

Infrastrukturen für die Archivierung digitaler Dokumente

Ein Tagungsbericht

Inka Tappenbeck

Mit der Entstehung digitaler Medien haben sich auch neue Formen der Produktion, Publikation und Distribution wissenschaftlicher Information entwickelt: Elektronische Sekundärveröffentlichungen, aber auch genuin elektronische Publikationen, Publishing-on-Demand sowie die Möglichkeit des Zugriffs auf Dokumente im Pay-per-View-Verfahren sind nur einige Beispiele für diese Veränderungen. Mit der steigenden Anzahl und Relevanz digitaler Ressourcen nimmt aber auch die Frage nach den Möglichkeiten und Bedingungen ihrer zuverlässigen Archivierung an Dringlichkeit zu. Dies betrifft sowohl die Sicherung der Datenspeicherung als auch den zukünftigen Zugriff auf die in ihnen enthaltenen Informationen. Die Faktoren, die einer einfachen Lösung dieser Aufgabe entgegenstehen, sind vielfältig: Der rasante Technologiewechsel im IT-Bereich erschwert den Zugriff auf ältere Datenformate und erfordert eine Emulation der ursprünglichen technischen Umgebung eines Dokuments bzw. die Migration der Daten selbst. Vor allem aber fehlen Standards für die Archivierung digitaler Ressourcen, die vorgeben, welche Informationen für die Langzeitarchivierung erforderlich sind, und die einen Rahmen für die Erstellung und Pflege dieser Informationen bereitstellen. Um sich einer Lösung dieser Probleme zu nähern, trafen sich Vertreter internationaler Institutionen aus dem Archiv-, Bibliotheks- und IT-Bereich vom 6. bis zum 8. Dezember 2000 in York anlässlich der von Cedars¹ ausgerichteten Konferenz „Preservation 2000“ und des dieser Konferenz vorausgehenden eintägigen Workshops „Information Infrastructures for Digital Preservation“.² Im folgenden soll ein Überblick über die Inhalte der im Rahmen beider Veranstaltungen gehaltenen Vorträge gegeben werden.

Im Anschluss an die Eröffnung des Workshops „Information Infrastructures for Digital Preservation“ durch *Robin Dale* (Vorsitzende der OCLC/RLG Working Group on Metadata for Digital Preservation³) stellte *Brian Lavoie* (OCLC⁴) den Arbeitsbericht der OCLC/RLG Working Group on Metadata for Digital Preser-

1 Cedars: CURL Exemplars for Digital ARchiveS <<http://www.curl.ac.uk/projects/cedars.html>>

2 <<http://www.ukoln.ac.uk/events/cedars-2000>>

3 <<http://www.rlg.ac.uk/pr/pr2000-oclc.html>>

4 Online Computer Library Center <<http://www.oclc.org>>

vation vor. Im Mittelpunkt stand das von der Gruppe erarbeitete „White Paper“, eine Zusammenfassung und Analyse der aktuellen internationalen Diskussion über die Verwendung von Metadaten für die Archivierung digitaler Dokumente. Ziel dieses Papiers ist es, funktionale Erfordernisse für ein auf möglichst breiter Ebene akzeptierbares Archivierungs-Metadatenmodell zu definieren. Als Grundlage eines solchen Modells wurde von der Arbeitsgruppe das vom Consultative Committee for Space Data Systems⁵ erarbeitete „Reference Model for an Open Archival Information System“ (OAIS)⁶ vorgeschlagen. Das OAIS-Modell stellt sowohl eine Terminologie als auch ein Strukturkonzept für Archivierungsmetadaten bereit und identifiziert zentrale für die Archivierung erforderliche Funktionen und Abläufe. Viele internationale mit Archivierungsfragen befasste Institutionen und Projekte orientieren sich bereits jetzt an diesem Modell, so etwa das Projekt Cedars⁷, die National Library of Australia⁸ und das EU-Projekt Nedlib (Networked European Deposit Library)⁹. In der sich an diesen Vortrag anschließenden Panel-Diskussion mit Vertretern der Harvard University¹⁰, des Cedars Projekts, der National Library of Australia¹¹ und des Nedlib Projekts¹² wurde deutlich, dass ein breiter Konsens hinsichtlich der Einschätzung der maßgeblichen Bedeutung des OAIS-Modells herrscht, seine praktische Adaption allerdings in den verschiedenen Institutionen und Projekten in durchaus unterschiedlicher Form geschieht. Diese Differenzen resultieren aus der Anpassung des OAIS-Modells an lokale Erfordernisse, Möglichkeiten und Absichten. In diesem Zusammenhang wurde intensiv diskutiert, wie das weithin geteilte Ziel der Interoperabilität von Archivierungsmetadaten mit lokalen Profilen verbunden werden kann.

Im Anschluss an diese Diskussion stellte *Inka Tappenbeck* (Niedersächsische Staats- und Universitätsbibliothek Göttingen¹³) den Metadatenentwurf des Projekts CARMEN AP 2/5¹⁴ vor. In diesem Arbeitspaket der Global Info-Sonderfördermaßnahme CARMEN (Content Analysis, Retrieval and Metadata:

5 <<http://www.ccsds.org>>

6 <<http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf>>

7 Cedars: Draft Specification: Cedars Preservation Metadata Elements <<http://www.leeds.ac.uk/cedars/MD-STR-5.pdf>>

8 NLA: Preservation Metadata for Digital Collections: Exposure Draft <<http://www.nla.gov.au/preserve/pmeta.html>>

9 Nedlib: Metadata for long term preservation <<http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>>

10 <<http://www.harvard.edu>>

11 <<http://www.nla.gov.au>>

12 <<http://www.kb.nl/nedlib>>

13 <<http://www.sub.uni-goettingen.de>>

14 <<http://harvest.sub.uni-goettingen.de/carmen>>

Effective Networking)¹⁵ geht es um die Entwicklung von Metadaten für die Rechteverwaltung und Archivierung und um die Erarbeitung eines Workflow für deren Erstellung. Auch dieses Projekt hat sich für eine leicht modifizierte Adaption des OAI-Metadatenkonzepts entschieden. Da die erforderliche Vielzahl und Komplexität dieser Metadaten jedoch von keiner einzelnen Institution bereitgestellt und gepflegt werden kann, hat das Projekt CARMEN AP 2/5 einen Workflow entwickelt, der zeigt, wie eine Zusammenarbeit zwischen Autoren, Verlegern, Bibliotheken und Archiven bei der kooperativen Erstellung und Pflege der erforderlichen Daten aussehen könnte. Dabei wurden auch Erfahrungen aus der Kooperation mit dem Springer Verlag, der sich zu einer solchen Kooperation im Rahmen des Projekts CARMEN AP 2/5 bereit erklärt hat, eingebracht.

Oya Y. Rieger (Cornell University Library¹⁶) präsentierte in ihrem Vortrag „Project Prism: Preservation Metadata Research“ die Langzeitstudie der Cornell University Library zur Archivierung digitaler Dokumente, wobei zugleich auch die mit den Dokumenten verbundenen Urheberrechte und die Rechte der Nutzer dieser Dokumente Beachtung finden. In diesem Archivierungs- und Sicherheitsservice wird eine technische Infrastruktur erarbeitet, die die Bereitstellung digitaler Materialien, die Sicherung ihrer Integrität und ihre Langzeitarchivierung in automatisierter Form ermöglicht.

Günter Mühlberger (Universität Innsbruck¹⁷) stellte in seinem Vortrag das Konzept des EU-Projekts „METAE: The Metadata Engine“¹⁸ vor. In diesem Projekt soll ein Verfahren entwickelt werden, das im Prozess der Digitalisierung gedruckter Bücher des 19. und 20. Jahrhunderts durch die Verwendung spezieller Software zur Layout-Analyse automatisch Metadaten zu den digitalisierten Dokumenten generiert. Dieser Ansatz beruht auf einer Analyse der typischen Struktur gedruckter Bücher dieses Zeitraums, der Anordnung und Gestalt von Titelseiten, Überschriften, Paragraphen, Fußnoten, Bildelementen und Seitenzahlen. Dieses Projekt will eine effiziente und qualitativ hochwertige Metadatenherstellung im Prozess der Digitalisierung selbst ermöglichen und so den intellektuellen Aufwand und die damit verbundenen Kosten minimieren. Die Grenzen dieses Verfahrens liegen daher naturgemäß dort, wo die Typik endet, etwa bei ungewöhnlichen Dokumentstrukturen. Die Reihe der Vortragenden dieses Workshops schloss mit *Margarete Byrnes* (National Library of Medici-

15 <<http://www.mathematik.uni-osnabrueck.de/projects/carmen>>

16 <<http://www.library.cornell.edu>>

17 <<http://www.uibk.ac.at>>

18 <<http://meta-e.uibk.ac.at>>

ne¹⁹), die in ihrem Beitrag „Categorising Commitment Levels“ die Ergebnisse einer Working Group der National Library of Medicine vorstellte.

Die sich an diesen Workshop anschließende zweitägige Konferenz „Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials“ wurde von der Direktorin des Cedars Projekts, *Clare Jenkins* (Imperial College of Science, Technology and Medicine²⁰), eröffnet. Den Einleitungsvortrag hielt *Lynne Brindley* (The British Library²¹), die die bisher im Bereich der Archivierung digitaler Ressourcen erreichten Fortschritte zusammenfasste und die diesbezügliche Entwicklung an der British Library darstellte. Für die Zukunft sah Brindley vor allem im Bereich der praktischen Umsetzung der in verschiedenen Projekten erarbeiteten Ergebnisse unmittelbaren Handlungsbedarf. Hierzu stellte sie den Entwurf eines Manifests²² zur Diskussion, das diese Umsetzung auf den Weg bringen soll. Wesentliche Aspekte dieses Manifests betreffen die Organisation der internationalen Zusammenarbeit im Bereich der Archivierung und die Positionierung des Themas „digitale Archivierung“ im öffentlichen Bewusstsein.

In der Session über „Models for Distributed Digital Archives“ stellte *Kelly Russell* (Consortium of University Research Libraries²³) die Erfahrungen des Projekts Cedars im Bereich der digitalen Archivierung dar. Dabei thematisierte sie vor allem die verschiedenen technologischen Strategien zur Archivierung digitaler Information: die Archivierung der ursprünglichen Hard- und Software, die Emulation der ursprünglichen technischen Umgebung eines Dokuments sowie die Migration der Daten, und diskutierte die je spezifischen Vor- und Nachteile der einzelnen Verfahren. Jedoch hängt der Erfolg der gewählten Strategie, so Russell, in erster Linie von den jeweiligen Bedingungen und Erfordernissen der archivierenden Institution ab, so dass es nicht sinnvoll ist, eine generelle Entscheidung für oder gegen eine bestimmte Archivierungsstrategie zu treffen.

Über einen interessanten dezentral konzipierten Ansatz zur Sicherung digitaler Ressourcen berichtete *Vicky Reich* (Stanford University Libraries²⁴). Unter dem Motto „Diffused Knowledge Immortalizes Itself“ (Sir James Mackintosh, 1765-1832) stellte sie die Initiative LOCKSS (Lots Of Copies Keep Stuff Safe)²⁵ vor, die eine Bereitstellung und Sicherung digitaler Publikationen auf dem Wege

19 <<http://www.nlm.nih.gov>>

20 <<http://www.ic.ac.uk>>

21 <<http://www.bl.uk>>

22 <<http://www.bl.uk/concord>>

23 <<http://www.curl.ac.uk>>

24 <<http://www-sul.stanford.edu>>

25 <<http://lockss.stanford.edu>>

ihrer Vervielfachung und Speicherung an verschiedenen Orten betreibt. Die Stärke eines solchen verteilten Systems liegt in seiner Redundanz, denn der Ausfall eines oder selbst mehrerer Sicherungsserver bedeutet noch keine Gefährdung der vorgehaltenen Information. Durch den automatischen kontinuierlichen Abgleich der auf den verschiedenen Servern gespeicherten Daten wird, so Reich, die Aktualität der Inhalte gewährleistet und Fehler beseitigt. Bisher haben sich weltweit zahlreiche Universitäten und andere wissenschaftliche Forschungseinrichtungen, aber auch einige internationale Verlage, als Testpartner für dieses Projekt zur Verfügung gestellt. Für weitere interessierte Institutionen ist eine Teilnahme am Test möglich.²⁶

In der Session „Perspectives of Managing National Digital Collections“ stellte *Helen Shenton* (The British Library) die Initiativen der British Library zur Archivierung digitaler und digitalisierter Informationsressourcen vor. Sie ging dabei vor allem auf die Veränderungen im Personaleinsatz ein, die diese neuen Aufgaben, etwa die Digitalisierung von Printmedien, für die mit ihnen befassten Institutionen mit sich bringen. Dabei wurde deutlich, dass diese Aufgaben nicht ohne einen zusätzlichen finanziellen Aufwand zu leisten sind, da vorhandenes Personal zunächst auf die neuen Aufgaben geschult bzw. neues Personal eingestellt werden muss. Zudem müssen neue Arbeitsabläufe, vor allem im Bereich der Digitalisierung, erst einmal etabliert werden. Dass dies alles nicht umsonst zu haben ist, war eine auf der Tagung von allen Beteiligten geteilte Erfahrung, die nun auch den Geldgebern der mit Archivierungsaufgaben betrauten Institutionen deutlich zu machen ist. Anschließend präsentierte *Lex Sijtsma* (Koninklijke Bibliotheek²⁷) unter dem Titel „A Model for Digital Deposits“ einige der Ergebnisse des Europa-Projekts Nedlib. In Form einer Multimediovorführung wurden die einzelnen Arbeitsabläufe der digitalen Archivierung, von der Aufnahme eines Dokuments über seine Archivierung bis zu seiner Bereitstellung, anschaulich dargestellt. Daran schloss sich der Vortrag von *Colin Webb* (National Library of Australia) über das Projekt PANDORA²⁸ an. In diesem Projekt der australischen Nationalbibliothek geht es um die Archivierung australischer Online-Publikationen. Das seit 1997 zu diesem Zweck aufgebaute Archiv umfasst bereits eine Auswahl von über 1.300 Dokumenten, die auch in die australische Nationalbibliographie aufgenommen wurden. Im Kontext seines Vortrags wies Webb auf die von der australischen Nationalbibliothek erstellte und gepflegte Internet-Seite PADI (Preservation Access to

26 Kontaktadresse: <vreich@stanford.edu>

27 <<http://www.kb.nl>>

28 <<http://pandora.nla.gov.au/pandora>>

Digital Information)²⁹ hin, die eine ausgezeichnete Sammlung von Informationsmaterialien zum Thema „digitale Archivierung“ enthält.

Margaret Jones (The Arts and Humanities Data Service³⁰) eröffnete die Session „Practicalities of Digital Preservation“. In ihrem Vortrag stellte sie das im Rahmen des gleichnamigen Projekts erarbeitete „Workbook for the Preservation Management of Digital Materials“³¹ vor. Dieses Buch stellt einen Leitfaden für die Planung von Digitalisierungsvorhaben dar und wird den Autoren zufolge in Zukunft zu einem Trainingsinstrument für mit Archivierungsaufgaben befasste Institutionen weiterentwickelt. Michele Cloonan und Shelby Sanett (University of California³²) präsentierten in ihrem gemeinsamen Vortrag das Projekt InterPARES³³, in dem eine exemplarische Kostenkalkulation für digitale Archivierungsvorhaben entwickelt wird. Die wirklichen Schwierigkeiten liegen dabei, so die Autorinnen, nicht im Bereich der Technik, sondern im ökonomischen und organisatorischen Bereich. Im Anschluss daran sprach Ellis Weinberger (Cedars) über „Intellectual Property Rights for Digital Preservation“. Ausgehend von der anglo-amerikanischen Rechtslage erläuterte er die spezifischen juristischen Aspekte, die insbesondere im Bereich der Urheberrechte bei der Archivierung digitaler Ressourcen zu beachten sind. Des Weiteren legte Weinberger eine „To-do-Liste“ für die Planung digitaler Archivierungsvorhaben vor, die angibt, welche rechtlichen Aspekte zu klären sind, bevor ein Dokument in den Archivierungsprozess aufgenommen werden kann.

Die darauf folgende Session befasste sich mit „Authenticity and Authentication for Digital Preservation“. Den Einleitungsvortrag sprach Nancy Brodie (National Library of Canada³⁴). Unter dem Titel „Authenticity, Preservation and Access in Digital Collections“ gab sie einen Überblick über die Archivierungsaktivitäten der kanadischen Nationalbibliothek, die bereits seit 1994 elektronische Ressourcen in ihre Sammlung aufnimmt. Darüber hinaus thematisierte sie die verschiedenen Dimensionen der Authentizität digitaler Dokumente (Originalität, Integrität, Zitierfähigkeit) und die sich daraus ergebenden Authentifizierungserfordernisse zum Schutz und Erhalt der Dokumente. In diesem Zusammenhang sprach Brodie auch die Risiken an, die sich aus dem öffentlichen Zugriff auf ungeschützte digitale Dokumente ergeben und stellte verschiedene Möglichkeiten der Authentifizierung vor (Verschlüsselung, digitale Signaturen, digitale Zertifikate). Für einige mit der Authentifizierung digitaler

29 <<http://www.nla.gov.au/padi>>

30 <<http://ahds.ac.uk>>

31 <<http://www.jisc.ac.uk/dner/preservation/workbook>>

32 <<http://www.ucla.edu>>

33 <<http://www.interpares.org>>

34 <<http://www.nlc-bnc.ca>>

Dokumente unweigerlich verbundene Probleme, wie etwa die in digitalen Kontexten häufig schwer zu beantwortende Frage nach dem „Originaldokument“, wurden jedoch auch in diesem Vortrag keine abschließenden Lösungen angeboten. Im Anschluss erläuterte *Kevin Ashley* (University of London Computing Centre³⁵) die verschiedenen Authentizitätsebenen, die ins Spiel kommen, wenn digitale Dokumente durch Informationsanbieter an Nutzer geliefert werden: die Authentifizierung der Nutzer zu Zwecken der Belieferung und Abrechnung, die Authentizität der Dokumente (s.o.) sowie die Authentizität der Anbieter (Vertrauenswürdigkeit, Legalität etc.). Dieser Vortrag wurde ergänzt durch den Beitrag *George Barnums* (U.S. Government Printing Office, Federal Depository Library Program³⁶) über die Initiativen der U.S. Regierung zum Schutz der Authentizität offizieller digitaler Regierungsinformationen. Dabei sprach er auch die ungewisse Zukunft des Federal Depository Library Program an.

Die vorletzte Session leitete *Robin Dale* mit einem Beitrag zur internationalen Zusammenarbeit im Bereich der Entwicklung von Archivierungsmetadaten ein. Dale resümierte die Arbeit der RLG/OCLC Working Group on Metadata for Digital Preservation und betonte die Bedeutung des OAIS-Modells für die bisherigen und weiteren Arbeiten in diesem Feld. Sie hob hervor, dass zukünftige Aktivitäten insbesondere auf eine Klärung der Fragen abzielen müssten, welche Quantität an Metadaten für die erfolgreiche Archivierung erforderlich seien, welcher Anteil dieser Daten automatisch erstellt werden könne, welche Institutionen für die Bereitstellung derjenigen Metadaten verantwortlich seien, die nicht automatisch erstellt werden können und welcher Personalaufwand damit impliziert sei. Im Anschluss daran stellte *Neil Beagrie* (Joint Informations Systems Committee³⁷) die bisherigen und zukünftigen Initiativen des JISC im Bereich der digitalen Archivierung³⁸ vor. Auch hier wurde deutlich, dass die Herausforderungen für die Zukunft vor allem im Bereich der praktischen Umsetzung der bisher in verschiedenen Projekten erzielten Ergebnisse liegen.

Den Abschluss der Konferenz bildete der Vortrag des Präsidenten der Research Libraries Group (RLG)³⁹, *Jim Michalko*, über den Stand der internationalen Entwicklungen im Bereich der digitalen Archivierung. Dabei nannte er auch diejenigen Gebiete, auf die sich zukünftige Anstrengungen konzentrieren sollten, etwa die Formulierung von Kriterien für die Auswahl der

35 <<http://www.ulcc.ac.uk>>

36 <<http://www.gpo.gov/fdlpdesktop>>

37 <<http://www.jisc.ac.uk>>

38 <<http://www.jisc.ac.uk/dner/preservation>>

39 <<http://www.rlg.ac.uk>>

zu archivierenden Objekte und die Erarbeitung von Organisationsmodellen für den Archivierungsprozess sowie für die – unter Umständen auch kommerzielle – Bereitstellung der archivierten Dokumente.

Die weltweite Aktualität und Relevanz der Frage nach Standards und Infrastrukturen für die digitale Archivierung zeigte sich nicht nur an der Präsenz von über 150 Tagungsteilnehmern zahlreicher Nationen, sondern auch an dem Engagement, mit dem die vorgestellten Beiträge diskutiert wurden. Dabei fiel jedoch auf, dass die meisten Bemühungen und Aktivitäten aus dem anglo-amerikanischen Bereich kommen; in Deutschland sind dagegen noch relativ wenig Bestrebungen im Bereich der digitalen Archivierung auszumachen. Dies spiegelte sich auch in der Zusammensetzung der Teilnehmer wieder: Obwohl die Tagung in Europa stattfand, waren mehr australische als deutsche Teilnehmer anwesend. Für die Zukunft wäre eine stärkere Partizipation an der internationalen Entwicklung sicher wünschenswert, denn die erfolgreiche Aufbewahrung und Dokumentation unseres kulturellen Erbes, das in immer stärkerem Maße in digitaler Form vorliegt, ist nur in Form einer globalen Zusammenarbeit zu leisten.

