

Langzeitarchivierung digitaler Ressourcen

Ute Schwens
Hans Liegmann

Aus:

Grundlagen der praktischen Information und Dokumentation / begr. von Klaus Laisiepen, Ernst Lutterbeck u. Karl-Heinrich Meyer-Uhlenried. - 5., völlig neu gefasste Ausg. - München : Saur.

Bd. 1: Handbuch zur Einführung in die Informationswissenschaft und -praxis. - 2004. S. 567- 570.

(Die Zählung der Abschnitte stimmt nicht mit der Druckausgabe überein)

<http://nbn-resolving.de/urn:nbn:de:0008-2005110800>

B 20 Langzeitarchivierung digitaler Ressourcen

Ute Schwens und Hans Liegmann

B 20.1 Die digitale Welt, eine ständig wachsende Herausforderung

Nicht nur die Menge und Heterogenität der originär in digitaler Form vorliegenden Informationen wächst beständig an. In großem Umfang werden ursprünglich analog vorliegende Daten digitalisiert, um den Benutzerzugriff über Datennetze zu vereinfachen. Im Tagesgeschäft von Behörden, Institutionen und Unternehmen werden digitale Akten produziert, für die kein analoges Äquivalent mehr zur Verfügung steht. Sowohl die wissenschaftliche Fachkommunikation wie der alltägliche Informationsaustausch sind ohne digitale Informationsvermittlung nicht mehr vorstellbar.

Mit der Menge der ausschließlich digital vorliegenden Information wächst unmittelbar auch ihre Relevanz als Bestandteil unserer kulturellen Überlieferung sowie die Bedeutung ihrer dauerhaften Verfügbarkeit für Wissenschaft und Forschung. Denn das in der scientific community erarbeitete Wissen muss, soll es der Forschung dienen, langfristig verfügbar gehalten werden, da der Wissenschaftsprozess immer wieder eine Neubewertung langfristig archivierter Fakten erforderlich macht. Die Langzeitarchivierung digitaler Ressourcen ist daher eine wesentliche Bedingung für die Konkurrenzfähigkeit des Bildungs- und Wissenschaftssystems und der Wirtschaft. In Deutschland existiert eine Reihe von Institutionen (Archive, Bibliotheken, Museen), die sich in einer dezentralen und arbeitsteiligen Struktur dieser Aufgabe widmen.

Im Hinblick auf die heutige Situation, in der Autoren und wissenschaftliche Institutionen (Universitäten, Forschungsinstitute, Akademien) mehr und mehr selbst die Veröffentlichung und Verbreitung von Publikationen übernehmen, erscheint auch weiterhin ein verteilter Ansatz angemessen, der jedoch um neue Beteiligte an der Publikationskette erweitert werden muss.

B 20.1.1 Langzeitarchivierung im digitalen Kontext

Mit „Langzeitarchivierung“ meint man in diesem Zusammenhang mehr als die Erfüllung gesetzlicher Vorgaben über Zeitspannen, während der steuerlich relevante tabellarisch strukturierte Daten verfügbar gehalten werden müssen. „Langzeit“

ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen werden. Dabei spielen nach bisheriger Erfahrung das Nutzerinteresse der Auf- und Abwärtskompatibilität alter und neuer Systemumgebungen eine Rolle nur dann, wenn dies dem Anbieter für die Positionierung am Markt erforderlich scheint. „Langzeit“ bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.

Der Bedeutungsinhalt von „Archivierung“ müsste hier nicht näher präzisiert werden, wäre er nicht im allgemeinen Sprachgebrauch mit der fortschreitenden Anwendung der Informationstechnik seines Sinnes nahezu entleert worden. „Archivieren“ bedeutet zumindest für Archive, Museen und Bibliotheken mehr als nur die dauerhafte Abspeicherung digitaler Informationen auf einem Datenträger. Vielmehr schließt es die Erhaltung der dauerhaften Verfügbarkeit digitaler Ressourcen mit ein.

B 20.2 Substanzerhaltung

Eines von zwei Teilzielen eines Bestandserhaltungskonzeptes für digitale Ressourcen ist die unversehrte und unverfälschte Bewahrung des digitalen Datenstroms: die Substanzerhaltung der Dateninhalte, aus denen digitale Objekte physikalisch bestehen. Erfolgreich erreicht wird dieses Teilziel dann, wenn die aus heterogenen Quellen stammende und auf unterschiedlichsten Trägern vorliegenden Objekte möglichst früh von ihren originalen Träger getrennt und in ein homogenes Speichersystem überführt werden. Die verantwortliche archivierende Institution wird vorzugsweise ein funktional autonomes Teilsystem einrichten, dessen vorrangige Aufgabe die Substanzerhaltung digitaler Ressourcen ist. Wichtige Bestandteile dieses Systems sind automatisierte Kontrollmechanismen, die den kontinuierlichen systeminternen Datentransfer überwachen. Die kurze Halbwertszeit technischer Plattformen macht auch vor diesem System nicht

halt und zwingt zum laufenden Wechsel von Datenträgergenerationen und der damit möglicherweise verbundenen Migration der Datenbestände.

Dauerhafte Substanzerhaltung ist nicht möglich, wenn die Datensubstanz untrennbar an einen Datenträger und damit an dessen Schicksal gebunden ist. Technische Maßnahmen zum Schutze der Verwertungsrechte (z. B. Kopierschutzverfahren) führen typischerweise mittelfristig solche Konfliktsituationen herbei. Ein digitales Archiv wird in Zukunft im eigenen Interesse Verantwortung nur für solche digitalen Ressourcen übernehmen, deren Datensubstanz es voraussichtlich erhalten kann. Ein objektspezifischer „Archivierungsstatus“ ist in dieser Situation zur Herstellung von Transparenz hilfreich.

B 20.3 Erhaltung der Benutzbarkeit

Substanzerhaltung ist nur eine der Voraussetzungen, um die Verfügbarkeit und Benutzbarkeit digitaler Ressourcen in Zukunft zu gewährleisten. „Erhaltung der Benutzbarkeit“ digitaler Ressourcen ist eine um ein Vielfaches komplexere Aufgabenstellung als die Erhaltung der Datensubstanz. Folgen wir dem Szenario eines „Depotsystems für digitale Objekte“, in dem Datenströme sicher gespeichert und über die Veränderungen der technischen Umgebung hinweg aufbewahrt werden, so steht der Benutzer/die Benutzerin der Zukunft gleichwohl vor einem Problem. Er oder sie ist ohne weitere Unterstützung nicht in der Lage, den archivierten Datenstrom zu interpretieren, da die erforderlichen technischen Nutzungsumgebungen (Betriebssysteme, Anwendungsprogramme) längst nicht mehr verfügbar sind. Zur Lösung dieses Problems werden unterschiedliche Strategien diskutiert, prototypisch implementiert und erprobt.

Der Ansatz, Systemumgebungen in Hard- und Software-Museen zu konservieren und ständig verfügbar zu halten, wird nicht ernsthaft verfolgt. Dagegen ist die Anwendung von Migrationsverfahren bereits für die Substanzerhaltung digitaler Daten erprobt, wenn es um einfachere Datenstrukturen oder den Generationswechsel von Datenträgertypen geht. Komplexe digitale Objekte entziehen sich jedoch der Migrationsstrategie, da der für viele Einzelfälle zu erbringende Aufwand unakzeptabel ist. Aus diesem Grund wird mit Verfahren experimentiert, deren Ziel es ist, Systemumgebungen lauffähig nachzubilden (Emulation). Es werden mehrere Ansätze verfolgt, unter denen die Anwendung formalisierter Beschreibungen von

Objektstrukturen und Präsentationsumgebungen eine besondere Rolle einnimmt.

Die bisher genannten Ansätze setzen durchgängig erst zu einem späten Zeitpunkt auf, zu dem das digitale Objekt mit seinen für die Belange der Langzeitarchivierung günstigen oder weniger günstigen Eigenschaften bereits fertig gestellt ist. Darüber hinaus wirken einige wichtige Initiativen darauf hin, bereits im Entstehungsprozess digitaler Objekte die Verwendung langzeitstabiler Datenformate und offener Standards zu fördern. Welche der genannten Strategien auch immer angewandt wird, die Erhaltung der Benutzbarkeit wird nicht unbedingt korrespondieren mit der Erhaltung der ursprünglichen Ausprägung des „originalen“ Objektes. Es wird erforderlich sein, die Bemühungen zu konzentrieren auf die Kernfunktionen digitaler Ressourcen, vordringlich das, was ihren wesentlichen Informationsgehalt ausmacht.

B 20.4 Infrastruktur digitaler Archive

Der ISO-Standard 14721:2001 – Reference Model for an Open Archival Information System – OAIS (Lit. 01) beschreibt die Infrastruktur eines digitalen Archivs in Form eines Modells. Durch die Abgrenzung und eindeutige Benennung von Funktionsmodulen, Schnittstellen und Typen von Informationsobjekten ist es gelungen, eine über die Grenzen der Anwendergemeinschaften Archive, Datenzentren und Bibliotheken hinweg geltende allgemeine Sicht auf die Kernfunktionen eines digitalen Archivs zu schaffen. Dies bildet eine wertvolle Grundlage für die Nutzung von Synergien bei der Beauftragung, Planung und Implementierung produktiver Systeme durch die verantwortlichen Systembetreiber. OAIS beschreibt mehrere Funktionsmodule, die dem Datenfluss und den Arbeitsabläufen des Archivs entsprechend angeordnet sind: Eingangsbearbeitung (ingest), Metadatenverwaltung (data management), Objektspeicherung (archival storage), Erhaltung der Langzeitverfügbarkeit (preservation planning), Administration und Bereitstellung (access).

Das Funktionsmodul „Erhaltung der Langzeitverfügbarkeit“ besteht aus wiederum vier Teilsystemen, deren Aufgabe es ist, die Umgebungsbedingungen des Archivsystems zu beobachten, Auswirkungen von technischen Veränderungen möglichst früh zu erkennen und die Planungsgrundlage für die Erhaltung der Langzeitverfügbarkeit der im System gespeicherten Objekte zu liefern.

Die Funktion „Monitor Designated Community“ stellt durch Interaktion mit der Außenwelt des Archivsystems sicher, dass aktuelle Informationen über die Nutzerbedürfnisse gesammelt werden. Die gesammelten Informationen dienen dazu, die Bereitstellungsverfahren des Archivsystems laufend an die sich ändernden Gewohnheiten der Nutzerwelt anpassen zu können. Dies betrifft z.B. bevorzugte Datenformate, Zugriffsprotokolle und die allgemeine Kommunikation über die Ausschnittsstelle des Systems.

Mit der Funktion „Monitor Technology“ wird die Entwicklung digitaler Technologien in der Außenwelt des Systems beobachtet. Es sollen diejenigen Entwicklungen frühzeitig identifiziert werden, die schädliche Auswirkungen auf die Benutzbarkeit der im System gespeicherten Objekte haben können.

Die Funktion „Develop Preservation Strategies and Standards“ empfängt die Informationen der beiden ersten Module und setzt sie in Aktivitätsempfehlungen an die Systemadministration um. Solche Empfehlungen können sich z.B. auf die Anwendung neuer Standards bei der Aufnahme von Publikationen in das Archiv beziehen.

Über die Funktion „Develop Packaging Designs and Migration Plans“ werden die Vorgaben der Systemadministration technisch implementiert. Dazu gehört die Umsetzung von Migrations- und Emulationsstrategien. Die Informationspakete, die an den Schnittstellen der einzelnen Module unseres Archivsystems ausgetauscht werden, erhalten durch diese Funktionseinheit die Vorschriften über ihre innere Struktur.

B 20.5 Technische Metadaten

Die Erhebung und strukturierte Speicherung technischer Metadaten ist eine wichtige Voraussetzung für die automatisierte Verwaltung und Bearbeitung digitaler Objekte im Interesse ihrer Langzeitarchivierung. Zu den hier relevanten Metadaten gehören z.B. Informationen über die zur Benutzung notwendigen Systemvoraussetzungen hinsichtlich Hardware und Software und die eindeutige Bezeichnung und Dokumentation der Datenformate, in denen die Ressource vorliegt. Spätestens zum Zeitpunkt der Archivierung sollte jedes digitale Objekt über einen eindeutigen beständigen Identifikator (persistent identifier) verfügen, der es unabhängig vom Speicherort über Systemgrenzen und Systemwechsel hinweg identifiziert und dauerhaft nachweisbar macht.

B 20.6 Vertrauenswürdige digitale Archive

Digitale Archive stehen erst am Beginn der Entwicklung, während Archive für traditionelles Schriftgut über Jahrhunderte hinweg Vertrauen in den Umfang und die Qualität ihrer Aufgabenwahrnehmung schaffen konnten. Es werden deshalb Anstrengungen unternommen, allgemein akzeptierte Leistungskriterien für vertrauenswürdige digitale Archive aufzustellen, die bis zur Entwicklung eines Zertifizierungsprogramms reichen. Die Konformität zum OAIS-Referenzmodell spielt dabei ebenso eine wichtige Rolle wie die Beständigkeit der institutionellen Struktur, von der das Archiv betrieben wird. Es wird erwartet, dass Arbeitsmethoden und Leistungen der Öffentlichkeit präsentiert werden, sodass aus dem möglichen Vergleich zwischen inhaltlichem Auftrag und tatsächlicher Ausführung eine Vertrauensbasis sowohl aus Nutzersicht wie auch im Interesse eines arbeitsteiligen kooperativen Systems entstehen kann.

B 20.7 Verteilte Verantwortung bei der Langzeitarchivierung digitaler Ressourcen

B 20.7.1 National

Hinsichtlich der Überlegungen zur Langzeitarchivierung digitaler Quellen in Deutschland muss das Ziel sein, eine Kooperationsstruktur zu entwickeln, die entsprechend der Strukturen im analogen Bereich

- die Bewahrung und Verfügbarkeit aller digitalen Ressourcen gewährleistet, die in Deutschland, in deutscher Sprache oder über Deutschland erschienen sind,
- die Bewahrung und Verfügbarkeit der wichtigsten Objekte jedes Fachgebiets organisiert, unabhängig davon, ob es sich um Texte, Fakten, Bilder, Multimedia handelt,
- sowie die Bewahrung und Verfügbarkeit digitaler Archivalien garantiert. Das Auffinden der Materialien soll dem interessierten Nutzer ohne besondere Detailkenntnisse möglich sein, d. h. ein weiteres Ziel einer angestrebten Kooperationsstruktur muss sein,
- die Verfügbarkeit durch Zugangsportale zu erreichen, die die Nutzer dorthin lenken, wo die

Materialien liegen, wobei unter Umständen Zugriffsrechte, Kosten u. a. verwaltet werden müssen.

Beim Aufbau einer solchen Struktur sind vor allem die Bibliotheken, Archive und Museen gefordert. In Deutschland sollen in ein entstehendes Kompetenznetzwerk Langzeitarchivierung aber auch die „Produzenten“ digitaler Ressourcen, d. h. Verlage, Universitäten, Forschungseinrichtungen, Wissenschaftler sowie technische Dienstleister wie Rechen-, Daten- und Medienzentren und Großdatenbankbetreiber einbezogen werden.

B 20.7.2 Internationale Beispiele

Ein Blick ins Ausland bestärkt den kooperativen Ansatz. In Großbritannien ist im Jahr 2001 die Digital Preservation Coalition (DPC) eingerichtet worden mit dem Ziel, die Herausforderungen der Langzeitarchivierung und -verfügbarkeit digitaler Quellen aufzugreifen und die Langzeitverfügbarkeit des digitalen Erbes in nationaler und internationaler Zusammenarbeit zu sichern. Die DPC versteht sich als ein Forum, welches Informationen über den gegenwärtigen Forschungsstand sowie Ansätze aus der Praxis digitaler Langzeitarchivierung dokumentiert und weiterverbreitet. Die Teilnahme an der DPC ist über verschiedene Formen der Mitgliedschaft möglich.

In den USA ist im Jahr 2000 ein Programm zum Aufbau einer nationalen digitalen Informationsinfrastruktur und ein Programm für die Langzeitverfügbarkeit digitaler Ressourcen in der Library of Congress (LoC) verabschiedet worden (Lit. 02). Die Aufgaben sollen in Kooperation mit Vertretern aus anderen Bibliotheken und Forschung sowie kommerziellen Einrichtungen gelöst werden. Darüber hinaus hat die LoC in Folge ihrer Jubiläumskonferenz im Jahre 2000 einen Aktionsplan aufgestellt, um Strategien zum Management von Netzpublikationen durch Bibliothekskataloge und Metadatenanwendungen ergeben. Der Ansatz einer koordinierten nationalen Infrastruktur, auch unter den Rahmenbedingungen einer äußerst leistungsfähigen Nationalbibliothek wie der LoC, bestätigt die allgemeine Einschätzung, dass zentralistische Lösungsansätze den künftigen Aufgaben nicht gerecht werden können.

Im Archivbereich wird die Frage der Langzeitverfügbarkeit digitaler Archivalien in internationalen Projekten angegangen. Besonders zu erwähnen ist

das Projekt ERPANET, das ebenfalls den Aufbau eines Kompetenznetzwerks mittels einer Kooperationsplattform zum Ziel hat. InterPares ist ein weiteres internationales Archivprojekt, das sich mit konkreten Strategien und Verfahren der Langzeitverfügbarkeit digitaler Archivalien befasst. Die Anlage und Zielsetzung dieser Projekte aus dem Archivbereich verdeutlichen, wie ähnlich die Herausforderungen der digitalen Welt für alle Informationsanbieter und Bewahrer des kulturellen Erbes sind und lassen Synergieeffekte erwarten.

Ein umfassender Aufgabenbereich von Museen ist das fotografische Dokumentieren und Verfahren von Referenzbildern für Museumsobjekte. Die Sicherung der Langzeitverfügbarkeit der digitalen Bilder ist eine essentielle Aufgabe aller Museen. Im Bereich des Museumswesens muss der Aufbau von Arbeitsstrukturen, die über einzelne Häuser hinausreichen, jedoch erst noch nachhaltig aufgebaut werden.

B 20.8 Rechtsfragen

Im Zusammenhang mit der Langzeitarchivierung und -verfügbarkeit digitaler Ressourcen sind urheberrechtlich vor allem folgende Fragestellungen relevant:

- Rechte zur Durchführung notwendiger Eingriffe in die Gestalt der elektronischen Ressourcen im Interesse der Langzeiterhaltung,
- Einschränkungen durch Digital Rights Management Systeme (z. B. Kopierschutz),
- Konditionen des Zugriffs auf die archivierten Ressourcen und deren Nutzung.

Die EU-Richtlinie zur Harmonisierung des Urheberrechts in Europa greift diese Fragestellungen alle auf, die Umsetzung in nationales Recht muss aber in vielen Ländern, darunter auch Deutschland, noch erfolgen. Erste Schritte sind in dem „ersten Korb“ des neuen deutschen Urheberrechtsgesetzes erfolgt.

Literatur

- 01 Open Archival Information System – Reference Model. ISO 14721:2003.
- 02 Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program. Washington: Library of Congress, 2002. [http://www.digitalpreservation.gov/rep/ndiipp_plan.pdf]