Anwendung finden kann oder in fachlich sinnvoller Weise zu modifizieren ist.

Obwohl also praktisch alle Fragen in Bezug auf den künftigen Umgang mit den Steuerakten der nordrhein-westfälischen Finanzverwaltung offen sind, erschien die Vorstellung dieses Themas im Rahmen einer Tagung angemessen, die sich mit den Auswirkungen elektronischen Archivguts auf das archivarische Berufsbild widmet. Denn das Beispiel verweist darauf, dass die Archivierung digitaler Unterlagen nicht nur die mittlerweile bekannten technischen Schwierigkeiten aufwirft, sondern auch zur erneuten Beschäftigung mit altherwürdigen archivarischen Grundfesten wie dem Provenienzprinzip Anlass geben kann.

2) Die Archivierung von Websites der Polizeibehörden

Seit dem Juli letzten Jahres existiert in Nordrhein-Westfalen ein Projekt zur Bewertung der Unterlagen der nordrhein-westfälischen Polizeibehörden, dessen Ziel die Erstellung eines landesweit anzuwendenden Überlieferungs- und Bewertungsmodells ist. Im Zuge dieses Projektes kam seitens der beteiligten Kolleginnen und Kollegen der Wunsch auf, auch die Webauftritte der Polizeibehörden einer Bewertung und ggf. Archivierung zu unterziehen.¹

Gemäß den vorläufigen, von der Projektgruppe entwickelten Kriterien erscheinen vor allem jene Inhalte archivwürdig, die die spezifische Selbstdarstellung der jeweiligen Polizeibehörde betreffen. Dazu gehören Begrüßungseiten, Berichte zu einzelnen Ereignissen, vor allem überregionaler

¹ Auch hinsichtlich des Intranets der Polizeibehörden besteht unsererseits der Wunsch, ausgewählte Inhalte zu übernehmen – allerdings fehlt es hier zur Zeit noch an der Bereitschaft der Polizei, dem Archiv hier einen entsprechenden Zugang zu gewähren.

Bekanntheit (exemplarisch sei hier der „Fall Neisius“ aus Köln genannt)² sowie spezielle, von einer Polizeibehörde eigenständig entwickelte Konzepte zu Prävention und Vorbeugung. Bislang unterhält jede Polizeibehörde ihre eigene Webseite, ein einheitlicher, standardisierter Internet-Auftritt ist geplant, ein Zeitplan für die Realisierung aber noch nicht in Sicht.

Die Archivierung von Internetauftritten unterliegt in unserem Fall bestimmten Prämissen:

- Der Anstoß zur Archivierung geschieht im Allgemeinen durch das Archiv: Die Webauftritte aller Polizeibehörden des Sprengels werden in regelmäßigen Abständen nach archivwürdigen Inhalten durchgesehen.
- Die ausgewählten Seiten müssen je nach ihrer Struktur gegebenenfalls durch die Behörde vor der Übergabe einer gewissen Bearbeitung unterzogen werden. So müssen etwa (interne) Verlinkungen geändert werden.
- Dies ist bedingt durch die grundsätzliche Entscheidung, nicht komplette Internetauftritte zu archivieren, sondern nur deren als archivwürdig angesehene Inhalte. Es besteht also ein grundsätzliches Primat des Inhalts gegenüber der Form, das heißt in diesem Fall gegenüber Design und Funktionalitäten. Anders ausgedrückt: Webdesign und Funktionalitäten gehen bei diesem Projekt partiell verloren.

Um die Archivierung von Internet-Seiten der Polizeibehörden einmal an einem Beispiel praktisch zu erproben, haben wir Kontakt zum Kölner Polizeipräsidium aufgenommen und die dortige Internet-Redaktion aufgesucht. Man konnte sich auf die folgende, zunächst recht simple Vorgehensweise verständigen:

1. Das zuständige Dezernat im Hauptstaatsarchiv Düsseldorf teilt der Polizeibehörde mit, welche Inhalte der aktuellen Webseite archiviert werden sollen.
2. Für die Übertragung der durch das Archiv als archivwürdig angegebenen Inhalte gibt es zwei Wege:
 - a. Die Behörde legt die Seiten als Zip-Archiv in einem gesonderten Ordner ab, auf den das Hauptstaatsarchiv einen passwort-geschützten Zugang über das Internet erhält.
 - b. Handelt es sich um größere Datenmengen, soll der Zugriff per FTP über das Landesverwaltungsnetz geschehen.

Die Übernahme geschieht zunächst in Form von html-Dateien.

Bei diesem Vorgehen wird, wie gesagt, in Kauf genommen, dass nicht mehr alle ursprünglich vorhandenen Funktionalitäten, also Verknüpfungen, auch weiterhin funktionieren. Für externe Verknüpfungen kann dies generell nicht in Frage kommen, die Beibehaltung interner Verknüpfungen richtet sich nach der jeweiligen Archivwürdigkeit. So werden z. B. nicht nur die „Frequently Asked Questions“ (FAQ) übernommen, sondern sinnvollerweise auch die dazu gehörigen Antworten. Um die Verknüpfung beider html-Dateien zu gewährleisten, müssen lediglich alle miteinander in Verbindung stehenden Dateien in einem Verzeichnis abgelegt werden. Bei komplexeren Fällen, wie den Informationen über den Fall „Neisius“, die mit über den Webauftritt verstreuten Presseartikeln verlinkt sind, ist etwas Vorarbeit der Behörde vonnöten, die die ursprünglich verzweigten Verlinkungen „glätten“ muss. Seitens der Polizeibehörde Köln sieht man hier kein Problem, diese Arbeit zu leisten. Inwieweit weitere Polizeibehörden sich dazu bereit finden, wird noch zu ergründen sein. Auf diese Weise kann aber zumindest für die als archivwürdig bewerteten Inhalte die ursprüngliche Funktionalität wieder hergestellt werden.

Trotzdem ist dieses Vorgehen nicht ausreichend und nicht völlig befriedigend. Es sollten deshalb, wie zum Beispiel bei der Archivierung von Datenbanken auch, Funktionalitäten und Erscheinungsformen, also in gewissem Ausmaß auch das Design der Webseite, zumindest dokumentiert werden. Hierzu gibt es verschiedene Möglichkeiten, die im Einzelnen noch zu diskutieren sind: So könnte man etwa die

² Es handelt sich hier um den Fall eines 31-jährigen Mannes, der nach seiner Festnahme auf einer Kölner Polizeiwache durch Beamte misshandelt wurde. Der Mann fiel ins Koma und starb wenige Tage darauf. Die Staatsanwaltschaft hat mittlerweile Anklage gegen einige der Beamten erhoben. Der Fall wurde ausführlich auf der Homepage des Polizeipräsidiums Köln dokumentiert.

eigentliche homepage, also im Sinne der Startseite, mit allen ihren weiter leitenden Links mit archivieren. Gegebenenfalls ist auch an die Herstellung und Archivierung von Screenshots zu denken. Wichtiger Bestandteil einer Dokumentation ist selbstverständlich darüber hinaus das Impressum, quasi als Provenienzangabe.

Hinsichtlich der eigentlichen Langzeitarchivierung lautet natürlich auch hier das Zauberwort XML. Hierzu muss eine geeignete DTD erarbeitet oder vielleicht eine schon bestehende für die eigenen Belange entsprechend abgeändert werden, so dass die Struktur der Website abgebildet werden kann. Da, wie oben bereits geschildert, die Absicht besteht, für alle Polizeibehörden auf der Grundlage eines einheitlichen Bewertungsmodells immer die gleichen Inhalte zu archivieren, dürfte die Erarbeitung einer „Standard“-DTD für die zu überliefernden Website-Inhalte keine allzu große Schwierigkeit darstellen. Auch die Dokumentation kann in Form von Metadaten – inklusive etwaiger vorliegender Screenshots – in die XML-Struktur eingebaut werden. Dieser Teil allerdings ist noch in der Planungsphase, hier muss der Werkstattbericht für diesen Teilbereich enden.

3) Digitalisierte Pläne der Deutschen Bahn AG

Durch das Hauptstaatsarchiv Düsseldorf wurden bislang ca. 63.000 Karten und Pläne der Deutschen Bahn AG, vornehmlich aus den Bereichen Vermessungswesen, konstruktiver Ingenieurbau und – allerdings zu einem geringeren Teil – Hochbau übernommen. Laut Ankündigung der DB Netz vom Oktober letzten Jahres sind noch einmal ca. 50.000 Pläne für die Anmietung vorgesehen. Hintergrund dieser umfangreichen Anmietungen ist das Digitalisierungsprojekt „IZ Plan“ („Informationszentren-Plan“). Im Rahmen dieses Projekts werden seit 1997 alle Karten und Pläne eingescannt, mit Metadaten versehen und über ein Netzwerk den einzelnen Einrichtungen der Bahn AG zur Verfügung gestellt. Seitens der Bahn werden die Pläne, sobald sie digitalisiert sind, nicht mehr benötigt und sollen den zuständigen staatlichen Archiven angeboten oder – falls diese kein Interesse anmelden – gegebenenfalls vernichtet werden.

Nun stellen diese Pläne für unser Haus sowohl lagerungstechnisch, schlicht aufgrund ihrer Masse, als auch bestandserhalterisch ein massives Problem dar. Viele sind bereits heute so beschädigt, dass sie Benutzern nicht mehr vorgelegt werden können. Da lag der Gedanke nahe, gleich die digitalisierten Pläne mitsamt den dazugehörigen Metadaten zu übernehmen. Wie bereits erwähnt, wird zu jedem eingescannten Plan in einer Oracle-basierten Datenbank namens PRO*FILE ein eigener Datensatz mit allen zugehörigen Metadaten angelegt. Verknüpft werden Bilddateien und Metadaten über einen bei der Digitalisierung auf den Plänen angebrachten Barcode (inkl. Nummer). Es gibt eigene Datensätze zu Plänen sowie zu den auf diesen abgebildeten Strecken. Dabei treten zwischen Plan- und Streckendatensätzen 1:n-, gegebenenfalls auch n:1-Beziehungen auf. Das heißt, dass auf einem Plan mehrere Strecken abgebildet sein können oder es zu einer Strecke mehrere Pläne geben kann.

Die Übernahme der Daten stellt kein technisches Problem dar. Die Pläne werden als unkomprimierte TIFF-Dateien abgespeichert, die eine mittlere Größe von 0,5 bis 1 MB aufweisen. Bei 110.000 - 120.000 Plänen ist also mit einem Gesamtdatenvolumen von 80 - 90 GB zu rechnen. Die Metadaten können problemlos aus PRO*FILE nach MS-Excel exportiert werden, die n:n-Verknüpfungen bleiben dabei erhalten und nachvollziehbar. Diese Metadaten fungieren zur Zeit als Abgabeliste für die Pläne, die bereits in analoger Form nach ihrer Digitalisierung abgegeben wurden.

Momentan stößt die Übernahme der digitalisierten Pläne noch auf rechtliche Bedenken bei Teilen der DB Netz, da es sich bei diesen Plänen größtenteils um noch aktives, in der Benutzung befindliches Bestandsplanwerk handelt und ein archivrechtlicher Anspruch auf die Übernahme der erst nach der Privatisierung der Bahn 1993 hergestellten Digitalisate nicht besteht. Diese Fragen werden zur Zeit mit der Konzernzentrale geklärt. Sollte die Bahn AG ihr Einverständnis erklären, würden Übernahme, Lagerung und Benutzung der Pläne im Archiv in erheblicher Weise erleichtert: Die Metadaten könnten

den Grundstock der Bewertung und anschließenden Verzeichnung bilden, die Pläne selbst würden den Benutzern, sofern sie keinen Sperrfristen mehr unterliegen, als Bilddateien, evt. in das leichter zu handhabende JPG-Format konvertiert, im Lesesaal zur Verfügung gestellt. Hinsichtlich der Langzeitsicherung und -archivierung wäre an eine Ausbelichtung der TIFF-Dateien auf Mikrofilm zu denken, die Metadaten in den Excel-Dateien wären ohne Weiteres im ASCII-Format abzulegen.

An dieser Stelle möchte ich einige kurze Anmerkungen hinsichtlich der generellen Verwaltung und Aufbewahrung solcher Datenbestände anschließen. Ein Datenvolumen von knapp 100 GB darf heutzutage niemanden mehr schrecken. Allerdings handelt es sich hier nur um eines von mehreren Vorhaben, bei dem großen Datenmengen anfallen werden. Allein bei einem weiteren, in unserem Hause in der Vorbereitung befindlichen Projekt, der Digitalisierung eines Bestandes von Reichskammergerichtsakten und -urkunden, ist mit einem Datenvolumen von weiteren 100 - 120 GB zu rechnen. Wir werden hier also sehr schnell in den Terabyte-Bereich vorstoßen, eine Datenmenge, die zur Zeit von einem einzelnen Haus weder von der technischen noch von der personellen Ausstattung her zu handhaben sein wird. Seitens des Hauptstaatsarchivs Düsseldorf wird deshalb in dieser Hinsicht eine enge Kooperation mit dem nordrhein-westfälischen Landesamt für Datenverarbeitung und Statistik angestrebt, das in unmittelbarer räumlicher Nachbarschaft angesiedelt ist. Datenhaltung und -verwaltung inklusive der Entwicklung einer technisch und kostenmäßig tragfähigen Migrationsstrategie wären dann dort zu leisten. Der Zugriff auf die Daten muss aus Gründen des Datenschutzes und des Archivrechts jedoch ausschließlich beim Archiv verbleiben. Erste Gespräche über eine solche Zusammenarbeit finden zur Zeit statt.

Dies war ein kurzer Überblick über einige Problemstellungen, mit denen wir im Hauptstaatsarchiv Düsseldorf im Laufe des letzten Jahres konfrontiert wurden. Zu Anfang des Beitrages stand die Metapher des Eisbergs. Tatsächlich haben wir die Erfahrung gemacht, dass dort, wo man nachfragt und nachhakt, auch oft etwas vorhanden ist, um das man sich kümmern muss. Gerade in der letzten Woche erhielt unser Haus von einer der Behörden unseres Sprengels eine ganze Liste von EDV-Anwendungen, die allein in dieser Behörde in Gebrauch sind. Man gewinnt den Eindruck, dass der größte Teil des Problems noch unter der Wasseroberfläche vor sich hin zu schlummert. Die Flüchtigkeit oder Virtualität der Daten in Verbindung mit der rasanten Fortentwicklung der Hard- und Softwareumgebungen gemahnt uns zur Eile. Probleme haben wir also genug, es ist abzusehen, dass weitere auf uns zukommen werden. Trotzdem soll unser abschließender Appell – und hier wollen wir das Bild vom Eisberg zu Ende führen – lauten: „Keine Panik auf der Titanic.“