

Hans Liegmann

Selbstbedienung oder Lieferung frei Haus?

Netzpublikationen auf ihrem Weg in Die Deutsche Bibliothek

Das Thema Langzeiterhaltung digitaler Dokumente und Publikationen hat Eingang in die fachliche und öffentliche Diskussion gefunden. Wir freuen uns darüber, denn Die Deutsche Bibliothek kann sich dieser Diskussion stellen und wird ihren Anteil zur Erhaltung des digitalen kulturellen Erbes beitragen. Im Oktober 2001 hat eine weitere Phase in den Bemühungen um die Ausweitung der Sammlung, Archivierung und Verzeichnung von Netzpublikationen an unseren Standorten begonnen¹⁾.

Archivbibliotheken und Archive, die Verantwortung für die Erhaltung von Kulturgut und Zeitdokumenten übertragen bekommen, können diese Verantwortung nur wahrnehmen, wenn die zu bewahrenden Objekte in den „Hoheitsbereich“ der jeweiligen Institution überführt werden. Erst, wenn die erforderliche Verfügungsgewalt über den Gegenstand gegeben ist, können Migrations- und Emulationsverfahren zur Sicherung der langfristigen Benutzbarkeit Anwendung finden. Physisch vorhandene Medien wie Bücher, Musiknoten, Landkarten, Mikrofilme, Tonträger sowie Aktenberge bereiten bei der Überführung in ein Archivsystem wenig Schwierigkeiten. Sie nehmen ihren Weg vom Produzenten über Transportunternehmen in die Poststelle der

Archivbibliothek und gelangen von diesem Ausgangspunkt in die meist material- oder publikationstypbezogenen Geschäftsgänge. Die ankommenden Objekte treffen in Verpackungsformen und Liefereinheiten ein, die innerhalb eines gewissen Variantenreichtums normiert sind. Geregelt Verfahren haben sich eingespielt und es ist selbstverständlich, dass Beleg- oder Pflichtexemplare frei Haus an Die Deutsche Bibliothek übersandt werden.

Für Netzpublikationen gelten jedoch andere Rahmenbedingungen. Auch sie sollen zukünftig in Der Deutschen Bibliothek archiviert und dauerhaft erhalten werden. Dazu ist es erforderlich, eine möglichst authentische Kopie der Originalversion anzufertigen und diese in ein System zu überführen, das die Langzeitbeständigkeit der Archivversion garantiert. Verändern sich die Originalversionen elektronischer Ressourcen (z. B. elektronische Zeitschriften und dynamische Websites), so müssen diese Änderungen in möglichst kurzen Abständen auf die Archivversion abgebildet werden, um die Originalversion in ihrer Änderungsgeschichte oder in ihrem jeweils aktuellen Stand widerzuspiegeln. Die erforderlichen Archivierungsaktivitäten enthalten immer einen Teilprozess, der als Transfer der Publikation bezeichnet wird. Es gibt „transferfreundliche“ digitale Dokumente, denen diese günstige Eigenschaft explizit vom Produzenten verordnet wird. Dies ist immer dann der Fall, wenn die Versandfähigkeit eine wesentliche Voraussetzung für die Marktchancen der Publikation sind. Elektronische Bücher (E-Books in Dateiform für die entsprechenden

Lesegeräte) sind Musterbeispiele für Transferfreundlichkeit. Leider weisen sie jedoch andere weniger günstige Eigenschaften auf, die ihnen eine eher ungünstige Prognose für die Langzeitverfügbarkeit geben. E-Books können komplexe Gebilde aus einer Vielzahl von Dateien sein, sie haben jedoch immer eine standardisierte äußere Hülle und eine innere Struktur, die alle notwendigen Bestandteile zusammenhält und die Konsistenz des Ganzen sichert. Sowohl Bezahlung als auch Auslieferung der meist individualisierten Kundenexemplare erfolgen über das Internet, aus Kundensicht passiv durch den Versand elektronischer Post oder aktiv durch das Herunterladen von einem Buchhandelsserver. Leider ist diese Gruppe der immanent transferfreundlichen Objekte derzeit extrem klein. Wer würde es außer den Produzenten selbst wagen, heute die Zukunft von E-Books zu prognostizieren?

Der erfolgreiche Online-Zugriff auf eine digitale Publikation sowie die Möglichkeit des teilweisen Herunterladens durch den Endnutzer sind keine Garantien für die problemlose Übermittlungsfähigkeit der Gesamtheit in ein Archivsystem. Es folgen zwei Beispiele, mit denen Die Deutsche Bibliothek Erfahrungen gesammelt hat.

Elektronische Publikationen erscheinen in einer außerordentlichen Vielfalt technischer Realisierungsformen. Die über die Transferfähigkeit entscheidenden technischen Eigenschaften stehen meistens in einem gewissen Zusammenhang mit der gewählten Präsentationsform. Bildet ein Verleger oder Urheber

die traditionelle Printvorlage (hierarchische Gliederung in Jahrgang, Heft, Artikel etc.) in der elektronischen Ausgabe ab oder entscheidet er sich für die zurzeit moderne „Portal-Lösung“? Letztere realisiert ihre Datenhaltung in der Regel durch ein Datenbanksystem auf dem Ursprungsserver. Benutzer erhalten entweder Antworten aufgrund individueller Suchanfragen, die über Formulare an das Portal übermittelt werden oder durch den Abruf personalisierter Profile. Probleme ergeben sich durch die vielfältigen proprietären Datenhaltungssysteme, die nicht in ihren Eigenheiten in den Archivsystemen nachgebaut werden können. Für den Transfer solcher datenbankbasierten Publikationen fehlen derzeit die erforderlichen Standards und Konventionen. Bilaterale Absprachen zwischen Produzent und Archiv müssen wegen fehlender Eignung zur produktiven Anwendung im Massenbetrieb vermieden werden.

Bei traditionell hierarchisch angelegten elektronischen Publikationen kommen häufig einfachere technische Lösungen zum Einsatz, die den Transfer relativ leicht ermöglichen. Häufig findet sich an dieser Stelle ein zwar stark verzweigtes, aber im Aufbau statisches System von Dateien in den internetüblichen Formaten. Alle Inhalte sind über Hyperlinks miteinander verknüpft und können über diese erreicht werden. Dabei ist es für die Beurteilung der Transfermöglichkeiten unerheblich, ob es sich um bild-, ton- oder textbasierte Informationen handelt.

Wo die traditionellen Einteilungsmuster der Welt gedruckter Publikationen nicht mehr vorhanden sind, treffen wir einerseits auf eine Vielfalt gut zu bewältigender und andererseits komplizierter zu bearbeitender Publikationen. Websites, die für das Sammelspektrum Der Deutschen Bibliothek relevant sind, sollen unabhängig von ihren technischen Eigenschaften archiviert werden können. Aus der Präsentationsform können wenig Rückschlüsse auf die Transferfähigkeit gezogen werden. Die Präsentation einer Website kann in einem großen Umfang von proprietären Software-Leistungen ihres Ursprungsservers abhängen, sodass der Transport der Inhalte in ein Archiv zwar möglich ist, dort aber die Einzelinformationen so zusammenhanglos ankommen, dass Rückschlüsse auf das Aussehen der Originalversion nicht mehr möglich sind. In der gegenwärtigen Situation ist eine starke Tendenz zur Dynamisierung digitaler Publikationen mit den entsprechenden Folgeproblemen zu erkennen. Aus Sicht einer Archivbibliothek mag dies bedauerlich sein. Jedoch ist es nur konsequent, die vorhandenen technischen Möglichkeiten für die Gestaltung neuer Medienformen optimal einzusetzen. Die Bedenken der Kulturgüterhalter gegenüber der Kurzlebigkeit der Publikationen werden sich an den erweiterten Nutzungsmöglichkeiten der gegenwärtigen Nutzer messen müssen.

Nationalbibliotheken, die Netzpublikationen übernehmen und archivieren, setzen derzeit zwei unterschiedliche Methoden zum Transfer von Internetdokumenten ein. Die Selbstbedienungsvariante wird üblicherweise als „web

harvesting“ bezeichnet. Wesentliches Element dieses Verfahrens ist ein Softwareroboter, der Dateien in den Weiten des Internet aufsuchen kann und diese anschließend in den Datenspeicher des Archivs kopiert, ohne dass der Originalserver einen besonderen Beitrag dazu leisten muss. Dabei wird der Softwareroboter so gesteuert, dass er nur einen bestimmten national begrenzten Ausschnitt des Internet einsammelt (z. B. Webseiten aller Server, deren Name mit „uk“ endet). Web harvesting weist als Methode zur Erfassung von Internetdokumenten bestimmte Schwierigkeiten auf sowie Vor- und Nachteile, die in einem Artikel der nächsten Ausgabe von Dialog mit Bibliotheken diskutiert werden. Hinsichtlich der Übermittlung von Dokumenten in das Archivsystem gilt, dass die als „unfreundlich“ bezeichneten Publikationen auch von einem web harvester nicht erfasst werden können. Ein Roboter, der an beliebige Datenhaltungssysteme die jeweils notwendigen Anfragen stellen könnte um alle enthaltenen Informationen zu extrahieren, steht noch nicht zur Verfügung. Man stelle sich nur vor, es sollten alle Katalogisate im Online-Katalog einer Bibliothek über Suchanfragen ermittelt und heruntergeladen werden.

Die Lieferung frei Haus basiert auf dem konventionellen Modell autonomer, voneinander abgrenzbarer Publikationen. Der Produzent (Lieferant) übernimmt die Verantwortung für die Konsistenz der für das Archivsystem bestimmten Kopie. Damit übernimmt er eine wichtige Teilaufgabe im Prozess der Datenübermittlung. Er stellt Datenpakete zusammen

und überträgt diese an die Archivbibliothek. Lieferungskonventionen und Transferstandards sind allgemeingültig und werden nicht zwischen Produzent und Archiv bilateral vereinbart. Die Leistungsfähigkeit des Verfahrens im Hinblick auf transferunfreundliche Publikationen hängt einerseits davon ab, ob Vereinbarungen über allgemein anwendbare Standards für ihre Übermittlung definiert werden können und ob andererseits den Produzenten der Aufwand abverlangt werden kann, dieses definierte Übergabeformat zu erzeugen.

Die Deutsche Bibliothek hat im Projekt „Networked European Deposit Library – NEDLIB“²⁾ (1998 - 2000) den dort entwickelten web harvester getestet und seitdem die weitere Entwicklung verfolgt. Nach Auffassung Der Deutschen Bibliothek ist mit dem aktuell erreichten Stand der Technik ein in Qualität und Vollständigkeit befriedigendes Ergebnis durch den Einsatz eines Software-Roboters nicht zu erreichen. Die kurzfristig erzielbaren Datenmengen mögen zwar beeindruckend sein, können jedoch die Unsicherheiten im Hinblick auf das Qualitätsniveau nicht ausgleichen. Die Deutsche Bibliothek hat sich deshalb auf die Einrichtung einer Schnittstelle für die Übermittlung von Metadaten und Netzpublikationen (Lieferung frei Haus an Die Deutsche Bibliothek) konzentriert. Seit Oktober 2001 ist diese Schnittstelle mit einer ausgewählten Gruppe von Verlagen und verlegenden Stellen in Betrieb. Sie ist auf Skalierbarkeit und weitgehende Automatisierung der Transferabläufe angelegt. Über Er-

gebnisse aus der Evaluation der Ausbaustufen dieses Verfahrens wird Die Deutsche Bibliothek zu gegebener Zeit berichten. Für die Zukunft sieht Die Deutsche Bibliothek einen dringenden Bedarf in der Formulierung möglichst allgemeiner und mit vertretbarem Aufwand anwendbarer Standards für den Transfer dynamischer Publikationen. Experimente in diesem Zusammenhang werden durchgeführt. Web-harvesting-Techniken werden als ergänzende Maßnahme für bestimmte Publikationstypen in Zukunft erwogen. Unternehmungen mit dem Ziel, eine vollständige Archivierung des Internets zu erreichen, wird Die Deutsche Bibliothek genau beobachten und die versprochenen mit den tatsächlichen Leistungen vergleichen³⁾.

Anmerkungen

1

<<http://deposit.ddb.de>>

2

<<http://www.konbib.nl/coop/nedlib>>

3

<<http://www.archive.org>>