# Zur Langzeitverfügbarkeit digitaler Ressourcen

Susanne Dobratz
susanne.dobratz@cms.hu-berlin.de

## Bereitstellung und Langzeitverfügbarkeit

Neben der Unterstützung der Open-Access-Bereitstellung wissenschaftlicher Publikationen ist es für Betreiber von Institutional Repositories essentiell, sich auch um die Langzeitverfügbarkeit der digitalen Publikationen zu bemühen. Dies kann entweder dadurch geschehen, dass das Institutional Repository zu einem vertrauenswürdigen digitalen Langzeitarchiv ausgebaut wird oder aber gezielt Kooperationen mit Archivierungsinstitutionen eingegangen werden.

Die Bereitstellung digitaler Publikationen kann relativ problemlos auf einem Dokumentenserver geschehen, indem der Autor selbst seine Dokumente und einige zugehörige Metadaten auf den Server aufspielt. Es ist jedoch ungleich schwieriger, Garantien über die zukünftige Lesbarkeit der digitalen Dokumente abzugeben. Derartige Garantien hängen von organisatorischen und technischen Belangen des jeweiligen Dokumentenservers ab:

- vom Dateiformat, in dem die Publikationen gespeichert werden,
- von der Qualität und dem Umfang der zu den Dokumenten gespeicherten Metadaten,
- von der Qualität der Sicherung der Authentizität und Integrität der gespeicherten Dokumente und
- von der finanziellen Planung einer Universität für das digitale Langzeitarchiv.

Der Aufbau eines eigenen vertrauenswürdigen Archivs entsprechend dem 2002 von der Research Libraries Group publizierten Anforderungsprofil [1] und den im OAIS-Standard[1] definierten Eigenschaften [2] ist sehr aufwendig und kann nicht von jedem Betreiber eines Dokumentenservers geleistet werden. Daher entschließen sich Betreiber von Institutional Repositories häufig dazu, die Aufgabe der Langzeitverfügbarkeit abzugeben. Ein Beispiel ist die Universität Southampton, die sich zwar dafür verantwortlich fühlt, die Dokumente der dort betriebenen *eprints*-Server einer Langzeitarchivierung zuzuführen, dies aber über eine Kooperation mit der British Library realisieren möchte. Zu diesem Zweck arbeitet sie im Projekt *PRESERV*[2] daran, entsprechende Schnittstellen der *eprints*-Server zum vertrauenswürdigen digitalen Archiv der British Library zu implementieren, vgl. [3].

## Langzeitverfügbarkeit auf dem edoc-Server

Die Problematik der Langzeitarchivierung wurde beim Aufbau des Dokumentenservers der HU von vornherein als ein wichtiges Thema angesehen, und die dazu erforderlichen Maßnahmen wurden und werden schrittweise umgesetzt.

> In diesem Zusammenhang bedeutet »Langzeit« bzw. »langfristig« nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten technologischen Wandel bewältigen können. [4]

### Kooperation mit der DDB

Langfristig soll der edoc-Server zu einem vertrauenswürdigen Archiv ausgebaut werden. Bis dies realisiert ist, ist die HU Berlin eine Kooperation mit Der Deutschen Bibliothek (DDB) eingegangen. Die über den edoc-Server publizierten digitalen Dissertationen und Habilitationsschriften werden in das DDB-eigene digitale Langzeitarchiv übernommen.

### Betrieb des edoc-Servers

Einige Eigenschaften eines vertrauenswürdigen digitalen Langzeitarchivs besitzt der edoc-Server aber bereits schon heute. Dazu gehört an erster Stelle die Garantie eines zuverlässigen und organisatorisch gesicherten Betriebes. Diese Garantie hat die HU Berlin dadurch gegeben, dass sie den edoc-Server als festen Bestandteil in das Dienstleistungsangebot von CMS und UB eingegliedert und den Betrieb durch zugehöriges Personal abgesichert hat.

### Leitlinien

An zweiter Stelle wurden für den edoc-Server Leitlinien definiert, die beschreiben, welche Publikationen unter welchen Bedingungen über den edoc-Server veröffentlicht werden und unter welchen Bedingungen Garantien für eine zukünftige Lesbarkeit und Benutzbarkeit der Dokumente gegeben werden.

### Sicherheit

Drittens werden der Zugang zum und der Zugriff auf den edoc-Server über diverse organisatorische und technische Maßnahmen kontrolliert und gesichert. So steht der Server selbst im Rechnerraum des CMS und ist damit in die Sicherheits-Infrastruktur des CMS eingebunden. Durch den Einsatz von digitalen Signaturen wird sichergestellt, dass die Authentizität und Integrität der Dokumente gewahrt wird und jede nachträgliche Änderung der Autorenschaft, des Veröffentlichungsdatums oder gar des Inhaltes bemerkt wird. Des Weiteren wird durch tägliche Dateisicherungen über das zentrale Backupsystem des

1 Open Archival Information System
2 Preservation Eprint Services, www.jisc.ac.uk/index.cfm?name=project_preserv

CMS dem Verlust von Dokumenten vorgebeugt.

### Dokumentation

Ein vertrauenswürdiges digitales Langzeitarchiv bedient sich transparenter Technologien und Methoden zur Speicherung, Migration und Bereitstellung von Dokumenten und Metadaten. Daher arbeitet die AG Elektronisches Publizieren derzeit im Rahmen des Projektes *reUSE* daran, eine umfassende Dokumentation des edoc-Servers mit sämtlichen angewendeten Verfahren zu erarbeiten.

### Dokumentformate

Die Problematik der Langzeitarchivierung in Abhängigkeit vom verwendeten Dateiformat stellt einen weiteren wichtigen Arbeitsbereich der AG Elektronisches Publizieren dar. Für Publikationen, die mit unserer Unterstützung aufbereitet und veröffentlicht werden, zielen wir auf ein XML-basiertes Speicherformat ab, z. B. in Form von xDiML-Dokumenten. Bei den von den Autoren über die Open Access-Schnittstelle selbst eingereichten Publikationen ist es das Ziel, diese in einem Format zu archivieren, welches bereits Eigenschaften besitzt, die eine turnusmäßige Migration der Dokumente in jeweils aktuell lesbare Formen erlauben. Dafür kommen beispielsweise XHTML sowie PDF-X oder PDF-A in Betracht.

### Metadatenmanagement

Ein Manko für ein langfristiges Management der im digitalen Archiv befindlichen Dokumente bilden zurzeit noch die Metadaten. Hier sind bisher vorrangig bibliographische und organisatorische Metadaten erfasst. Notwendig sind die Erarbeitung von Metadaten zur Erfassung technischer Parameter und der Einsatz von Tools, um diese Art von Metadaten automatisch generieren zu können. Standards, die an dieser Stelle zum Einsatz kommen könnten, wären das Metadatenmodell der PREMIS-Abeitsgruppe[3], das METS-Metadatenmodell[4], LMER – das Metadatenmodell der DDB[5] sowie XADES[6], ein vom W3C herausgegebener Standard zur Langzeitarchivierung digitaler Signaturen. Eine Software, die genutzt werden könnte, um technische Metadaten automatisch aus Dokumenten auszulesen, wäre das von der Harvard-Universität entwickelte System JHOVE[7]. Mit den Projekten *reUSE* und *Sun Center Of Excellence for Trusted Digital Repositories* unternimmt die AG Elektronisches Publizieren erste Schritte zur Aufarbeitung dieser Problematik.

## nestor

Die AG Elektronisches Publizieren ist Mitinitiator und Projektpartner in *nestor*[8], dem »Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland«. Dieses wurde ins Leben gerufen,

- um diejenigen zusammenzubringen, die sich hierzulande mit dem Problem der dauerhaften Erhaltung digitaler Objekte beschäftigen,
- um eine Informations- und Kommunikationsplattform als zentrale Anlaufstelle für alle Fragen zur digitalen Langzeitarchivierung zu schaffen und
- um einen nach außen sichtbaren Fokus in Deutschland zu bilden, der Kooperationen vermittelt und als Einstiegspunkt für internationale Allianzen zur Verfügung steht.

Das langfristige Ziel von nestor ist der nachhaltige Aufbau einer kooperativen Infrastruktur, in der vielfältige Fachkompetenzen bei der Problemlösung zusammenwirken. Zum aktuellen Stand des Projektes siehe [5].

Dabei beteiligt sich die AG Elektronisches Publizieren insbesondere daran, für die Bundesrepublik Deutschland Kriterien vertrauenswürdiger digitaler Langzeitarchive zu formulieren und ein Zertifizierungsverfahren zu erarbeiten. Dabei sollen nicht nur Dokumentenserver betrachtet werden sondern auch digitale Langzeitarchive, die von Archiven, Museen, Datenzentren, Verlagen und in Firmen betrieben werden.

## Literatur


[1] RLG/OCLC Working Group on Digital Archive Attributes: *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA, RLG, 2002, www.rlg.org/en/pdfs/repositories.pdf (28.04.2005).

[2] Consultative Committee for Space Data Systems (CCSDS): *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, BLUE BOOK, 2002, www.ccsds.org/documents/650x0b1.pdf (28.04.2005).

[3] Hitchcock, S.: *Preservation metadata for IRs. Vortrag auf dem DCC Workshop on Long-term Curation within Digital Repositories, 06.07.2005*, www.dcc.ac.uk

[4] Schwens, U., Liegmann, H.: Die digitale Welt – eine ständige Herausforderung. In: Kuhlen, R., Seeger, T., Strauch, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. Bd. 1 Preprint, 5. völlig neu gefaßte Ausgabe, München, Saur, 2004, www.langzeitarchivierung.de/downloads/digitalewelt.pdf

[5] Dobratz, S., Neuroth, H., Schoger, A., Strathmann, S.: nestor – Entwicklungsstand des Kompetenznetzwerkes zur Langzeitarchivierung digitaler Ressourcen in Deutschland. *Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) Themenheft Langzeiterhaltung digitaler Informationen – Von Projekten zu Konzepten*, 2005 (im Druck).


3 PREservation Metadata: Implementation Strategies, www.oclc.org/research/projects/pmwg
4 Metadata Encoding and Transmission Standard, www.loc.gov/standards/mets
5 Langzeitarchivierungsmetadaten für Elektronische Ressourcen, www.ddb.de/standards/pdf/lmer1.pdf
6 XML Advanced Electronic Signature, www.w3.org/TR/XAdES
7 JSTOR/Harvard Object Validation Environment, hul.harvard.edu/jhove
8 Network of Expertise in Long-Term Storage of Digital Resources, www.langzeitarchivierung.de