



Thesen zur Langzeitarchivierung digitaler Dokumente in Deutschland

Susanne Dobratz – HU Berlin

Dr. Inka Tappenbeck – SUB Göttingen

Unter Mitarbeit von

Hans Liegmann – DDB Frankfurt/Main

12. April 2002, Augsburg, 92. Deutscher Bibliothekartag

Zur Entstehung dieses Beitrags

- Digital Library Forum
 - <http://www.dl-forum.de>
- Arbeitskreis „Infrastrukturen für Digitale Bibliotheken
- Ergebnispapier (21.2.2001/ 2.2.2002)
 - Rahmenkonzept für DL-Dienste
 - AG Metadaten
 - AG Langzeitarchivierung



Ergebnispapier DL-Forum

Dieses Papier skizziert den Entwurf eines Rahmenkonzeptes für Digital Library-Dienste, so wie ihn die Beteiligten als sinnvoll erachten. Damit dokumentieren die Beteiligten ihre Bereitschaft, an der Gestaltung der Digitalen Bibliotheken aktiv mitzuwirken und hierbei eine Strategie zu verfolgen, die über die Grenzen der einzelnen Institutionen hinausgeht. Zugleich wird gesehen, dass auch dieser Versuch nicht alleine steht.

Ergebnispapier DL-Forum 2

Mit diesem Papier soll keineswegs ein mit anderen Versuchen konkurrierendes Kooperationsmodell vorgeschlagen werden. Es soll vielmehr aus Sicht der Beteiligten formuliert werden, welche Problemfelder vorrangig angegangen werden. Es soll vielmehr aus Sicht der Beteiligten formuliert werden, wie eine Verbesserung der elektronischen Fachinformation für den Nutzer erreicht werden kann, welche Problemfelder vorrangig angegangen werden sollten und wie die Prozesse zur Umsetzung dieser Ziele moderiert werden können.

Die Herausforderung

- Steigende Anzahl, Heterogenität und wissenschaftliche Relevanz digitaler Publikationen erfordert zuverlässige Archivierungsstrategien:
 - Standards für die :
 - dauerhafte Datenspeicherung (Trägermedium)
 - Sicherung des künftigen Zugriffs (Datenformate)
 - dauerhafte Nutzbarkeit (Erschließung)



Problem und Aufgaben

- Es fehlen **einheitliche Standards** als Rahmen für die Bewältigung der Zukunftsaufgabe
- Ziel: **Anschlußfähigkeit** der deutschen Entwicklung an die aktuellen internationalen Standardisierungsprozesse

Definition „Archivierung“

- Unter „Archivierung“ bzw. „Langzeitarchivierung“ soll dabei die erfolgreiche Gewährleistung der Langzeitverfügbarkeit einer Ressource verstanden werden.
- Ein wichtiges Ziel ist dabei die Sicherung der Authentizität (Vertrauenswürdigkeit) des archivierten Dokuments.

Authentizität: Anforderungen

- Idealerweise: **volle Originalität** des zu archivierenden Dokuments
- Realiter: „**relative Authentizität**“ in Abhängigkeit von der Beschaffenheit des zu archivierenden Dokuments und der gegebenen technischen Möglichkeiten

Authentizität: Kriterien

- Welche **Eigenschaften** einer Klasse von Dokumenten sind **relevant** für deren **Authentizität**?
- **Wie wichtig** sind diese **Eigenschaften** für die zukünftige Präsentation des Dokuments?
- Welche **Erhaltungsstrategie** ist den prioritär zu erhaltenden Eigenschaften dieses Typs von Dokument angemessen?
- Bsp.: Referenzierung externer Quellen

Archivierungsplanung

- Bestandsaufnahme, Analyse und Auswertung der internationalen Entwicklungen in Bezug auf die deutsche Situation.
- Entwicklung von **Norm-Standards** in Übereinstimmung mit internationalen Archivierungsinitiativen.
- Aufbau einer **dezentralen** und **kooperativen Archivierungsinfrastruktur** in Deutschland.

Internationale Archivierungsinitiativen

- Cedars
(CURL Exemplars for Digital Archives)
 - <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>
- NedLib
(Networked European Deposit Library)
 - <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>
- OCLC/RLG WG on Metadata for Digital Preservation
 - <http://www.rlg.ac.uk/pr/pr2000-oclc.html>
- National Library of Australia
 - <http://www.nla.gov.au/preserve/pmeta.html>

Standardisierung I: Infrastruktur

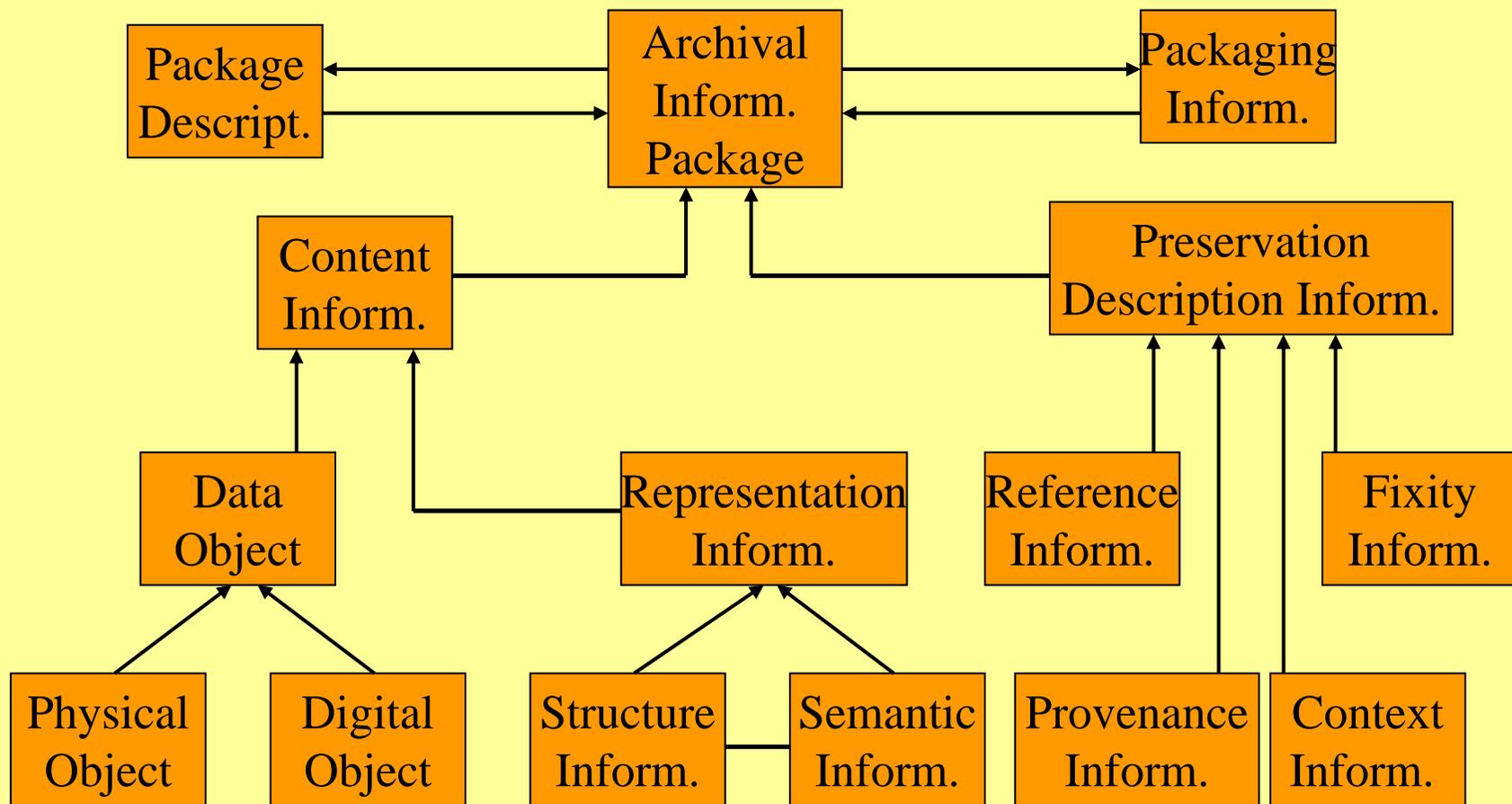
- Reference Model for an Open Archival Information System (OAIS). Red Book, Version Juli 2001:
 - www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf
- **Funktionen** und **Abläufe** eines Archivsystems
- Terminologie und **Strukturkonzept** für Metadaten
- **Neutralität** gegenüber Archivierungstechniken
- dezentrale Implementierung möglich



OAIS: Logisches Modell

www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf

Dobratz/Tappenbeck: Langzeitarchivierung



Standardisierung II: Dokumentstandards

- Nutzung von **präsentationsunabhängigen Auszeichnungssprachen** um Kosten für Emulations- oder Migrationsmodelle zu senken und inhaltliche Dokumentstrukturierungen zu bewahren
 - Verwendung von Norm-Standards statt de-facto Firmenstandards
 - SGML (ISO 8879) seit 1986 !!!!!
 - XML (extensible Markup Language)
 - XLS (eXtensible Style Language)
 - XLink (eXtensible Linking Language)



SGML/XML

- Standard-Dokumenttypdefinitionen (DTD)
 - Mathematische Formeln: MathML
 - Chemische Formeln: CML
 - Vektorgrafiken: SVG
 - ISO DocBook DTD
 - Open e-book Standard
 - Dissertationen Online (DiML-Entwicklungen)
 - u.s.w.
 - Siehe WWW-Repositories:
 - Z.B. <http://www.schema.net>



Standardisierung III: Metadaten

- Dublin Core Metadata Element Set (DCMES).
 - Version 1.1, Juli 1999
 - dublincore.org/documents/dces/
- einheitliche Beschreibung verschiedener Objekttypen
- Retrieval nach gemeinsamen Prinzipien
- Interdisziplinarität
- Internationalität



Dublin Core Metadata Element Set

dublincore.org/documents/dces/

dc.title

dc.creator

dc.subject

dc.description

dc.publisher

dc.contributor

dc.date

dc.type

dc.format

dc.identifier

dc.source

dc.language

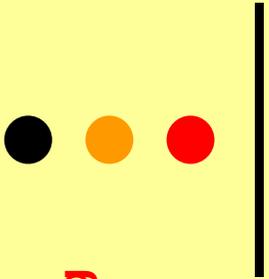
dc.relation

dc.coverage

dc.rights

Problemstellungen und Handlungsbedarf

1. Archivierungsfreundliche Dokumentformate
 - Ausbildung von Autoren
 - Standardisierte Werkzeuge / Schnittstellen
 - Unterstützung dynamischer Komponenten
 - Unterstützung von Datenmodellen, die medienneutrales Publizieren ermöglichen



Problemstellungen und Handlungsbedarf

2. Transferstandards und –protokolle

- Standardisierung eines allgemein akzeptierten Transferformats
- Transferoptimierung durch geschlossene Transfercontainer, die alle relevanten Daten und Metadaten des Dokuments enthalten
- Automatische Übermittlung der Dokumente und ihrer Metadaten

Problemstellungen und Handlungsbedarf

3. Authentizität und Integrität

- Aufbau von Public Key Infrastrukturen (PKI) in Universitäten, Bibliotheken und Archiven
- Unterstützung der Entwicklung und Nutzung digitaler Signaturen und Zeitstempel
- Förderung von Lösungen zur Sicherung der Authentizität dynamischer Dokumente

Problemstellungen und Handlungsbedarf

4. Technische Archivierungskonzepte

- „technology watch“: Marktbeobachtung
- Entwicklung von Qualitätskriterien und Zertifizierungsverfahren für Archivsysteme
- technische Verfahren für die automatisierte Migration von Massendaten
- Metadaten zur Steuerung von Konversionsprozessen
- Forschungsförderung für Emulationsverfahren

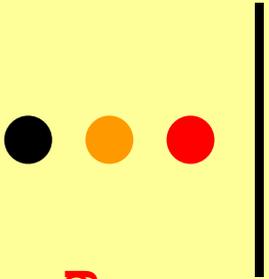
Problemstellungen und Handlungsbedarf

5. Metadatenmanagement

- Entwicklung von Archivierungsmetadaten als Erweiterung des Dublin Core Metadata Element Sets
- Erarbeitung eines gemeinsamen Standards für die Implementierung von Metadaten in XML/RDF
- Entwicklung eines Workflows zur dezentralen Erstellung und Pflege von Metadaten
- Praktische Tests in Zusammenarbeit aller an der „information chain“ Beteiligten

Problemstellungen und Handlungsbedarf

6. Funktionale und organisatorische Erfordernisse
 - Ausgang von geltenden Zuständigkeiten bei der Archivierung analoger Publikationen
 - gemeinsame Abstimmung der inhaltlichen Zuständigkeiten, sicherheitstechnisch erforderlichen Redundanzen und der technischen Instrumentarien (OAIS)



Problemstellungen und Handlungsbedarf

...6.

- Aufbau einer Koordinationsstruktur ähnlich der Digital Preservation Coalition in Großbritannien
- Förderung von archivierungsfreundlichen Publikationsmodellen und Geschäftsgängen

Problemstellungen und Handlungsbedarf

7. Rechtliche Fragen

- Entwicklung von Modellen zur aktiven und passiven Rechteverwaltung (DRM)
- Entwicklung semantischer und technischer Standards zur Beschreibung der rechtlichen Eigenschaften eines Dokuments und zur Abwicklung von Transaktionen
- Schaffung einer rechtlichen Basis für alle Fragen der digitalen Archivierung



Schlussfolgerungen I

- Stärkung der öffentlichen Bewusstseinsbildung für die Relevanz der Archivierung digitaler Informationsressourcen
- Analyse der internationalen Entwicklungen
- Aktive Partizipation an den aktuellen Initiativen
- Koordination der bereits in Deutschland begonnenen Archivierungsaktivitäten
- Nutzung von Standards, Vermeidung von Doppelentwicklungen



Schlussfolgerungen II

- Ausweitung praktischer Tests von Konzepten für die verteilte Archivierung digitaler Dokumente
- Zusammenführung und Qualifizierung aller an der Produktionskette Beteiligten, z.B. in Form von Projekten
- Einrichtung einer dezentralen Koordinationsstruktur mit klarer Verantwortungsteilung

Die ersten Schritte zur Realisierung

- Workshop: LZA digitaler Dokumente in Deutschland unter Beteiligung der DDB, der SUB, der BSB und der HU
- Initialzündung zur Erarbeitung eines gesamtdeutschen Konzepts
- Etablierung einer Koordinationsstruktur für die deutsche Situation
- Entwicklung von Richtlinien und Empfehlungen, auch für die Förderpolitik des BMBF



Fragen?

Dobratz/Tappenbeck: Langzeitarchivierung

**Vielen Dank für Ihre
Aufmerksamkeit!**

Weitere Informationen:

- susanne.dobratz@rz.hu-berlin.de
- tappenbeck@mail.sub.uni-goettingen.de

Siehe auch:

Zeitschrift für Bibliothekswesen und
Bibliographie 6/2001