Erschließung und Nutzbarmachung digitaler Unterlagen im Bundesarchiv

Von Bettina Martin-Weber

Das Bundesarchiv gehört zu den wenigen Archiven in der Bundesrepublik Deutschland, die über praktische Erfahrung bei der Erschließung und Nutzbarmachung digitaler Unterlagen verfügen. Im vorliegenden Werkstattbericht werden gegenwärtige Praxis, Neuansätze und Perspektiven beschrieben. Die Ausführungen konzentrieren sich auf die Erschließung von Daten aus IT-Fachanwendungen in Dienststellen des Bundes und der DDR. Hierbei handelt es sich um Daten aus File-Systemen, modernen Datenbanksystemen und um Statistikdaten.[1]

1. Erschließung digitaler Archivobjekte

Alle relevanten technischen und inhaltlichen Informationen zu archivierten Daten aus einer IT-Fachanwendung einer Provenienzstelle werden im Bundesarchiv bislang in einem sog. Findbuch zusammengestellt. Das „Findbuch" bezieht sich auf eine digitale Archivalieneinheit, die sich inhaltlich und materiell-physisch gegen andere Dateneinheiten abgrenzen lässt und im folgenden als digitales Archivobjekt bezeichnet wird.[2] Als Produkt der archivischen Erschließung unterscheidet sich dieses „Findbuch" von herkömmlichen Findbüchern, die üblicherweise auf der Ebene der Bestände aufsetzen und die Summe der Archivalieneinheiten einer Provenienzstelle, zumindest aber einen umfänglicheren Teil derselben, strukturiert nachweisen. Es enthält zwar durchaus vertraute Elemente, z.B. einen Überblick über die Geschichte der Provenienzstelle, beschreibt aber dann die Entstehung der Daten, den Datenaufbau, die Dateien, die Datenformate, die Datensatzstrukturen, den Feldaufbau, die Kodierungen usw. und dokumentiert die technische Bearbeitung der Daten im Archiv.

Die gebundene Form des Findbuchs setzte der Einbindung von Dokumentationsunterlagen bislang Grenzen. Aktuelle Überlegungen zielen darauf, inhaltliche Erschließung, technische Beschreibung und Dokumentationsunterlagen zu digitalen Archivobjekten in strukturierten, auf Fortschreibung und Zuwachs hin offenen Dossiers[3] zu organisieren. Die offene Form bietet sich auch an, da eine Publikation der Dossiers im allgemeinen nicht zweckmäßig ist, da sie nicht für die Recherche, sondern erst in einem fortgeschrittenen Stadium der Benutzung tatsächlich herangezogen werden. Gleichwohl steht einer online-Version der Textteile der Dossiers nichts entgegen.

Bislang liegen Findbücher für eine Vielzahl, aber keineswegs für alle digitalen Archivobjekte vor; sie repräsentieren jeweils vorläufig die vollständige Erschließung eines Objektes. Ihre Aufbereitung ist arbeitsintensiv und wird nie abgeschlossen sein, da die technische Bearbeitung (z.B. Migrationen) hierin nachzuhalten ist.

Neben diesen Findbüchern – bzw. künftig Dossiers – werden in verschiedenen Tabellen alle Bearbeitungen, Um- und Sicherungskopierungen sowie Konvertierungen und auch Anonymi-

---

[1] Das Bundesarchiv hat bisher keine elektronischen Akten übernommen.

[2] Der Begriff lehnt sich an den im Referenzmodell OAIS (Open Archival Information System) gewählten Terminus des „Digital Information Object" (vgl. Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System, CCSDS 650.0-W-5.0 White Book, 1999) an und meint die übernommenen primären Daten ohne weitere Darstellungs-, Meta- oder Erschließungsinformationen. Es kann sich hierbei um eine u.U. sehr umfangreiche Datei, aber auch um mehrere Dateien handeln.

[3] Zu dieser Organisationsform und zum Begriff regte ein Besuch im Schweizer Bundesarchiv im Sommer 2000 an.

sierungen dokumentiert. Diese technische Dokumentation dient sowohl dem Nachweis der Bearbeitung als auch der internen Arbeitsplanung.

In der Archivdatenbank BASYS (Bundesarchiv – System), die v.a. als zentrales Erschließungsinstrument für Schriftgut seit den frühen 90er-Jahren im Bundesarchiv aufgebaut wurde, wird digitales Archivgut bislang nur in den sog. „Stammdaten", in Informationen auf Bestandsebene und zu den Provenienzstellen nachgewiesen.[4] Digitale Archivobjekte werden in BASYS bislang nicht erschlossen.

Eine Beständeübersicht über die maschinenlesbaren Daten im Bundesarchiv bietet der sog. Datenbestandskatalog (DBK), der die digitale Überlieferung der Provenienzstellen auf der Ebene der digitalen Archivobjekte nachweist. Sie ist im Intranet des Bundesarchivs für den internen Gebrauch einsehbar und wird Benutzern derzeit als Ausdruck oder Word-Datei zur Verfügung gestellt. Im Internet sind die maschinenlesbaren Bestände des Bundesarchivs im Beständebaum mit kurzen Beschreibungen auf der Ebene der digitalen Archivobjekte vollständig nachgewiesen.

2. Funktionen der Erschließung

Archivische Erschließung hat den Zweck, Benutzern das Auffinden gesuchter Archivalien zu ermöglichen und strukturelle Bezüge zu verdeutlichen. Die Erschließung digitaler Archivobjekte hat weitergehende Funktionen. Während sich konventionelles Archivgut im allgemeinen dem (kundigen) Benutzer sehr weitgehend selbst erschließt, da es eine Fülle relevanter Kontextinformationen in sich trägt, sind digitale Archivobjekte nur benutzbar, interpretierbar und analysierbar, wenn dem Benutzer vorab differenzierte Kontextinformationen und technische Daten an die Hand gegeben werden, die sich aus dem digitalen Archivobjekt selbst nicht erschließen lassen. Erst die Erschließung der Objekte stellt ihre Benutzbarkeit her.

Darüber hinaus hat Erschließung die Funktion, die Authentizität der digitalen Überlieferungsform nachzuweisen. Die Authentizität digitaler Objekte lässt sich nicht durch Ab- oder Vergleich mit anderen Originalen, Kopien, paralleler Überlieferung o.Ä. verifizieren oder falsifizieren. Allenfalls mangelnde Plausibilität der Daten lässt an ihrer „Echtheit" zweifeln. Es müssen daher andere Wege des Nachweises der Authentizität der Unterlagen gefunden werden.

Dass unter den Bedingungen der Migration keine unbedingte Authentizität – im Sinne der Wahrung von Identität – möglich ist, liegt auf der Hand. Erst in der archivischen Praxis aber wird deutlich, in welchem Umfang Modifikationen an übernommenen Daten notwendig sein können, sollen sie in für die Langzeitarchivierung und Nutzung nach heutigem Stand geeignete Formate und Strukturen konvertiert werden. Betroffen sind nicht nur die Datenformate. Auch Eingriffe bei Dateistrukturen, Datensatzstrukturen und dem Aufbau von Datenfeldern können erforderlich werden. Dies gilt insbesondere für Daten aus älteren DV-Anwendungen, die häufig nicht heute üblichen Normalformen entsprechen.

Bei der Langzeitarchivierung von Daten aus IT-Fachanwendungen durch Migration – elektronische Akten sind bewusst ausgenommen – geht es nicht um die Bewahrung des „look and feel", sondern um die Sicherung des Informationsgehaltes der Daten und den Erhalt der Auswertungsoptionen. Dies entspricht einer Form der Authentizität, die ich als „informationelle Authentizität" charakterisieren möchte. Eine lückenlose, präzise Dokumentation der technischen Bearbeitung der Daten ist daher als Nachweis ihres in den Konvertierungs- und Kopierungsprozeduren unveränderten Informationsgehaltes unverzichtbar.

---

[4] Die maschinenlesbaren Daten einer Provenienzstelle bilden hierbei einen eigenen Bestand, der sich durch den Zusatz „MD" von der für den Schriftgutbestand gebildeten Archivsignatur unterscheidet (DA 5 Staatsrat der DDR, DA 5 MD Staatsrat der DDR – maschinenlesbare Daten).

Die vorhandenen Erschließungsinstrumente, vor allem Datenbanken, sind bislang nicht auf die besonderen Erfordernisse der Erschließung digitaler Überlieferung abgestimmt. Im Bundesarchiv erfolgte die Erschließung digitaler Archivobjekte in der beschriebenen Form bislang getrennt von der übrigen schriftlichen Überlieferung einer Provenienzstelle. Diese Entwicklung wurde durch organisatorische Rahmenbedingungen verstärkt.[5] Es wird nun aber darauf ankommen, den Erschließungszusammenhang von konventioneller und digitaler Überlieferung einer Provenienzstelle herzustellen. Es werden Modelle zu entwickeln sein, mittels derer der Nachweis und die Erschließung von Daten aus IT-Fachanwendungen in Beziehung zur Erschließung der Schriftgutüberlieferung einer Provenienzstelle erfolgen kann. Für den Nachweis elektronischer Akten wird dies in Zukunft ohnehin unabdingbar sein.

Für die interne Informationspolitik im Bundesarchiv ist der Nachweis der digitalen Überlieferung in einer zentralen und allgemein zugänglichen Erschließungsumgebung aufgrund der Aufbauorganisation und der dislozierten Struktur von ganz besonderer Bedeutung.

Aber auch für den externen online-Zugriff auf Erschließungsinformationen, sei es in einer Datenbank-Replikation im Internet oder in datenbankgestützten Online-Findbüchern, ist der Nachweis digitaler Archivobjekte in BASYS wünschenswert.

3. Erschließungsinformationen im Benutzungsprozess

Umfängliche Erschließungsinformationen zu digitalen Archivobjekten sind notwendig, aber nicht alle Informationen sind in den unterschiedlichen Stadien eines Benutzungsprozesses von Relevanz.

Der Benutzungsvorgang lässt sich in drei Phasen unterscheiden:

Erste Phase: Die Recherche in einem zentralen, allgemein zugänglichen Findhilfsmittel (Beständeübersicht, Datenbestandskatalog, Datenbank) führt den Benutzer zu einem digitalen Archivobjekt. Hierbei sollte auch die Struktur der parallelen Schriftgutüberlieferung einer Provenienzstelle wahrzunehmen sein.

Zweite Phase: Der Benutzer informiert sich in einem auf das einzelne Objekt bezogenen Dossier detailliert über Inhalt, Struktur, IT-technische und rechtliche Nutzungsbedingungen, Auswertungsoptionen etc. Vor allem die zweite Phase unterscheidet derzeit die Benutzung digitaler Überlieferung von einer gewöhnlichen Schriftgutbenutzung.

Dritte Phase: Der wissenschaftliche Benutzer wertet die Daten aus und verarbeitet sie weiter. In der Regel erfolgt dies nicht im Archiv, sondern mit Benutzungskopien am heimischen Rechner. Hiervon zu unterscheiden sind Anfragen nach Einzelinformationen, häufig privater Natur, die im allgemeinen nur durch gezielte Recherchen durch Archivpersonal zu beantworten sind.

In der ersten Phase reichen zum Nachweis der digitalen Überlieferung wenige Metadaten aus. Zu erfassen sind die Provenienzstelle und die Archivsignatur, die sprechende Bezeichnung der IT-Fachanwendung, die Laufzeit der Daten bzw. der Entstehungszeitpunkt der Daten, ggf. die Bezeichnung einzelner Dateien.

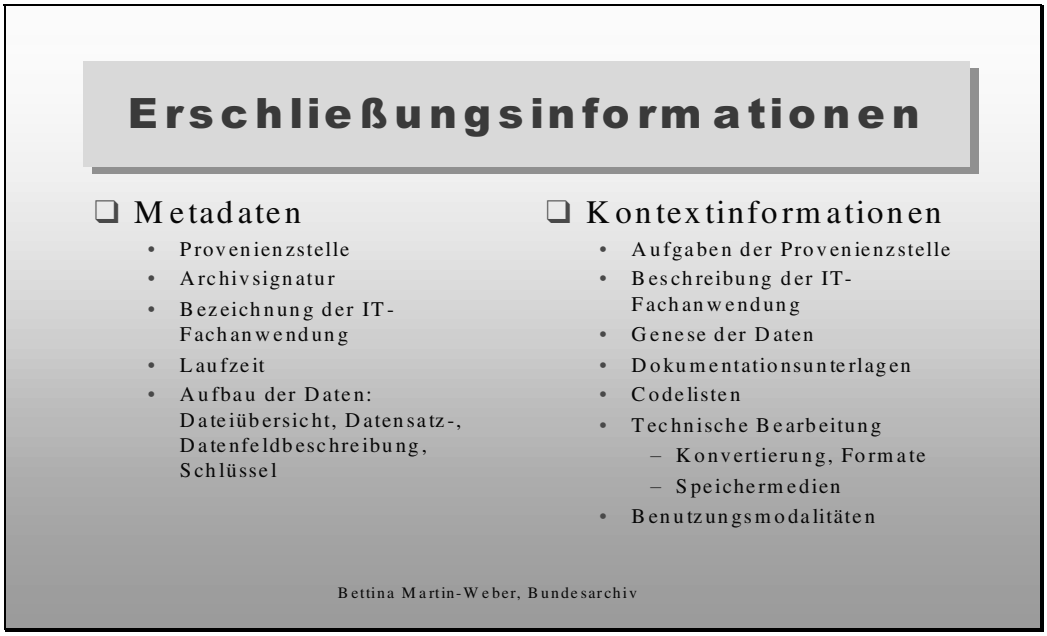
Nutzbare oder für die Aufnahme vergleichbarer Informationen bestimmte Felder existieren in der Erschließungsdatenbank im Bundesarchiv. Die im allgemeinen aussagekräftige Bezeichnung einer IT-Fachanwendung kann z.B. als Titel erfasst werden (z.B. *Kriminalstatistik*, *Zentraler Kaderdatenspeicher, Todesursachenstatistik*). Die Signatur eines digitalen Archivobjektes und die Laufzeit einer zugrundeliegenden IT-Fachanwendung sind problemlos zu

[5] Die Zuständigkeit des Referates für maschinenlesbare Dateien erstreckt sich auf drei, an verschiedenen Orten untergebrachte Abteilungen des Bundesarchivs – DDR in Berlin, Bundesrepublik in Koblenz, Militärarchiv in Freiburg – und wird derzeit aus einer vierten Abteilung LS heraus in Koblenz wahrgenommen.

erfassen. „Sprechende“ Bezeichnungen einzelner, zu einem Archivobjekt gehörende Dateien können z.B. in den Enthält-Vermerken nachgewiesen werden.

Im Zuge der Entwicklung eines Datenmodells für eine Erschließung der digitalen Überlieferung im Kontext der Schriftgutüberlieferung werden die in BASYS vorhandenen Erschließungsansätze zu prüfen sein. Dabei wird kritisch zu hinterfragen sein, welche Informationen hierin zweckmäßigerweise aufzunehmen sind.

Erst in der zweiten Phase der Benutzung werden die übrigen Metadaten und Kontextinformationen dem Benutzer zur Verfügung gestellt. Sie sollen den Entstehungszusammenhang der Daten evident halten, ihren Aufbau und Inhalt beschreiben, die technische Bearbeitung dokumentieren sowie die Benutzungsmodalitäten festhalten und sind häufig um Dokumentationsunterlagen, Codepläne u.ä. angereichert. Es handelt sich hierbei im wesentlichen um die oben beschriebenen Findbücher bzw. Dossiers.



4. „Erschließungslandschaft“

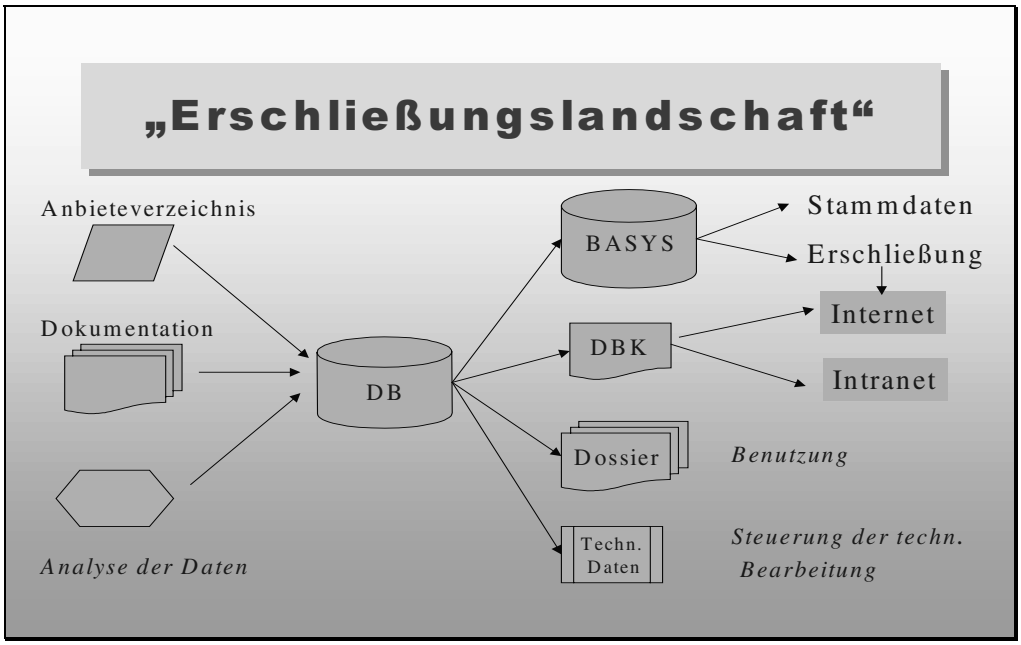
Es ist sinnlos, Daten zu archivieren, zu denen von den abgebenden Stellen keine aussagekräftigen Dokumentationsunterlagen geliefert werden, weil die Benutzbarkeit der Daten im allgemeinen nicht herzustellen ist. Daten tragen im Unterschied zu anderen Unterlagen meist nur wenige Metadaten und keine Kontextinformationen in sich. Die Archive sind darauf angewiesen, dass ihnen von den datenabgebenden Stellen alle relevanten Informationen zur Verfügung gestellt werden. Dies sollte in elektronischer Form in einem standardisierten Anbieteverzeichnis erfolgen, das Grundlage der Bewertung ist und für die spätere Erschließung durch Dokumentationsunterlagen zu ergänzen ist.[6]

Künftig sollen im Bundesarchiv die Informationen aus Anbieteverzeichnis, Auswertung der Dokumentationsunterlagen, Analyse der Daten und Protokollierung der technischen Bearbeitung in einer Datenbank zusammengeführt werden, die zunächst als „Insellösung“ in der zu-

[6] Das Bundesarchiv führt derzeit in einem Pilotprojekt im Geschäftsbereich des Bundesministeriums für Verteidigung eine Erhebung zu IT-Fachanwendungen durch, die zunächst als Bewertungsgrundlage und dann als Basis eines differenzierteren Anbieteverzeichnisses für archivwürdige Anwendungen dienen soll. Um Aufwände bei den Behörden zu minimieren, wurden in der ersten Phase keine technischen Details erhoben. Diese sollen nach erfolgter Bewertung nur für die archivwürdigen Anwendungen erfasst werden.

ständigen Organisationseinheit aufgebaut wird, dann aber als Grundlage eines Datenmodells fungieren soll, das perspektivisch in BASYS zu realisieren ist. Aus der Datenbank als Informationspool sollen die jeweils relevanten Informationen in die verschiedenen Erschließungsinstrumente und Präsentationsformen fließen und die Steuerung der technischen Bearbeitung erfolgen.

Metadaten werden nach BASYS sowohl auf der Ebene der „Stammdaten" der Bestände und Provenienzstellen als auch auf der Ebene der Archivalieneinheiten kopiert. Die Erschließungsdatenbank soll im Internet gespiegelt werden. Aus der Datenbank wird die Bestandeübersicht „Datenbestandskatalog" erstellt, die ihrerseits Grundlage der Präsentation von Bestandsinformationen im Intranet und Internet sein kann. Metadaten, Kontextinformationen und technische Beschreibung bilden die Grundlage des Dossiers. Die technische Beschreibung dient zur Steuerung der technischen Bearbeitung, also vor allem der Konvertierung und Sicherungskopierung.



5. Archivierungsstrategie und technisches Konzept

Im Bundesarchiv werden digitale Archivobjekte in nicht-proprietären Formaten als ASCII-Flat Files mit Feld- und Satztrennzeichen (csv-Dateien) gespeichert, derzeit als reine Primärdaten ohne weitere Metadaten und Kontextinformationen.

Da es sich bei den bisher archivierten Daten ganz überwiegend um DDR-Daten handelt, die nicht im Standard-EBCDIC, sondern zum Teil komprimiert und binär gepackt vorliegen, kann die Konvertierung nicht mit Standard-Tools durchgeführt werden. Vielmehr sind spezielle Konvertierungsprogramme erforderlich. Diese werden von externen Programmierern auf der Basis von Werkverträgen erstellt. Zum Auftragsvolumen gehört häufig die gleichzeitige Erstellung von Programmen zur Aufbereitung anonymisierter Benutzungsversionen personenbezogener Datenbestände. Die Konvertierung der Daten selbst und die Sicherung bzw. Umkopierung der Daten erfolgt im Bundesarchiv im Fachreferat.[7]

[7] Für die Sicherungsroutinen ließe sich ein Outsourcing denken, allerdings sind hier auch noch nicht die Quantitäten angefallen, die die Vergabe zur kostengünstigeren Variante werden ließen.

Die Daten werden derzeit auf CD-R und DAT gesichert und können auf diesen Medien auch Benutzern zur Verfügung gestellt werden.[8]

Die Schutzfristen für DDR-Bestände werden bei wissenschaftlichen Benutzungsvorhaben regelmäßig verkürzt. Damit ist auch die Benutzung von Datenbeständen im Bundesarchiv bereits Realität. Allerdings besteht vor allem hinsichtlich der noch keineswegs abgeschlossenen Erschließung und technischen Bearbeitung der DDR-Daten ein drängender Handlungsbedarf.[9]

Bei der Übernahme von Daten aus der Bundesverwaltung steht das Bundesarchiv am Anfang und muss feststellen, daß sich die für die Übernahme der Akten eingeübten Routinen keineswegs auf die Übernahme von Daten übertragen lassen, sondern in Einzelfällen auch aufgrund der mit der Datenübergabe verbundenen Aufwände mit nicht unerheblichen Widerständen bei den anbietepflichtigen Stellen zu rechnen ist.

Angesichts der Rückstände bei der Bearbeitung bereits übernommener und benutzbarer Daten und der gewaltigen Aufgaben, die vor uns stehen, ist kritisch zu fragen, was in der Nutzbarmachung digitaler Überlieferung geleistet werden muss – und was geleistet werden kann.

Auf die Bedeutung einer sorgfältigen und vielschichtigen Erschließung wurde hingewiesen. Festgestellt habe ich in diesem Zusammenhang, dass die komplexe und zeitintensive Benutzung der Daten in aller Regel nicht im Archiv erfolgt. Hinzu kommt, dass die Benutzbarkeit der DDR-Daten eine Ausnahme darstellt, für bundesrepublikanische Daten aber i.d.R. Schutzfristen einer heutigen Nutzung entgegenstehen. Es macht daher derzeit keinen Sinn, archivierte Daten in einem ständigen Zugriff auf schnellen Speichermedien, auf Jukeboxen oder in Datenbanken zu halten, wie sie die „Archivierungskonzepte“ kommerzieller Anbieter für die kurz- oder mittelfristige Sicherung und Bereitstellung der Daten anbieten. Das Verfahren, die Daten auf Speichermedien vergleichbar der konventionellen Überlieferung im Magazin zu lagern und bei Bedarf auszuheben und zu kopieren, hat sich im Bundesarchiv bewährt. Dies gilt sowohl für die Nutzung als auch für die technische Bearbeitung im Rahmen der Sicherungsstrategie.

Auch dem Anspruch an online-Verfügbarkeit der Daten kann mit skeptischer Zurückhaltung begegnet werden. Der Aufwand, den die Präsentation sehr großer Datenbestände im Internet bedeutet, sowohl in der Aufbereitung als auch in der Datenpflege, dürfte in keinem Verhältnis zum Nutzen stehen.

Eine andere Frage ist die Aufbereitung der Benutzungsdateien. Sicher ist es wünschenswert, Daten komfortabler in einer Datenbankumgebung oder mit Statistikprogrammen aufbereitet zu nutzen. Vor allem die Anschaulichkeit der Daten im Archiv wäre so zu verbessern, auch interne Recherchen und Analysen würden erleichtert. Unseren Erfahrungen nach sind wissenschaftliche Benutzer aber in der Regel an der gesamten Datenbasis interessiert, die in längerfristig angelegten Projekten Grundlage umfangreicher Auswertungen und Bearbeitungen ist, die nicht im Archiv stattfinden. Hierfür sind die Daten in einer Form zur Verfügung zu stellen, die ihre unproblematische Weiterverarbeitung in Datenbanken oder mit Statistikprogrammen gewährleistet. Mehr muss nicht geleistet werden. Die Recherchen nach Einzelinformationen aber, wie sie bei den personenbezogenen DDR-Dateien zu erwarten, bisher allerdings

---

[8] Für die Bereitstellung der Daten fallen an Auslagen 80 DM für DAT und CD und 11 DM für Disketten an. Vgl. die Verordnung über Kosten beim Bundesarchiv (Bundesarchiv-Kostenverordnung) vom 29. September 1997 (BGBl I 2380), geändert durch die 1. Verordnung zur Änderung der Bundesarchiv-Kostenverordnung vom 7. Nov. 2000 (BGBl I 1495).

[9] Die Zeit drängt, weil häufig nur durch die Nachfrage bei Personen, die mit der Entstehung der Daten befasst waren, die erforderlichen Informationen für das technische und inhaltliche Verständnis der Daten zu beschaffen sind. Auch zeigen jüngste Erfahrungen, dass keineswegs alle vorhandenen Unterlagen tatsächlich erfasst und gesichert werden konnten, sondern dass auch heute noch Dokumentationsunterlagen zu DDR-Daten in Bundesbehörden liegen.

eher selten aufgetaucht sind, können bei einer schlichten Bereitstellung der Daten z.B. als Flat Files dem Benutzer kaum zugemutet werden, weil in jedem Fall Rechner, Software und Know-how vorhanden sein müssten. Diese Recherchen sind im Archiv mit eigenen Kräften durchzuführen.

Wenn ich im Falle der Aufbereitung der Daten auch angesichts begrenzter Ressourcen für Zurückhaltung plädiere, so scheinen mir auf dem Gebiet der Qualifizierung der Archivare Investitionen erforderlich zu sein. Ich möchte diese Forderung mit dem Begriff Methodenkompetenz beschreiben. Sowohl für eine weitsichtige, inhaltlich und technisch wohl begründete Bearbeitung der Daten als auch für die Erschließung und die Benutzerberatung halte ich profunde IT-Kenntnisse der archivischen Mitarbeiter und entsprechende Schulungsangebote für unverzichtbar.

XML als Perspektive

Als Vorstudie zu dem im vergangenen Jahr in Mannheim vorgestellten Projekt zur „Elektronischen Archivierung und Nutzung (ELAN)“ hat das Bundesarchiv gemeinsam mit der Forschungsstelle für Verwaltungsinformatik der Universität Koblenz im Rahmen einer Diplomarbeit die Beschreibungs- und Auszeichnungssprache XML auf ihre Anwendbarkeit für die Langzeitarchivierung einerseits und die Webpräsentation und Online-Recherche andererseits hin untersucht.[10] Ausgewählt wurde hierfür der Archivdatenbestand „Datei Grenzzwischenfälle“ (DVH 32 MD/1). Es handelt sich um eine Datei, in der sämtliche Vorkommnisse an den Grenzen der DDR zur Bundesrepublik, zu West-Berlin, Polen und zur Tschechoslowakei zwischen dem 1. Dezember 1971 und dem 30. September 1990 erfasst sind. Pro Jahr wurden durchschnittlich 35.000 Meldungen der Grenzposten durch das Kommando der Grenztruppen der DDR aufgenommen. Insgesamt umfasst der Datenbestand 650.000 Datensätze, in ASCII-csv-Dateien ca. 180 MB.[11] Die „Datei Grenzzwischenfälle“ bot sich als Objekt an, weil sie bereits konvertiert als Flat File im Standard-ASCII-Format vorlag. Der Datenbestand wurde auch ausgewählt, weil er im Unterschied zu vielen anderen DDR-Datenbeständen des Bundesarchivs keine personenbezogenen Daten enthält und auch ohne zusätzliche Bearbeitung (Anonymisierung) keine besonderen Benutzungsprobleme aufwirft. Außerdem konnte unterstellt werden, daß gerade an diesen Daten neben dem wissenschaftlichen auch ein allgemeines öffentliches Interesse – publizistisch, lokalgeschichtlich, persönlich motiviert – besteht, eine Online-Nachfrage also wahrscheinlich ist.

Die prinzipiellen Vorteile von XML wurden bereits in verschiedenen Vorträgen anlässlich dieser Tagung thematisiert.

Ich meine, dass XML nicht nur neue Perspektiven für den systemunabhängigen Datenaustausch bei archivischen Findmitteln aufzeigt, sondern auch neue Möglichkeiten für die Langzeitarchivierung bietet, da durch die Einbettung von Informationen in ein XML-Dokument Informationsverluste begrenzt und gleichzeitig die Lesbarkeit und Verständlichkeit der archivierten Daten selbst erhöht werden kann. Einem Defizit der derzeitigen, oben beschriebenen Archivierungsstrategie, der Verlust struktureller und funktionaler Informationen bei komplexen Datenobjekten, der nur durch die Bereitstellung entsprechend aussagekräftiger Dokumentationsunterlagen und Beschreibungen zu kompensieren ist, kann mit XML wirksam begegnet werden.

---

[10] Die Diplomarbeit von Björn Dehms, Langzeitarchivierung einfacher, relationaler Datenbanken. Entwicklung eines Prototypen zur Migration nach XML, wurde im Dezember 2000 vom Fachbereich Informatik der Universität Koblenz-Landau, Abteilung Koblenz angenommen.

[11] Im Rahmen der Tagung stellte Ulrich Meyer von der Forschungsstelle für Verwaltungsinformatik das Konvertierungsverfahren ausführlich da. Vgl. den Beitrag von Dehms, Engel und Meyer in diesem Band.

Mittels XML lassen sich semantische und strukturelle Informationen mit den Primärdaten verbinden und in einem Objekt darstellen und Metadaten, Informationen zur Provenienzstelle, Dateibezeichnungen, Angaben zur Laufzeit, zur Datenstruktur usw. im datentechnischen Verbund mit den Primärdaten in ein XML-Dokument einbinden[12]. Bislang sind die archivierten Daten nur aufgrund der auf dem Speichermedium aufgebrachten Signatur und der Dateibezeichnung identifizierbar.

Es ist allerdings einschränkend festzustellen, dass die Präsentation von und die Recherche in großen Datenbeständen mit XML-eigenen Werkzeugen ohne Datenbankunterstützung nicht zu realisieren ist. Die Aufgabenstellung der Diplomarbeit ging von der Überlegung aus, dass das XML-Dokument, das für die Webpräsentation aufbereitet wurde, gleichzeitig den Anforderungen für die Langzeitarchivierung genügen sollte, also softwareunabhängig realisiert werden sollte. Auf den Einsatz eines proprietären Datenbanksystems wurde daher verzichtet. Der Versuch, die Präsentation der Daten in einer Webversion ohne Unterstützung durch ein Datenbankprogramm zu verwirklichen, schlug fehl. Mit akzeptabler Performance ließ sich nur ein sehr kleiner Ausschnitt der Daten präsentieren. Die für eine derartige Webpräsentation – die für den Benutzer nur exemplarischen Wert hat – erforderlichen Aufwände sind angesichts des zu erwartenden Nutzens nicht vertretbar.[13]

Das Bundesarchiv wird XML als Werkzeug zur Langzeitarchivierung im Auge behalten. Allerdings ist vor allem die Entwicklung leistungsfähigerer Tools zur Präsentation und Migration der Daten aus XML-Dokumenten abzuwarten. Vor allem aber dürfte die Entwicklung nativer XML-Datenbanken ganz neue Möglichkeiten eröffnen.

Das im Rahmen der Diplomarbeit erstellte Konvertierungsprogramm wurde inzwischen auf der Grundlage eines Werkvertrags leicht modifiziert und soll an weiteren Beständen des Bundesarchivs getestet werden.

Fazit

Erst die Erschließung digitaler Archivobjekte stellt deren Benutzbarkeit sicher. Vor allem in dieser Hinsicht unterscheidet sie sich von traditioneller Schriftguterschließung. Für die Bereitstellung und Verbreitung der Erschließungsinformationen sind geeignete Formen zu entwickeln. Sie sollten sich an der Funktion von Information in veränderten Benutzungsverfahren orientieren und die Zusammenführung aller Erschließungsinformationen zur gesamten Überlieferung einer Provenienzstelle zum Ziel haben. Dies ist die vordringliche Aufgabe. Langzeitarchivierung in XML würde die Voraussetzungen für die Erschließung erneut drastisch verändern, weil Metadaten, Struktur- und Kontextinformationen nicht in Findmitteln bereitgestellt, sondern weitgehend in die digitalen Archivobjekte selbst eingebunden würden. XML wird in Archiven derzeit vor allem für den Austausch archivischer Erschließungsinformationen getestet und eingesetzt. Ob es auch für die Langzeitarchivierung taugt, wird die Zukunft zeigen.

---

[12] Damit würden die archivierten Objekte sich den in OAIS als Archival Information Packages (AIP) beschriebenen Informationspaketen annähern.

[13] Die im Rahmen der Diplomarbeit konzipierte Online-Auskunft war kurzzeitig auf der Webseite der Universität Koblenz einsehbar. Sie basiert auf Windows NT und ist auf dem Linux-Webserver des Bundesarchivs derzeit nicht lauffähig. Eine Anpassung der Daten steht noch aus.