

Zeitschrift: Geschichte und Informatik = Histoire et informatique
Herausgeber: Verein Geschichte und Informatik
Band: 13-14 (2002-2003)

Artikel: Langzeiterhaltung digitaler Ressourcen in einer Archivbibliothek
Autor: Liegmann, Hans
DOI: <https://doi.org/10.5169/seals-118854>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 11.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Langzeiterhaltung digitaler Ressourcen in einer Archivbibliothek

Hans Liegmann

Résumé

Les changements provoqués par l'avènement des publications électroniques au sein des bibliothèques d'archive sont présentés dans cette contribution. Ils englobent divers mécanismes nouveaux, qui doivent être créés pour la saisie, la collection, la description et l'archivage des publications électroniques. Tout en prêtant une attention particulière aux attentes spécifiques des historiennes et historiens, l'article présente les expériences faites dans le domaine de l'archivage de publications académiques et la coopération avec les éditeurs et institutions en Allemagne. Des recommandations en sont issues qui devraient permettre d'assurer l'archivage des ressources numériques malgré des intérêts divergents.

Zusammenfassung

Der Beitrag thematisiert die Veränderungen, die sich im Tätigkeitsfeld einer Archivbibliothek durch das Aufkommen des elektronischen Publizierens ergeben haben. Diese umfassen mehrere neuartige Mechanismen, die für die Erfassung, Sammlung, Erschliessung und Archivierung elektronischer Publikationen geschaffen werden müssen. Auf die vermutete spezifische Erwartungshaltung von Historikerinnen und Historikern an eine Archivbibliothek wird dabei insbesondere eingegangen. Der Beitrag stellt ausserdem die Erfahrungen dar, die bei der Archivierung von Hochschulschriften und bei der Kooperation mit publizierenden Verlagen und Institutionen in Deutschland gesammelt wurden. Daraus werden Empfehlungen für Rahmenbedingungen abgeleitet, wie die Langzeiterhaltung digitaler Ressourcen in einem Spannungsfeld widerstreitender Interessen gewährleistet werden kann.

Die Deutsche Bibliothek – Aufgabe und Sammelauftrag

Die Deutsche Bibliothek ist die zentrale Archivbibliothek und das national-bibliografische Informationszentrum Deutschlands und erfüllt die Funktion einer Nationalbibliothek. Sie entstand 1990 mit der Wiedervereinigung Deutschlands aus den Vorgängereinrichtungen Deutsche Bücherei Leipzig (gegr. 1912) und Deutsche Bibliothek Frankfurt am Main (gegr. 1947), zu der seit 1970 das Deutsche Musikarchiv Berlin gehört. Sie hat die Aufgabe, alle deutschen und deutschsprachigen gedruckten und trägergebundenen elektronischen Publikationen sowie Musikalien und Tonträger ab 1913 zu sammeln, dauerhaft zu archivieren, umfassend zu dokumentieren, sowie öffentlich und uneingeschränkt zugänglich zu machen.

Das «Gesetz über die Deutsche Bibliothek» vom 31. März 1969 in der Fassung des Einigungsvertragsgesetzes vom 23. September 1990¹ gibt ihr das Pflichtexemplarrecht für Deutschland, das jeden gewerblichen oder nicht gewerblichen Verleger verpflichtet, von seinen Neuerscheinungen zwei Exemplare abzuliefern, die an einem der Standorte bibliografisch erschlossen und an jeweils zwei Standorten archiviert werden.

Netzpublikationen als neue Herausforderung

Bislang erfasst das Gesetz über Die Deutsche Bibliothek diejenigen deutschen Publikationen, die unter Verwendung eines physischen Informationsträgers (Papier, Mikroform, Datenträger für magnetische und optische Aufzeichnungsverfahren) hergestellt und vertrieben werden.

Das «trägerlose» digitale Publizieren mit ausschliesslicher Verteilung über Kommunikationsnetze hat jedoch mittlerweile als eigenständige Plattform unverzichtbare Bedeutung für den wissenschaftlichen Informationsaustausch erlangt. Als Grundlagenarbeit und im Rahmen der Vorbereitung einer Gesetzesnovellierung zum Einschluss digitaler Publikationen in den Sammelauftrag entwickelt und erprobt Die Deutsche Bibliothek seit 1998 Verfahren zur Sammlung ausschliesslich digital veröffentlichten Kulturgutes und beabsichtigt, den dauerhaften Zugriff im Interesse der öffentlichen und wissenschaftlichen Nutzung zu erhalten.

Der Begriff «Netzpublikation» wurde im Rahmen der Diskussion um die Gesetzesnovellierung als Arbeitsterminus für eine nicht in körperlicher Form erschienene Publikationseinheit geschaffen und als Gegenstand des

1 Gesetz über die Deutsche Bibliothek (DBIBLG), <<http://www.ddb.de/wir/pdf/gesetz.pdf>>

Interesses für die Ausdehnung der Sammelaktivitäten benannt. Auf den ersten Blick vermittelt er eine sehr traditionelle Sicht der Begriffsschöpfer: Die Zeiten der bibliografischen Einheit, die von anderen bibliografischen Einheiten eindeutig abgrenzbar und autonom existiert, sind vorbei. Vielmehr gelten für Archivbibliotheken alle Informationen, die im Internet einem uneingeschränkten Personenkreis zugänglich sind, als «publiziert». «Netzpublikationen» bilden eine völlig neue Publikationsgruppe mit einer eigenen, nach verschiedenen Kriterien aufzuschlüsselnden Typologie und neuartigen Eigenschaften hinsichtlich Herstellungsverfahren, Zugangs- und Nutzungsformen.

Eingeschlossen sind also trotz der konservativen Terminologie alle Formen netzbasierter Veröffentlichungen, unabhängig davon, ob sie von einem Verlag oder einer Privatperson publiziert werden, ob ihre Nutzung kosten- oder lizenzpflichtig ist und unabhängig von ihren technischen Eigenschaften. Was bedeutet dies für eine Nationalbibliothek, die ihr bisheriges territoriales, objektives und weitgehend wertungsfreies Sammelprinzip weiterverfolgen will?

Ablieferungspflichtig sind grundsätzlich alle Netzpublikationen von Rechteinhabern, die ihren Sitz in Deutschland haben. Schon für Publikationen auf analogen Trägern (z.B. Schallplatten und Audio-CDs) kam es in der Vergangenheit in Einzelfällen zum Rechtsstreit mit multinationalen Konzernen darüber, ob auch nationale Vertriebsniederlassungen im Gesetzessinne als «Sitz» anzusehen sind. Hier werden sich in Zukunft bei der Kooperation von Hosts und Content Providern vergleichbare Konfliktfälle für Netzpublikationen entwickeln, denen wir durch eine Sammlung nach einem unter den Nationalbibliotheken abgestimmten Herkunftslandprinzip begegnen wollen. In der Kontinuität ihrer Sammeltradition als Archivbibliothek sieht es Die Deutsche Bibliothek als ihre Aufgabe an, diejenigen Netzpublikationen zu verzeichnen und zu bewahren, die der Vermittlung von Informationsgehalt dienen. Dies schließt instrumentale Software wie Textverarbeitungsprogramme, Betriebssysteme und Gerätetreiber aus. Erfahrungen aus der Anwendung der Pflichtexemplarregelungen haben über Jahrzehnte zu einem fein differenzierten Komplex von Sammelrichtlinien für gedruckte Publikationen geführt. Diese werden von Der Deutschen Bibliothek offen gelegt². Wichtig war Der Deutschen Bibliothek in der

2 Sammelrichtlinien für Die Deutsche Bibliothek. - Stand: 1. September 1997. - 3., überarb. Aufl. - Leipzig; Frankfurt am Main; Berlin: Die Deutsche Bibliothek, 1997

Vergangenheit, keine inhaltlichen, sondern ausschliesslich objektiv nachvollziehbare formale Kriterien bei der genauen Bestimmung des Sammelumfanges anzulegen. Die Höhe der vervielfältigten Auflage, die Anzahl der Druckseiten, die Zweckbestimmung von Publikationen als Akzidenzien³ oder als Werbung für den Vertrieb einer Ware sind bewährte Kriterien, die es erlaubt haben, die Lücke zwischen mangelnder personeller Kapazität und der Masse der Publikationen auf eine allgemein akzeptable Art und Weise zu verkleinern.

Zur angemessenen Bewältigung der Netzpublikationen stellen sich alle Fragen nun aufs Neue. Eine Auflagenhöhe als Massstab gibt es nicht mehr. Eine Netzpublikation steht typischerweise nur auf einem einzigen Server zur Verfügung und wird entweder einmal pro Jahr oder mehrere tausend Mal pro Tag zur Nutzung aufgerufen, ohne dass die Nutzungsfrequenz als objektiver Indikator der Relevanz für die Archivierung zur Verfügung stünde. Eine Abgrenzung nach Mindestumfang ist wegen der Vielfalt der Datenformate und ihres unterschiedlichen Speicherbedarfs ebenso unmöglich. Die Information einer Druckseite kann digital sowohl in Hunderten wie auch in Millionen von Bytes dargestellt werden.

Welche Erwartungen haben künftige Nutzer und insbesondere Historiker an diejenigen Nationalbibliotheken, die digitale Publikationen heute zu ihrem Sammelgegenstand machen? Es scheint, dass diese Frage noch beantwortet werden muss, denn zu vielfältig sind derzeit noch die Ansätze, als dass ein homogenes Erscheinungsbild des angestrebten Ergebnisses deutlich würde. Im Folgenden soll anhand unterschiedlicher produktiver und prototypischer Sammelmethode der Deutschen Bibliothek deutlich werden, von welcher Bedürfnislage ausgegangen wird und welche Problematik der jeweiligen Methode zu Eigen ist.

Das Internet wird von Geschichtswissenschaftlern vielfältig genutzt. Es ist Publikationsmedium, Informationsquelle und Diskussionsforum.

Digitales Publikationsmedium und Erhalt der Langzeitverfügbarkeit

Die Prognose digitaler Publikationen im Hinblick auf ihre dauerhafte Erhaltung und Verfügbarkeit wird bereits im Erstellungsprozess festgelegt. Dies ist vor allem für diejenigen Wissenschaften von grosser Bedeutung, deren Publikationen über einen relativ langen Zeitraum nach Erscheinen einem Nutzungsinteresse unterliegen. Es seien aus Sicht einer Archiv-

3 Z. B. Werbeanzeigen, Geschäfts- und Privatdrucke.

bibliothek einige Empfehlungen an die publizierenden Historiker formuliert, die einen günstigen Einfluss auf die zukünftige Beständigkeit digitaler Publikationen haben und die Aufgaben der Archivbibliotheken zu bewältigen helfen.

Einhaltung von Standards bei Herstellung und Vertrieb von Netzpublikationen

Anstrengungen zur Erhaltung der Langzeitverfügbarkeit digitaler Ressourcen können nur dann wirksam umgesetzt werden, wenn Vielfalt und Grad der individuellen Ausprägung der einzelnen Objekte möglichst gering ist. Archivbibliotheken wie Die Deutsche Bibliothek arbeiten im Verbund mit anderen Gedächtnisinstitutionen daran mit, die Standardisierung auf allen für die Langzeiterhaltung relevanten Gebieten wie z.B. Metadaten, Dokumentenformate, Datenaustauschformate und technische Verfahren zur Bestandserhaltung (Emulation, Migration) voranzutreiben. Uns ist bewusst, dass dies im Einklang mit den Rahmenbedingungen der Produktion und des Vertriebs zu geschehen hat, um Erfolg versprechend zu sein. So muss ein Kompromiss zwischen den Anforderungen an die Individualität der Produktmerkmale und die «langzeitarchivierungsfreundlichen» Produkteigenschaften gefunden werden. Derzeit ist die Verwendung von digitalen Editionstechniken verständlicherweise primär vom innovativen Interesse geprägt. Bei der Planung entsprechender Vorhaben sollten Vorgaben zur Standardisierung und Interoperabilität berücksichtigt werden.

Daten- und Dokumentformate

Bei der Bereitstellung von Netzpublikationen sollte darauf geachtet werden, dass offene und weit verbreitete Daten- und Dokumentformate Verwendung finden. Proprietäre Formate mögen im Einzelfall die Publikation aus ihrem Marktumfeld herausheben, schmälern aber die Chancen auf Langzeitverfügbarkeit erheblich.

Die Verwendung XML-basierter Formate mit einer strikten Trennung von Inhalt und Layout ist Stand der Technik und bietet eine gute Prognose auch dann, wenn das Systemumfeld wechselt. In der Realität wird derzeit ganz überwiegend PDF zur Bereitstellung von Texten an Endnutzer verwendet. Für eine auf Seitenbasis zitierfähige stabile Darstellung in unterschiedlichen Umgebungen gibt es zum Format «PDF» derzeit keine Alternative, obgleich es im strengen Sinne nicht zu den «offenen» Formaten gezählt werden kann.

Aktivitäten zur Einführung eines auf die langfristige Archivierung von Dokumenten spezialisierten PDF-Subsets «PDF/A»⁴ messen wir grosse Bedeutung bei und wir beabsichtigen, in Standardisierungsgremien daran mitzuwirken.

Metadaten

Digitale Ressourcen eröffnen die Möglichkeit, dass beschreibende und identifizierende Metadaten das Produkt von seiner Entstehung beim Autor über die Vermarktung bis zur Archivierung begleiten. Im Umfeld kommerzieller Verlage setzt sich gerade das Format ONIX⁵ als Standard durch. In Wissenschaft und Forschung wird das Metadatenset der «Dublin Core Metadata Initiative - DCMI»⁶ angewendet.

Bereits so frühzeitig wie möglich sollte in der Prozesskette einer dieser Standards durchgängig verwendet werden. Das Datenhaltungssystem des originär publizierenden Servers sollte in der Lage sein, Metadaten in einer dieser standardisierten Strukturen auszugeben bzw. dauerhaft und zugriffsfähig im Kontext des Publikationsangebotes vorzuhalten. Für die Langzeitarchivierung digitaler Ressourcen sind insbesondere technische Metadaten von Bedeutung, die Informationen über die notwendigen Systemvoraussetzungen für die Nutzung einer Netzpublikation liefern. Im Jahr 2003 ist es selbstverständlich, eine CD-ROM ohne viel Nachdenken in das Laufwerk eines nach gegenwärtigen Kriterien definierten Standard-PC einzulegen. In unseren Archiven befinden sich jedoch bereits heute Medien, deren Systemumgebung vom Markt verschwunden sind (Video-Disks, Disketten alter Betriebssysteme). Die Zuordnung technischer Metadaten soll es in Zukunft ermöglichen, die gefährdeten Objekte (auf physischen Medien und als Netzpublikationen) als Teilmengen zu bestimmen, um sie geeigneten Verfahren der Langzeiterhaltung (Migration, Emulation) zu unterziehen.

Mehrere internationale Projekte identifizieren derzeit die für die Langzeitarchivierung notwendigen zusätzlichen Datenelemente für die oben genannten Standardformate.⁷

4 PDF-Archive Committee, <http://www.aiim.org/pdf_a>, 28. September 2003

5 ONIX for Books, <<http://www.editeur.org/onix.html>>, 28. September 2003

6 Website Dublin Core Metadata Initiative, <<http://dublincore.org>>, 28. September 2003

7 PREMIS (PREservation Metadata: Implementation Strategies) <<http://www.oclc.org/research/projects/pmwg/>>, 26. Januar 2004;

National Information Standards Organization and AIIM International (Hg.): «Data Dictionary – Technical Metadata for Digital Still Images», <http://www.niso.org/standards/resources/Z39_87_trial_use.pdf>, 26. Januar 2004;

Eindeutige und beständige Identifikation

Die eindeutige und beständige Identifikation von Netzpublikationen unabhängig vom Ort der Speicherung ist ein wichtiger Faktor bei der Erhaltung der Langzeitverfügbarkeit. Wenn Publikationen ihren Speicherort wechseln (Verlag oder Host ändert sich, Auslagerung der Netzpublikation wegen des Wegfalles der wirtschaftlichen Nutzung von einem Verlagsserver in ein Archivsystem), ändert sich zwangsläufig die Dokumentadresse im WWW (URL). Alle bislang veröffentlichten Zitate verlieren damit ihre Quellenangabe und resultieren in dem gut bekannten «HTTP 404 - File not found». Die Zuordnung einer beständigen Identifikation wie eines «Digital Object Identifier - DOI»⁸ oder einer «National Bibliography Number - NBN» des URN-Systems⁹ sollte zu den Rahmenbedingungen der Veröffentlichung einer Netzpublikation gehören wie sich dies im Hinblick auf die selbstverständliche Nutzung von ISSN oder ISBN für gedruckte Publikationen etabliert hat.

Digital Rights Management (DRM)

DRM-Verfahren werden eingesetzt, um die bestehenden wirtschaftlichen Verwertungsrechte an digitalen Publikationen zu schützen. DRM-Technologien behindern und gefährden jedoch gleichzeitig die Anstrengungen zur Langzeiterhaltung erheblich. Kopierschutzverfahren machen eine Migration auf langzeitstabile Datenträger unmöglich. Die Bindung an individuelle Systemumgebungen führt dazu, dass die Verfügbarkeit der Publikation mit dem Untergang der Systemumgebung endet. Beispiele dafür sind proprietäre eBook-Reader, Bindung an die Geräte-Identifikation eines Hardware-Bausteins oder eines Endgerätes mit zwangsläufig nur begrenzter Funktions- und Lebensdauer. Institutionen, die mit der Erhaltung der Langzeitverfügbarkeit betraut sind, müssen daher mit einer Version der Publikation ausgestattet werden, die die notwendigen langzeiterhaltenden Aktivitäten erlaubt. Wahrscheinlich wird sich diese Version von der marktüblichen dadurch unterscheiden, dass sie keine DRM-Schicht enthält. Sie wird im Archiv entsprechend sicher vor Missbrauch geschützt bewahrt werden müssen.

National Library of New Zealand (Hg.): «Metadata Standards Framework – Preservation Metadata», <http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf>, 26. Januar 2004

8 Digital Object Identifier, <<http://www.doi.org/>>, 28. September 2003

9 Die Deutsche Bibliothek (Hg.): Persistent Identifier, <<http://www.persistent-identifier.de>>, 28. September 2003

Transfer der Netzpublikation in die Archivbibliothek

Netzpublikationen werden üblicherweise nicht dort archiviert, wo sie produziert oder dem Markt zugänglich gemacht werden. Statische Angebote von Netzpublikationen in einer spiegelungsfähigen (mirror-fähigen) Form werden mehr und mehr durch dynamische Angebote abgelöst, die zum Zeitpunkt der Benutzeranfrage eine auf die Interessen des Benutzers angepasste individuelle Antwort liefern.

Als technische Infrastruktur dienen Datenhaltungssysteme (Content-Management-Systeme, CMS), deren Funktionalität überwiegend auf der Grundlage proprietärer Anwendungen und ebensolcher Datenstrukturen aufgebaut ist. Die in diesen Systemen gespeicherten Netzpublikationen (z.B. elektronische Zeitschriften-Artikel) entstehen erst auf Anforderung an der Oberfläche der Systeme zum Benutzer hin und liegen weder archivierungsfähig zur Auslieferung an noch wären sie zur Abholung durch Archivierungsinstanzen bereit. Die dynamische Generierung der Publikation durch serverseitige Implementierungen erst auf die aktuelle Benutzeranforderung hin kann durch den derzeitigen technischen Stand automatischer Sammelverfahren nicht bewältigt werden.

Hier liegt es im ureigenen Interesse der publizierenden Autoren, im Veröffentlichungsprozess auf die Transferfähigkeit der geschaffenen Dokumente hinzuwirken. Spezifische Rahmenbedingungen für bestimmte wissenschaftliche Publikationsformen (z.B. Editionen historischer Quellen) sollten definiert und durch die gemeinsame Anwendung in einem Wissenschaftskontext gestärkt werden.

Mittel- und langfristig ist nur die Standardisierung offener Transferformate erfolgversprechend, die in Kooperation mit den Produzenten zu definieren und von den eingesetzten Datenhaltungssystemen einzuhalten sind, um eine möglichst gleichartige Schnittstelle zu den Archivsystemen der Gedächtnisorganisationen zu gewährleisten.

Dienstleistung von Archivbibliotheken an Historiker

Welche Dienstleistungen erwarten Historiker von einer Archivbibliothek?

Roy Rosenzweig berichtet in einem «American Historical Review Forum Essay» vom Juni 2003¹⁰ über das Abschalten der Website «Satiric Bert Is Evil» am 11. Oktober 2001. Sie wurde wegen öffentlicher Entrüstung

¹⁰ Rosenzweig, Roy: «Scarcity or Abundance? Preserving the Past in a Digital Era», <<http://chnm.gmu.edu/assets/historyessays/scarcityp.html>>, 28. September 2003

über die «political incorrectness» ihrer Inhalte nach den Anschlägen des 11. September von ihrem Rechteinhaber aus dem Netz genommen. Wären die Inhalte auf Papier veröffentlicht worden, so fänden sie sich in einer oder mehreren Bibliotheken wieder und stünden morgen und in Zukunft (vielleicht unter erschwerten Zugangsbedingungen) zur Nutzung bereit. Als Netzpublikation gehen sie ohne entsprechende Vorsorge einfach von heute auf morgen unter.

Als Bibliothekar ist man mehr geübt darin, über die Interessen der vielschichtigen Benutzergruppen zu mutmassen, als eine tatsächliche Erwartungshaltung zu erfassen und durch Nutzerbefragungen zu evaluieren. So auch hier nur die Annahme, dass die Aufgabenwahrnehmung unserer nationalen Archivbibliothek in der besonderen, durch verteilte Zuständigkeiten geprägten Situation in Deutschland in den vergangenen Jahrzehnten die Bedürfnisse im Wesentlichen erfüllt hat. Die Sammlung von gedruckten Publikationen ist durch die möglichst umfassende und ohne wertende Selektion bestimmte Erfassung eines definierten geografischen und sprachlichen Raumes bestimmt. Die Verzeichnung folgt den im deutschsprachigen Raum üblichen Standards mit Kompatibilität zu international üblichen Regelungen. Für den physischen Erhalt der Materialien werden erhebliche Mittel eingesetzt, die in den Betrieb von Massensäuerungsanlagen, die Schaffung günstiger Lagerbedingungen und die Herstellung digitaler Benutzungssurrogate fließen.

Archivbibliotheken werden den Umfang dessen, was sie an Netzpublikationen in ihren Bestand aufnehmen wollen, präzise definieren müssen. Dabei werden sie mehrere Einflussfaktoren einbeziehen und in ihrem Verhältnis zueinander abwägen. Das bisherige Ziel der Wertungsfreiheit, alle Publikationen zu erhalten, um dem Historiker der Zukunft eine lückenlose Quellenlage präsentieren zu können, gerät in Konflikt mit der Menge des zu bewältigenden Materials, den technologischen Herausforderungen, die zur Erzielung einer konsistenten Präsentation in der Zukunft erforderlich sind und den gewohnt hohen Ansprüchen an den Umfang der formalen und inhaltlichen Erschliessung je bibliografischer Einheit.

Die angestrebte Neuregelung des Gesetzes über Die Deutsche Bibliothek wird den Anspruch der umfassenden Sammlung und Archivierung aufrechterhalten. Die Umsetzung dieses Anspruches für digitale Publikationen erfordert jedoch schon auf dem ersten Schritt ihres Weges vom Originalserver in das Archiv innovative Lösungen und Offenheit für die Entwicklungen der Zukunft.

Transferwege in die Archivbibliothek

«Körperlich» verbreitete Werke¹¹ gehen ihren Weg von der abgabepflichtigen Stelle über Transportunternehmen in die Poststelle der Archivbibliothek und werden von dort aus in die meist material- oder publikationstypbezogenen Geschäftsgänge weitergeschleust. Die ankommenden Objekte treffen in Verpackungsformen und Liefereinheiten ein, die innerhalb eines gewissen Variantenreichtums normiert sind. Geregelte Verfahren haben sich eingespielt und es ist sogar per Verordnung geregelt, dass Pflichtexemplare frei Haus an die Standorte Der Deutschen Bibliothek zu senden sind.

Für Netzpublikationen jedoch gelten andere Rahmenbedingungen. Es ist 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 Original-Version in ihrer Änderungsgeschichte oder in ihrem jeweils aktuellen Stand wiederzuspiegeln. Die erforderlichen Archivierungsaktivitäten beinhalten immer einen Teilprozess, den wir als «Transfer» der Publikation bezeichnen. 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 ist. Elektronische Bücher (eBooks in Dateiform für die entsprechenden Lesegeräte) sind gute Beispiele für Transferfreundlichkeit, aber wenig erfolgreich am Markt. eBooks können komplexe Gebilde aus einer Vielzahl von Dateien sein, sie haben jedoch immer eine standardisierte äussere Hülle und innere Struktur, die alle notwendigen Bestandteile zusammenhält und die Konsistenz der autonomen Einheit sichert.

Leider sind der erfolgreiche Online-Zugriff auf eine digitale Publikation und auch die Möglichkeit zum vollständigen oder ausschnittsweisen Her-

11 Als körperliche Verbreitung urheberrechtlich geschützter Werke gilt die Weitergabe auf einem anderen Datenträger. Hingegen gilt die Verbreitung in Netzwerken nicht als körperlich, weil der Nutzer kein physisch greifbares Werkexemplar erhält. Im deutschen Recht werden diese grundsätzlich als unterschiedliche Nutzungsarten betrachtet und die entsprechenden Rechte separat übertragen. Diese Theorien können nicht 1:1 auf das schweizerische Urheberrecht übertragen werden, da in der Schweiz teilweise andere gesetzliche Voraussetzungen vorhanden sind.

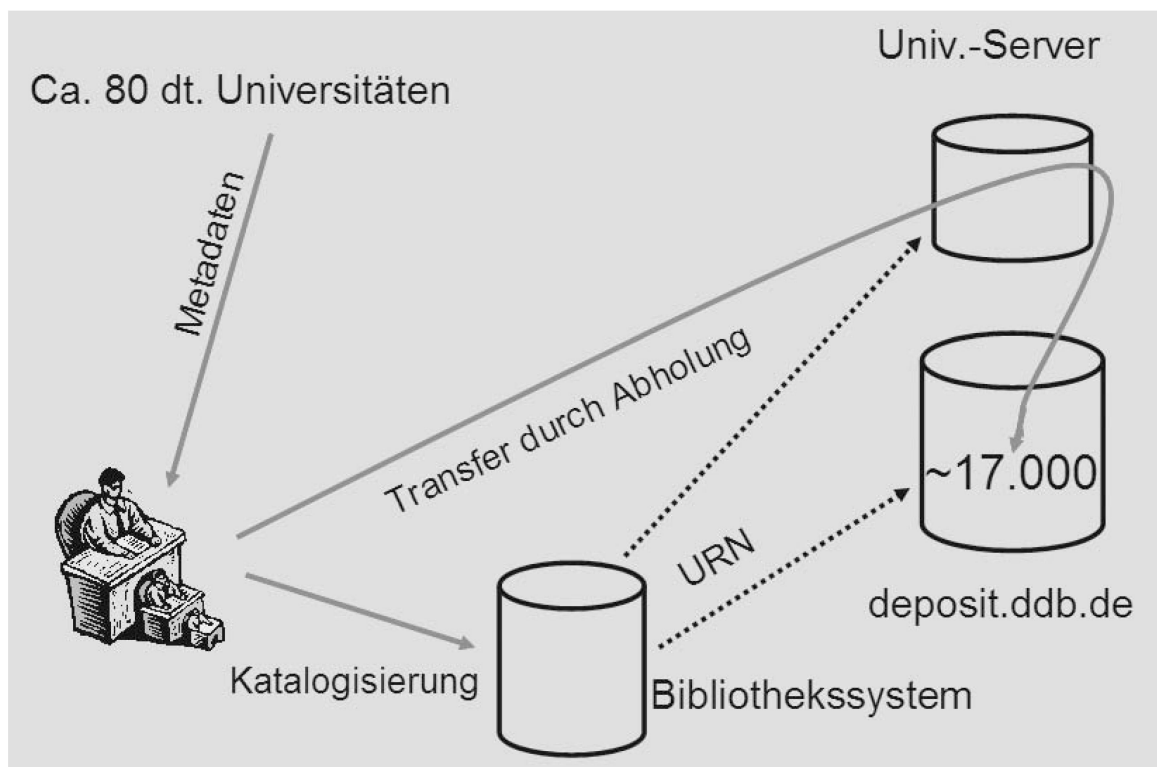
unterladen durch den Endnutzer keine Garanten für die problemlose Übermittlungsfähigkeit der Gesamtheit in ein Archivsystem.

Die Deutsche Bibliothek sammelt seit 1998 Erfahrungen mit einer Vielzahl von Transferwegen und setzt diese sowohl experimentell als auch produktiv ein.

Sammlung von Online-Hochschulschriften

Seit 1998 ist Die Deutsche Bibliothek in einem kooperativen und inzwischen von ihr koordinierten Projekt aktiv, in dem die Veröffentlichung digitaler Dissertationen in allen damit in Zusammenhang stehenden Aspekten gefördert wird.

Zur dauerhaften Archivierung der Dissertationen bei Der Deutschen Bibliothek wurde ein Verfahren zum Austausch der beschreibenden Metadaten und der Volltexte für Online-Hochschulschriften eingerichtet.



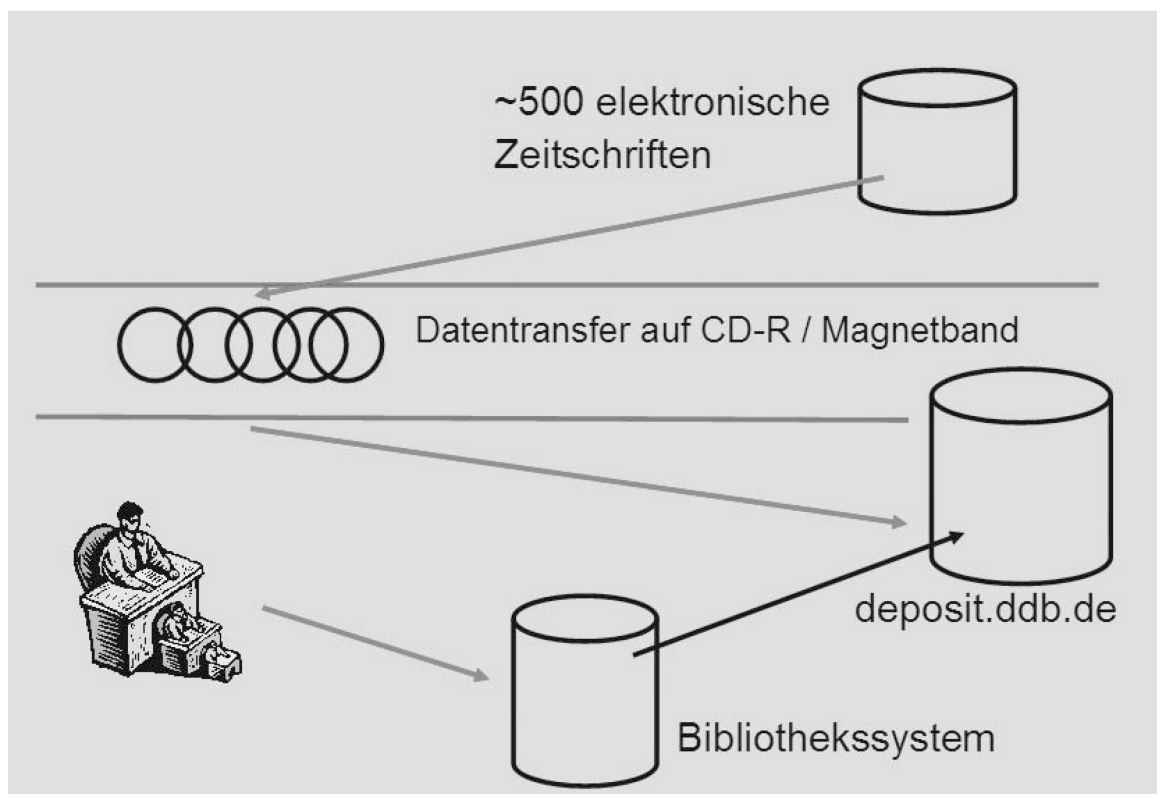
Circa 80 deutsche Universitäten liefern an Die Deutsche Bibliothek Metadaten von Online-Hochschulschriften in einem vereinbarten, auf Dublin Core beruhenden Metadaten-set auf verschiedenen Transferwegen (eMail, Anmeldeformular, Programmschnittstelle). In Der Deutschen Bibliothek werden unter Nachnutzung der angelieferten Metadaten Katalogisate auf nationalbibliografischem Niveau erstellt. Archivkopien der Online-Disser-

tationen werden ausgehend von den Originalen auf den universitären Publikationsservern erstellt und auf dem Archivserver¹² Der Deutschen Bibliothek dauerhaft gespeichert und verfügbar gehalten. Die Referenzierung der Volltexte zwischen Katalogisat, Original-Hochschulschrift und Archivkopie erfolgt durch einen beständigen Identifikator vom Typ URN (Uniform Resource Name).

Das Verfahren steht mittlerweile zur Revision und weiteren Automatisierung an: Metadaten sollen künftig über die Schnittstelle der Open Archives Initiative - Protocol for Metadata Harvesting OAI-PMH¹³ - aufgesammelt werden (siehe unten).

Archivierung des Informations-Servers SpringerLink

Für die elektronischen Publikationen des Springer-Verlags (Heidelberg, Berlin)¹⁴ wurde ein prototypischer Datentransfer zwischen dem Verlag und Der Deutschen Bibliothek eingerichtet.



¹² Archivserver Der Deutschen Bibliothek, <<http://deposit.ddb.de>>, 26. Januar 2004

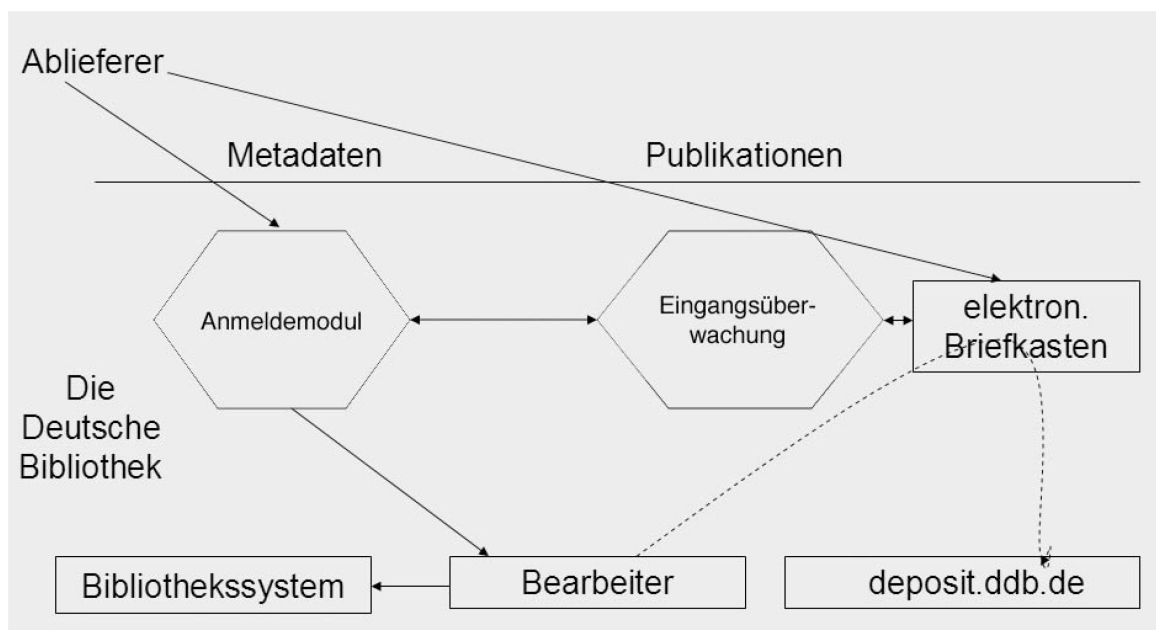
¹³ Website Open Archives Initiative: <<http://www.openarchives.org>>, 28. September 2003

¹⁴ Website Springer Link: <<http://www.springerlink.de>>, 28. September 2003

Die Datenlieferung erfolgte wegen der erheblichen Datenmenge auf physischen Trägern (CD-R, später DLT-Magnetbänder). Für jede zu archivierende elektronische Zeitschrift wurde ein Katalogisat im Bibliothekssystem Der Deutschen Bibliothek angefertigt. Die Archivierung erfolgt ebenfalls auf dem Archivserver, der Endnutterzugriff über das Katalogisat im Online-Publikumskatalog und die hierarchische Navigationsstruktur (Jahrgang-Heft-Artikel) der Publikation. Um dem Publikationstyp «elektronische Zeitschrift» gerecht zu werden, wurde ein Datenänderungsdienst implementiert, der in Zeitabständen für eine Synchronisation der Archivbestände sorgt. Das Verfahren steht zur Revision an, um zu einem produzentenneutralen Transferverfahren für elektronische Zeitschriftenartikel unabhängig von proprietären Eigenschaften der Ursprungssysteme entwickelt zu werden.

Freiwilliges Abgabeverfahren für Netzpublikationen

Seit Oktober 2001 bietet Die Deutsche Bibliothek ein Abgabeverfahren¹⁵ zur freiwilligen Ablieferung transferfähiger Netzpublikationen an.



Ein interaktives Formular ermöglicht die Übermittlung beschreibender und technischer Metadaten zu Einzelpublikationen oder auch zu elektronischen Zeitschriften. In Der Deutschen Bibliothek überwachen Prozesse (Anmel-

¹⁵ Abgabe von Netzpublikationen an Die Deutsche Bibliothek: <http://deposit.ddb.de/netzpub/web_abgabe_np_gesamt.htm>, 26. Januar 2004

demodul, Eingangsüberwachung) die einzelnen Aktivitäten und informieren die verantwortlichen Bearbeiter bei Vorlage der für die Katalogisierung und Archivierung notwendigen Daten und Informationen. Sobald Metadaten und die Publikationen in der Bibliothek eingetroffen sind, wird ein Katalogisat im Bibliothekssystem hergestellt und der Archivierungsvorgang ausgeführt. Die zu archivierende Kopie der Netzpublikation wurde zu diesem Zeitpunkt bereits in Verantwortung des Ablieferers in den elektronischen Briefkasten der Bibliothek transferiert, sodass die Bibliothek von wesentlichen Aufgaben der Konsistenzsicherung entlastet ist. Die Nutzungsfrequenz dieser Schnittstelle durch Ablieferer liegt leider unter den ursprünglichen Erwartungen. Als Ursache werden häufig der Umfang der von der Bibliothek geforderten beschreibenden Daten und der Aufwand zur Übermittlung genannt.

Abgabeverfahren für Newsletter

eMail-Newsletter haben bei fortlaufenden Publikationen häufig die konventionelle papiergebundene Vertriebsform vollständig abgelöst und es entsteht ein dringender Bedarf nach Erhaltung der Kontinuität in der Sammlung. Derzeit arbeiten wir an einem Verfahren, mit dem die Archivierung nach einer einmaligen Anmeldung des Newsletters vollautomatisch und kontrolliert durchgeführt wird. Der Anmelder hat lediglich die individuell seiner Publikation zugeteilte Mail-Adresse mit einem Exemplar zu versorgen.

Kooperation zwischen Publikations- und Archivservern

Ein wichtiges Element der künftigen Sammelstrategie Der Deutschen Bibliothek baut auf der engen Kooperation zwischen Dokument-/Publikations- und Archivservern. Unter Dokument-/Publikationsserver wird hier ein organisatorisches und technisches System verstanden, das Dokumente vor ihrer Veröffentlichung einem qualitätssichernden Prozess unterzieht (herkömmliche bzw. sich neu etablierende Formen des peer review). Aus technischer Sicht wird erwartet, dass ein Publikationsserver die oben formulierten Empfehlungen zur langzeitstabilen Gestaltung von Publikationen sinngemäss durchsetzt und die Autoren dabei unterstützt. Die Deutsche Bibliothek bereitet sich auf die Kooperation mit diesen Servern vor und richtet entsprechende Schnittstellen ein. Das bereits genannte OAI-Protokoll ermöglicht einen komfortablen Metadatentransfer. Die Voraussetzungen für eine daran anschliessende vergleichbar problemlose Übermittlung

der Volltexte aus der Umgebung der Produzenten in die Archivbibliotheken müssen in nächster Zukunft erst noch geschaffen werden.

Flächiges Einsammeln von Dateien - «web harvesting»

Die Deutsche Bibliothek hat sich in den vergangenen Jahren auf die Erarbeitung von Techniken für die selektive Sammlung von Netzpublikationen verbunden mit hohen Ansprüchen an Erschließungstiefe und Datenkonsistenz konzentriert. Andere Nationalbibliotheken haben währenddessen erhebliche Investitionen in Verfahren zur flächigen automatischen Einsammlung von Internet-Dateien (web harvesting) geleistet. Auf die umfangreichen Darstellungen in der einschlägigen Literatur¹⁶ wird deshalb hier nur verwiesen.

Die nationale und internationale Kooperation der Gedächtnisorganisationen macht es möglich, sich wechselseitig auf die Ergebnisse anderer zu beziehen und nahtlos auf den andernorts gemachten Erfahrungen aufzusetzen. So wird die zweite Säule unserer Sammelaktivitäten in Zukunft und nach Erteilung des entsprechenden gesetzlichen Auftrags darin bestehen, den deutschen Teil des Internets zu dokumentieren und dabei den Begriff der Netzpublikation mit dem bereits oben geschilderten Inhalt zu füllen. Auch dies wollen wir mit einem vorab definierten Anspruch an Konsistenz der Ergebnisse und der durch automatisierte Verfahren erreichbaren Erschließungstiefe durchführen. Auffällig ist, dass in den Nutzungsszenarien, die wir uns vorstellen, sehr häufig Historikern eine tragende Rolle zukommt. Bei den Bemühungen um die künftige Verfügbarkeit historischer Datenbestände des Internets und inzwischen verschollener Netzpublikationen führen wir den nach primären Quellen suchenden Historiker ins Feld, ohne ihn bislang tatsächlich gefragt zu haben, welche der mit Sicherheit nur unzureichend zu erfüllenden Bedürfnisse ihm die wichtigsten sind. Eine genügend weit vorausschauende Anforderungsdefinition durch unsere gegenwärtigen Nutzerinnen und Nutzer in Wissenschaft und Forschung wäre ein wichtiger Beitrag für den ökonomischen Einsatz unserer Mittel und eine an der zukünftig erwarteten Nutzung orientierte Festlegung der Prioritäten.

16 National Library of Australia (Hg.): Preserving Access to Digital Information (PADI) <<http://www.nla.gov.au/padi/>>, 28. September 2003

Leere Seite
Blank page
Page vide