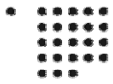


# Einführung in die Langzeitarchivierung digitaler Objekte

November 2010

FH Köln

Institut für Informationswissenschaft



Fachhochschule Köln  
Cologne University of Applied Sciences

Institut für Informationswissenschaft  
Institute of Information Science

Prof. Dr. Achim Oßwald  
achim.osswald@fh-koeln.de

**nestor**  
www.langzeitarchivierung.de



FH Köln

## Problemstellungen (1)

### ↳ Datensicherung im privaten Bereich



Wie lange werden wir diese  
Medien nutzen können?



Quelle der Grafiken: Wikipedia, s. Bildnachweise



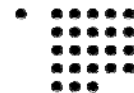
# Problemstellungen (2)

## ↪ Ursachen für Datenverlust:

- 3 % Naturkatastrophen
- 7 % Viren (6%)
- 14 % Softwarefehler (13%)
- 32 % menschliches Versagen (29%)
- 44 % Hardware- und Systemfehler (40%)

Angaben von data recovery company Ontrack (1996), nach R. Scheffel  
Quelle: Ross, Seamus, Gow, Ann: Digital Archaeology: Rescuing Neglected and Damaged Data Resources; JISC/NPO Studie 2/1999

In Klammer: Werte nach Smith, David M: The Cost of Lost Data.  
In: Graziado Business Report, Issue 3, 2003;  
vgl. <http://gbr.pepperdine.edu/033/dataloss.html> (22.11.10)



## Die Zielsetzung

heute



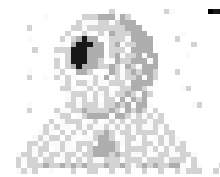
zur Nutzung nötig:  
Hardware

Betriebssystemsoftware

Anwendungssoftware

Daten

zukünftig

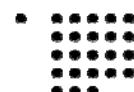


zur Nutzung nötig:  
Hardwarespezifische  
Informationen

Betriebssystemspezifische  
Informationen

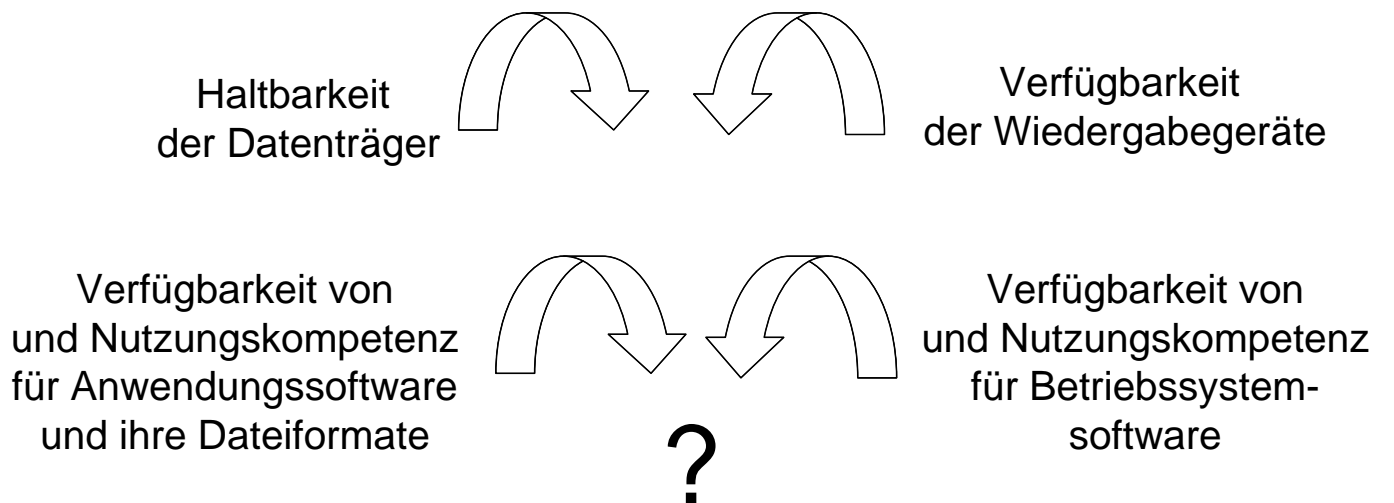
Anwendungssoftwarespezifische  
Informationen

Daten



# Problemstellungen: Überblick (4)

---



im Hinblick auf Langzeitverfügbarkeit  
Probleme bei Kompatibilität / Nutzbarkeit / Authentizität

---

## Überblick zu dieser Präsentation

---

### Aktuelle Fragen der digitalen Langzeitarchivierung (LZA) und Langzeitverfügbarkeit (LZV)

- ↪ Was ist eigentlich digitale LZA digitaler Objekte?
- ↪ Was soll gesichert werden?
- ↪ Wer soll die digitale LZA übernehmen?
- ↪ Wie soll gesichert werden?
- ↪ Welche Strategien und Verfahren zur digitalen LZA kommen zum Einsatz?
- ↪ Was wird / könnte digitale LZA kosten?
- ↪ Was läuft derzeit bzgl. digitaler LZA?
- ↪ Ausgewählte offene Fragen
- ↪ Weiterführende Quellen und Hinweise

# Was ist eigentlich digitale Langzeitarchivierung digitaler Objekte?

---

↪ **Langzeitarchivierung** digitaler Objekte umfasst alle Maßnahmen, die dazu dienen, digitale Objekte für die Nachwelt **dauerhaft** zu erhalten. Der Begriff korrespondiert mit jenem der **Langzeitverfügbarkeit**, die die dauerhafte Benutzbarkeit mehr in den Vordergrund stellt.

„**Langzeit**‘ ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche, nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten; Veränderungen, die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen können.“

**Zitiert nach:** Liegmann/Neuroth 2010, (aufgerufen: 22.11.10)  
[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_390.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_390.pdf)



## Was soll gesichert werden? (1)

---

### **Das digitale Erbe unserer Gesellschaft ...**

- „Das digitale Erbe besteht aus einzigartigen Quellen menschlichen Wissens und menschlicher Ausdrucksweisen. Es umfasst Quellen aus Kultur, Bildung, Wissenschaft und Verwaltung ebenso wie technische, rechtliche, medizinische und andere Arten von Informationen, die digital erstellt oder von existierenden analogen Datenträgern in digitale Form konvertiert wurden.“

*Zitiert nach: Art. 1 der UNESCO **Charta zur Bewahrung des digitalen Kulturerbes** (32. UNESCO-Generalkonferenz; 17. Oktober 2003);  
<http://www.unesco.de/444.html?&L=0> (aufgerufen: 22.11.10)*

- Digitale Materialien, d.h. Texte, Datenbasen und -banken, Foto, Film, Graphiken, Software, Webpages, wissenschaftliche Primärdaten

=> Unter Bezugnahme auf die Sammelrichtlinien der jeweiligen Einrichtung

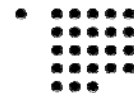


# Was soll gesichert werden? (2)

---

## **Objektbezogene Perspektive**

- ↪ **Verlagsprodukte**, d.h.
    - **originär digitale** Publikationen (born digital) auf **physischen Datenträgern** oder **online** verfügbar
    - **digitalisierte** Printprodukte
  - ↪ **Archivmaterialien** durch verwaltungsbezogen zuständige Archive
  - ↪ **Forschungsdaten**
  - ↪ **Museumsobjekte** durch besitzende Museum
  - ↪ **Websites** durch DNB und andere, gesetzlich verpflichtete oder freiwillig agierende Einrichtungen
  - ➔ **Sämtliche digitalen Zeugnisse** unseres täglichen Handelns
- 

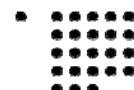


# Was soll gesichert werden? (3)

---

## **Nationale bzw. länderspezifische Perspektive (z.B. BRD): Sicherung der national relevanten digitalen Verlagsproduktion**

- ↪ **Digitale Publikationen auf physischen Datenträgern:**  
Sammelauftrag der DNB (bei deutscher Sprache und Bezug zu Deutschland) sowie anderer Pflichtexemplarbibliotheken (u.a. abhängig vom jeweiligen Landespflichtexemplarrecht oder entsprechenden Verordnungen)
- ↪ **Netzpublikationen, soweit öffentlich zugänglich:**  
Sammelauftrag der DNB seit 22.6.2006 ("Medienwerke in unkörperlicher Form"; soweit Bezug zu Deutschland); aber auch Auftrag und Zielsetzung einzelner Landesbibliotheken



# Was soll gesichert werden? (4)

---

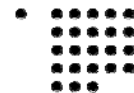
## **Exkurs:**

**Welche Netzpublikationen gehören zum Sammelauftrag der DNB und welche werden aktuell gesammelt?**

- ↪ „Zum Sammelgebiet Netzpublikationen gehören alle Darstellungen in Schrift, Bild und Ton, die in öffentlichen Netzen zugänglich gemacht werden. Die Abgabepflicht umfasst sowohl Internetpublikationen mit Entsprechung zum Print-Bereich als auch web-spezifische Medienwerke. Beispiele für Netzpublikationen sind elektronische Zeitschriften, E-Books, Hochschulprüfungsarbeiten, Digitalisate, Musikdateien oder auch Webseiten.“
- ↪ Gegenstand der momentanen Entwicklungsstufe ist die einzelobjektbezogene Ablieferung von **Netzpublikationen mit Entsprechung zum Printbereich**, also z. B. E-Books, elektronische Zeitschriften, Hochschulprüfungsarbeiten oder Digitalisate.

Quelle: [http://www.d-nb.de/netzpub/info/np\\_faq.htm](http://www.d-nb.de/netzpub/info/np_faq.htm)  
(Version vom 20.10.10; aufgerufen am 22.11.10)

- ↪ Die Konkretisierung des erweiterten gesetzlichen Auftrages erfolgte in der sog. **Pflichtablieferungsverordnung** v. 17.10.08, die durch die Neufassung der sog. **Sammelrichtlinien** am 1.6.2009 weiter konkretisiert wurde. (<http://files.d-nb.de/pdf/sammelrichtlinien.pdf> ; aufgerufen am 22.11.10)



# Was soll nicht gesichert werden?

---

## ***Nicht gesammelt werden sollen***

- ↪ „Zeitlich begrenzte Vorabveröffentlichungen, reine Software- oder Anwendungstools und auch Fernseh- und Hörfunkproduktionen ...

Ebenso werden Netzpublikationen, die rein gewerblichen Zwecken dienen, nicht gesammelt. Darunter fallen Webseiten mit Darstellungen der Dienstleistungen und Angebote einzelner Unternehmen für Kunden. Rein private Webseiten (Beispiel: private Fotos oder Urlaubsbeschreibungen) sind ebenso wenig sammelpflichtig wie Blogs ohne Themenbezug.“

Quelle: [http://www.d-nb.de/netzpub/info/np\\_faq.htm](http://www.d-nb.de/netzpub/info/np_faq.htm)  
(Version vom 20.10.10; aufgerufen am 22.11.10)



# Wer soll die dLZA übernehmen? (1)

---

↪ Prinzipiell kann jede Organisation diese Aufgabe übernehmen

↪ **Anforderungen gemäß der nestor-AG “Vertrauenswürdige Archive“**

„**nestor-Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive – Version 2**“ v. 13.11.08; urn:nbn:de:0008-2008021802  
<http://nbn-resolving.de/ResolverDemo.php>

## 1. Organisatorischer Rahmen

Zielentsprechende Verfügbarkeit von Ressourcen wie Personal, Finanzen, rechtliche und methodische Absicherung der LZA

## 2. Umgang mit den Objekten und ihren Metadaten

Integrität , Authentizität, Verfügbarkeit, Vertraulichkeit

## 3. Infrastruktur und Sicherheit

Technischen Aspekte des Gesamtsystems sowie die Aspekte der IT-Sicherheit

# Wer soll die dLZA übernehmen? (2)

---

↪ **Staatlich finanzierte Einrichtungen, z.B.:**

- Archive: Bundes-, Landes- oder regionale bzw. kommunale Ebene
- Museen (beliebige)
- Bibliotheken: Bundesebene = Deutsche Nationalbibliothek;  
Länderebene: jeweilige Landesbibliotheken;  
Fachebene: z.B. TIB, SSG-Bibliotheken

↪ **Privatrechtlich finanzierte Einrichtungen, z.B.:**

- Internet Archive
- PORTICO (<http://www.portico.org/>)

↪ **ABER:**

Einzelne Einrichtungen sind technisch und finanziell mit der Realisierung überfordert;  
die kooperative Entwicklung und Umsetzung der Methoden zur Erschließung und Archivierung bleibt eine gemeinsame Zielsetzung

# Wer soll die dLZA übernehmen? (3) Fallbeispiel

## Privatrechtlich finanzierte Einrichtung

↪ Internet Archive  
(<http://archive.org>)  
sammelt Momentaufnahmen von Webseiten und den darüber angebotenen Materialien

(z.B. 271 Versionen von [www.fbi.fh-koeln.de](http://www.fbi.fh-koeln.de)  
Stand: 22.11.10)

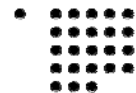
WaybackMachine

Enter Web Address: <http://> All Take Me Back info Search Contents Actions Pages

Searched for <http://www.fbi.fh-koeln.de> 271 Results

Note some duplicates are not shown. See alt.  
\* denotes when site was updated.  
Material typically becomes available here 6 months after collection. See FAQ.

Search Results for Jan 01, 1996 - Apr 27, 2009													
1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
0 pages	4 pages	4 pages	5 pages	8 pages	10 pages	10 pages	13 pages	13 pages	34 pages	38 pages	54 pages	5 pages	0 pages
	Apr 15, 1997	Jun 17, 1998	Jun 25, 1999	May 11, 2000	Jun 18, 2001	Mar 28, 2002	Feb 06, 2003	Jun 23, 2004	Jan 14, 2005	Jan 01, 2006	Jan 02, 2007	Jan 19, 2008	
	Apr 26, 1997	Jul 04, 1998	Feb 03, 1999	Jun 20, 2000	Feb 02, 2001	May 27, 2002	Feb 20, 2003	May 25, 2004	Feb 04, 2005	Jun 02, 2006	Feb 05, 2007	Feb 15, 2008	
	Jun 30, 1997	Dec 06, 1998	Feb 08, 1999	Jul 07, 2000	Mar 01, 2001	Jun 01, 2002	Apr 23, 2003	Jun 01, 2004	Feb 07, 2005	Jan 04, 2006	Feb 18, 2007	Mar 21, 2008	
	Oct 15, 1997	Dec 17, 1998	Apr 24, 1999	Aug 17, 2000	Mar 02, 2001	Jul 21, 2002	Jun 05, 2003	Jun 03, 2004	Feb 23, 2005	Jun 05, 2006	Mar 21, 2007	Apr 21, 2008	
			Apr 26, 1999	Oct 19, 2000	Apr 01, 2001	Aug 03, 2002	Jun 10, 2003	Jun 05, 2004	Feb 23, 2005	Jan 05, 2006	Apr 10, 2007	May 15, 2008	
				Nov 09, 2000	Apr 05, 2001	Aug 11, 2002	Jun 21, 2003	Aug 31, 2004	Mar 08, 2005	Jan 06, 2006	Apr 16, 2007		
				Nov 20, 2000	Apr 29, 2001	Sep 27, 2002	Jul 30, 2003	Sep 01, 2004	Mar 10, 2005	Jan 07, 2006	May 16, 2007		
					Jul 20, 2001	Sep 29, 2002	Aug 06, 2003	Sep 13, 2004	Mar 28, 2005	Jan 09, 2006	Jun 08, 2007		
					Sep 28, 2001	Nov 25, 2002	Sep 20, 2003	Sep 21, 2004	Apr 03, 2005	Jan 09, 2006	Jul 18, 2007		
					Dec 02, 2001	Nov 30, 2002	Oct 20, 2003	Oct 01, 2004	Apr 06, 2005	Jan 10, 2006	Aug 05, 2007		
							Oct 29, 2003	Sep 29, 2004	Apr 21, 2005	Jan 11, 2006	Aug 08, 2007		
							Nov 21, 2003	Oct 29, 2004	May 12, 2005	Jan 12, 2006	Aug 08, 2007		
							Dec 08, 2003	Nov 28, 2004	May 25, 2005	Jan 15, 2006	Aug 09, 2007		
									Jun 19, 2005	Feb 04, 2006	Aug 17, 2007		
									Jul 01, 2005	Feb 07, 2006	Aug 18, 2007		
									Jul 21, 2005	Feb 15, 2006	Aug 19, 2007		
									Jul 29, 2005	Apr 02, 2006	Aug 19, 2007		
									Aug 27, 2005	Apr 06, 2006	Aug 20, 2007		
									Sep 23, 2005	Apr 23, 2006	Aug 21, 2007		
									Oct 01, 2005	Apr 24, 2006	Aug 21, 2007		
									Oct 29, 2005	May 03, 2006	Aug 24, 2007		
									Nov 20, 2005	Jun 10, 2006	Aug 25, 2007		
									Nov 24, 2005	Jun 16, 2006	Aug 27, 2007		
									Dec 01, 2005	Apr 29, 2006	Aug 28, 2007		
									Dec 02, 2005	Jul 17, 2006	Aug 29, 2007		
									Dec 07, 2005	Jul 18, 2006	Aug 31, 2007		
									Dec 08, 2005	Aug 15, 2006	Sep 01, 2007		
									Dec 14, 2005	Aug 21, 2006	Sep 02, 2007		
									Dec 15, 2005	Aug 22, 2006	Sep 03, 2007		
									Dec 16, 2005	Aug 29, 2006	Sep 05, 2007		
									Dec 17, 2005	Oct 10, 2006	Sep 05, 2007		
									Dec 20, 2005	Oct 17, 2006	Sep 07, 2007		
									Dec 25, 2005	Nov 03, 2006	Sep 08, 2007		
									Dec 26, 2005	Nov 09, 2006	Sep 10, 2007		
									Dec 31, 2005	Nov 28, 2006	Sep 11, 2007		
										Dec 11, 2006	Sep 12, 2007		
										Dec 14, 2006	Sep 14, 2007		
										Dec 20, 2006	Sep 16, 2007		



## Exkurs: Beispiel Internet Archive

FH Köln - Fachbereich Bibliotheks- und Informationswesen (FBI) / Homepage - Mozilla Firefox

Willkommen am Fachbereich Bibliotheks- und Informationswesen (FBI) der Fachhochschule Köln (FH Köln)

[Aktuelles am FBI](#) [Der Fachbereich](#) [BID-Links](#) [Hilfe und Tips](#) [CLIR](#)

[Studienreform am Fachbereich](#)  
Sommerprogramm '07 der AKI Köln

Mit Hilfe der obigen Navigationsleiste erhalten Sie Zugriff auf folgende Angebote:

- [Aktuelle Themen am Fachbereich](#)
- [Der Fachbereich](#)
  - Fachschaff
  - Personen
  - Laborsangebau
  - Lehrveranstaltungen
- [Andere Bibliotheks-, Informations- und Dokumentationsstellen \(BID-Stellen\)](#)
- [Hilfe, Tips und Tricks für die Internet-Benutzung mit Ulrich Babak's \*\*Cockpit\*\*](#)
- [Informationsdienst Click!](#)

Außerdem erhalten Sie Informationen:

- [zu Veranstaltungen im BID-Bereich](#)
- [zum Berufsbild der Informationsfachleute](#)
- [zur Stadt Köln](#)

Fachhochschule Köln  
Cologne University of Applied Sciences

Institut für Informationswissenschaft

Willkommen auf den Seiten des Instituts für Informationswissenschaft

[Suche](#)

[English](#)

<p><b>Institut</b></p> <ul style="list-style-type: none"> <li>Kontakt</li> <li>Personen</li> <li>Projekte</li> <li>Publikationen</li> <li>Alumni</li> <li>Kooperationen</li> <li>Labore</li> <li>Fachschaft</li> </ul>	<p><b>Studium</b></p> <ul style="list-style-type: none"> <li>Studiengänge</li> <li>Lehrveranstaltungen</li> <li>Prüfungen</li> <li>Praxisphase</li> <li>Studieninformationen</li> <li>Studienreform</li> <li>Stud. Arbeitsergebnisse</li> <li>KOMPETENZPASS</li> </ul>	<p><b>Links</b></p> <ul style="list-style-type: none"> <li>IWS-Modüle</li> <li>Fachinformationen</li> <li>Benutzerkonten (ZVD)</li> <li>Bibliotheken der FH</li> <li>Mailinglisten am Institut</li> <li>Vervaltung der FH</li> </ul>	<p><b>ZBIW</b></p> <ul style="list-style-type: none"> <li>ZBIW - Aktuell</li> <li>Personen</li> <li>Programm</li> <li>Restplatzbörse</li> <li>Anmeldung</li> <li>Inhouse Seminare</li> </ul>
--	--	--	--

**Termine:**

- Vorlesungsverzeichnis WS 10/11 online - weitere Informationen
- Semesterterminplan WS 10/11 - weitere Informationen
- Semesterterminplan SoSe 11 - weitere Informationen

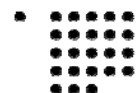
**Aktuelles:**

Die Absolventenfeier 2010 findet am 27.11.2010 um 19.30 in der Rotunde statt  
weitere Informationen

ZBIW: 15 Jahre bibliothekarische Fort- und Weiterbildung  
weitere Informationen

Das 3. Infoconomy Symposium zum Thema "Piraterie auf Informationsmärkten" findet am 28.10.2010 in der FH Köln statt.  
Ort: Rotunde der FH Köln, Claudiusstr. 1, Beginn: 9.30 Uhr  
weitere Informationen

Erfolgreicher Abschluss des Assessments-Verfahrens für den MAUIS-Jahrgang 2011  
weitere Informationen



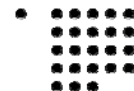
# Strategische Zielsetzungen der LZA + LZV digitaler Objekte

---

Abgestuftes Ziel unter dem Aspekt der Langzeitverfügbarkeit ist es, dem späteren Nutzer möglichst viel der ursprünglichen Inhalte eines digitalen Objektes und der damit verbundenen Anwendungserfahrungen zugänglich zu machen.

Diese Zielsetzung der LZA und LZV kann auf der Basis unterschiedlicher strategischen Zielsetzungen angestrebt werden:

- ↪ Sicherung primär des intellektuellen Content:  
"intellectual preservation" (Peter Graham; 1994)  
<http://www.clir.org/pubs/reports/graham/intpres.html>
- ↪ Sicherung des Content *mit Teilen der Funktionalität* der originalen Anwendungsumgebung
- ↪ Sicherung des Content *und* der originären Anwendungsumgebung



---

## Sicherungswürdige Aspekte eines digitalen Objekts

---

**Inhalt:** Information in Schrift, (bewegten) Bildern, Ton, ...

**Struktur:** Aufteilung und Abfolge der Daten (z.B. DTD)

**Layout:** Visualisierung des Inhalts + der Struktur (z.B. Stylesheets)

**Metadaten:** Objekteigenschaften + deren Veränderungen (z.B. DC)

Orientiert an einer Folie von S. Dobratz aus Neuroth 29.11.2005  
nestor-Seminare: Einführung in die Langzeitarchivierung digitaler Objekte, Göttingen,  
29.11.2005. Einführung in die Langzeitspeicherung und ... im Archivbereich,  
Göttingen, 13.1.2006



# Reduzierung der Komplexität der LZA digitaler Objekte

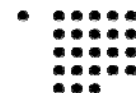
---

Die Komplexität der Aufgabe kann eingegrenzt werden durch:

- ↪ Vorauswahl bestimmter Objekttypen (z.B. nur Textdokumente)
- ↪ Vorauswahl bestimmter Dateiformate (z.B. PDF/A, XML o.ä.) vgl. die bisherige Präferenzregelung der DNB

1. PDF/A
2. Andere PDF-Versionen
3. HTML
4. PS
5. Weitere XML-basierte Formate, TXT
6. Sonstige (DVI, RTF, etc.)

- ↪ Vorauswahl bestimmter Anwendungsumgebungen (z.B. nur MS-Windows Betriebssysteme + Office-Umgebungen; nur OO-Anwendungen)



---

## Unterschied zwischen Verfahren zur Datensicherung und digitalen LZA

---

Sehr häufig werden **Verfahren zur Datensicherung** mit Maßnahmen zur digitalen LZA verwechselt! Bei der digitalen LZA geht es jedoch darum, digitale Objekte für die Nachwelt **dauerhaft** zu erhalten.

Datensicherung ist dafür eine Voraussetzung!

Beispiele für Datensicherungsverfahren (sg. Bitstream-Sicherung):

- ↪ Auffrischung der Datenträger (digital refreshing) = Datenerhaltung;  
*Sonderform*: Encapsulation (=Kapselung)  
d.h. image-file der ursprünglichen HW- und SW-Umgebung + der Anwendung, Ablauf ggf. auf virtuellem Computer
- ↪ LOCKSS-Konzept (Lots of Copies Keep Stuff Safe)  
d.h. viele Kopien einer Datei auf vernetzten Computern schaffen Redundanz und damit Datensicherheit



# OAIS: Referenzmodell für LZA+LZV

## ↪ Open Archive Information System (OAIS)

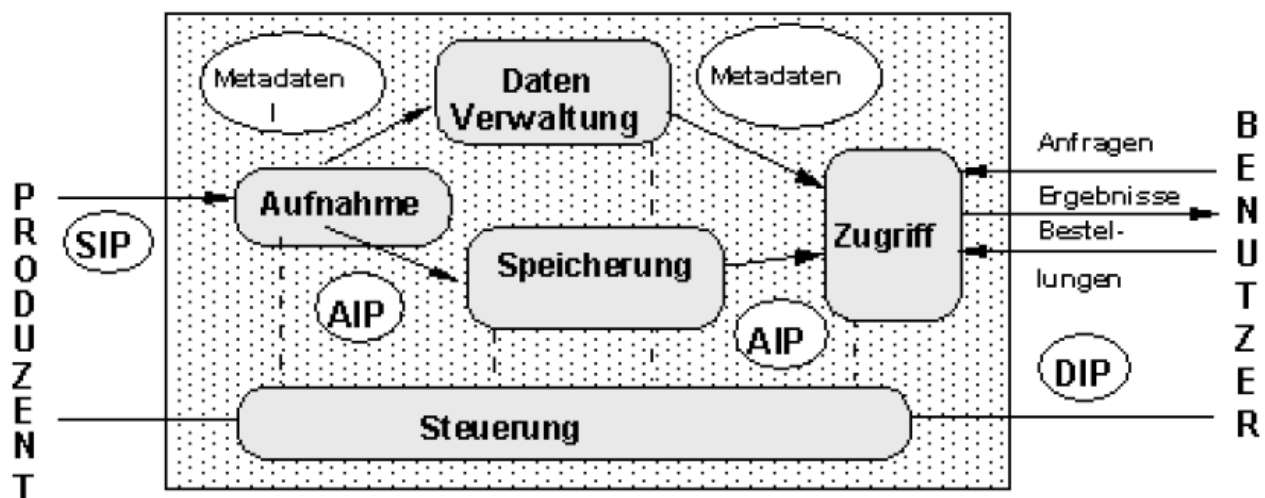
= **das** Referenzmodell zur Langzeitarchivierung und Verwaltung digitaler Objekte (es wird in fast jeder Präsentation gezeigt ☺)

- „Das als ISO 14721 verabschiedete Referenzmodell beschreibt ein Archiv als Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, Informationen zu erhalten und einer definierten Nutzerschaft verfügbar zu machen.“  
aus: Brübach, in: Das Referenzmodell OAIS – Open Archival Information Systemnestor-Handbuch, 2010; [ Version 2.3 ] Kap.4:3 (aufgerufen: 22.11.10)  
[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_474.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_474.pdf)

Das Referenzmodell spezifiziert weder das Design noch die Implementierung eines OAIS-konformen Archivs.

- Zentrale Begrifflichkeiten im OAIS-Modell sind:  
SIP = Submission Information Package  
AIP = Archival Information Package  
DIP = Dissemination Information Package

# OAIS: Referenzmodell für LZA+LZV



## Open Archive Information System (OAIS)

Auch interaktiv als javascript über die Nedlib-Homepage unter <http://nedlib.kb.nl/> =>results => Functional Process Model

# Aktuell relevante Verfahren zur digitalen LZA

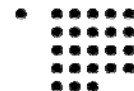
---

Verfahren zur **digitalen** LZA:

- ↪ Technisches Museum: ursprüngliche HW- und SW-Kombination wird in ihrer Funktion erhalten (... bis ein Bauteil fehlerhaft ist)
- ↪ Emulation: Nachahmung der ursprünglichen HW-, SW-Umgebung und des Betriebssystems mittels spezieller Emulationssoftware
- ↪ Migration: regelmäßige Übertragung des digitalen Objekts von einer HW- und SW-Umgebung in eine technologisch nachfolgende Systemumgebung (ggf. Zwang zum Formatwandel)  
=> dann erhöhte Gefahr funktionaler und inhaltlicher Verluste

*Sonderfall digital => analog:*

- ↪ Konversion: Übertragung auf analoge Speichermedien / Datenträger (Papier; MF), dabei zumeist Aufgabe der informationellen Mehrwerte der digitalen Anwendung



## Grundsätzliche Probleme

---

- ↪ Kopieren und manipulieren sind im digitalen Kontext extrem einfach  
=> gewollte oder ungewollte Veränderungen des originalen Content
  - ↪ Authentifizierung der Originalversion durch die Urheber unter Einbeziehung von Quelle, Erstellungstermin, Inhalt und Wiedergabeumgebung
  - ↪ Diverse Sicherungsverfahren in Erprobung:
    - Kryptographie – problematisch wg. Zugang
    - digitale Fingerprints mittels Hash-Code
    - Zeitstempel
    - verborgener Code z.B. Wasserzeichen
- Problem: Sicherungsverfahren kollidieren z.T. mit den technischen Archivierungsverfahren



# Was wird / könnte LZA kosten? (1)

---


## ↪ **Grundannahme:**

Archivierung und Verwaltung digitaler Objekte ist aufwändiger als bei gedruckten Publikationen

## ↪ Bislang kaum langfristige Erfahrungswerte, bestenfalls sind **Kostenmodelle** verfügbar

z.B.:

Kalkulation im Projekt JSTOR (The Scholarly Journal Archive) oder Preisstruktur des PORTICO-Dienstes;

ein systematisches Kostenmodell wird entwickelt im LIFE-Projekt (LIFE = **L**ife cycle **I**nformation **F**or **E**-literature) an der British Library 

# Was wird / könnte LZA kosten? (2)

---

**Kostenfaktoren** (nach Keller, Alice: Elektronische Zeitschriften, Grundlagen und Perspektiven, Wiesbaden 2005, 255f)

## ↪ **Einmalige Kosten:**

- "Auswahl des zu archivierenden Materials
- Prüfung der Rechtslage
- Festlegung der technischen Lösung ...
- Prüfung der Vollständigkeit der digitalen Objekte
- Herstellung oder Ergänzung von Metadaten"

## ↪ **Langfristig wiederkehrende Kosten:**

- "Erneuerung von Hard- und Software
- Speicherung der Dateien, regelmäßiges Überspielen oder Migration der Daten in neue Computerumgebungen
- Verwalten des Archivs
- Regelmäßiges Überprüfen der Sammel- und Archivierungspolitik"

# Was wird / könnte LZA kosten? (3)

---

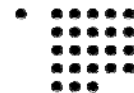
## Kostenfaktoren

(nach Wollschläger / Dickmann: Kosten

in: nestor-Handbuch, 2009, [ Version 2.0 ] Kap.14:2

[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_417.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_417.pdf)  
aufgerufen: 22.11.10)

- ↪ Initiale Kosten
- ↪ Beschaffungskosten
- ↪ Betriebskosten



## Was lief bislang bzgl. LZA?

### *... in Deutschland*

Nestor, der Berater der Griechen

in Troja, Symbolfigur für die beratende  
und unterstützende Funktion

### ↪ **Projekt nestor: Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland**

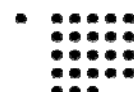
<http://www.langzeitarchivierung.de>

**Network of Expertise in long-term STOrage and availability of  
digital Resources in Germany**

Phase 1: 2003-2006; Phase 2: 2006-2009 (BMBF-Förderung)  
Phase 3: 2009 - heute (unsubventioniert)

*Projekthistorie:*

[http://www.langzeitarchivierung.de/ueber\\_uns/projektgeschichte.htm](http://www.langzeitarchivierung.de/ueber_uns/projektgeschichte.htm)



# Was läuft derzeit bzgl. LZA? (1)

## Aktivitätsbereiche von Projekten in Deutschland zur LZA digitaler Ressourcen:

### Projekte und Aktivitäten der ...

- ↪ **Deutschen Nationalbibliothek** in Kooperation mit **Landesbibliotheken** und weiteren Partnern zur Erprobung von Strategien und organisatorischen Abläufen für digitale Pflichtexemplare und Webseiten  
Projekte (u.a.): kopal; DP4lib (Digital Preservation for Libraries); LuKII eine Interoperabilität zwischen den Open-Source-Teilen von kopal und LOCKSS
- ↪ **Zentralen Fachbibliotheken** und andere **Spezialbibliotheken** fokussieren auf den Umgang mit digital(isiert)en Dokumenten, Forschungsdaten und Datenbanken
- ↪ **Staats- und regionalen Archive** sowie **Museen** zur Auswahl und Erprobung von Techniken für die Bewahrung digitaler Anwendungen und Objekte
- ↪ **IT-Zentren** und **-Experten** mit Blick auf die technische Umsetzung der Strategien und Techniken
- ↪ **Forschungseinrichtungen** zur kurz- und mittelfristigen Bewahrung von Forschungsdaten



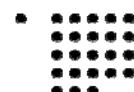
# Was läuft derzeit bzgl. LZA? (2)

## ... in anderen Ländern

- ↪ **PADI Preserving Access to Digital Information** der National Library of Australia; 1997 – 7/2010  
nun archiviert unter <http://nla.gov.au/nla.arc-10691>
- ↪ Gute Übersicht zu den Aktivitäten in einzelnen Ländern
- ↪ Methodisch & technologisch führend sind u.a. Australien, China, GB, NL und BRD
- ↪ siehe iPRES-Konferenzen seit 2004 jährlich, zuletzt 2010 in Wien



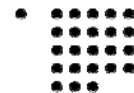
<http://rdd.sub.uni-goettingen.de/conferences/ipres/ipres-en.html>



# Ausgewählte offene Fragen (1)

---

- ↪ Strategische Konzeption der LZA + LZV: Inwieweit ist das OAIS-Reference Model tatsächlich langfristig tragfähig?
- ↪ Welche Techniken sind wirklich funktional und auch für den automatisierten Massenbetrieb geeignet?
- ↪ Standards zur Dokumenterstellung u. -kodierung:  
Ziel: Kostensenkung bei Emulation o. Migration durch Nutzung von Auszeichnungssprachen
- ↪ Metadatenstandards und Metadatenmanagement:  
Welche Standards sind geeignet + international konsensfähig?
- ↪ Anforderungen an Dokumentformate:  
“Archivierungsfreundlichkeit”
- ↪ LZA von Forschungsdaten



# Ausgewählte offene Fragen (2)

---

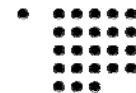
- ↪ Rechtliche Fragen des Zugriffs / DRM
- ↪ Dokumentidentifikation mittels persistenter Identifikatoren (z.B. DOI; URN)
- ↪ Transferstandards und -protokolle
- ↪ Kosten und Finanzierung der LZA
- ↪ Funktionsverteilung und Organisation zwischen Produzenten und Anbietern digitaler Objekte einerseits und Organisationen der LZA andererseits
- ↪ Formen der Zusammenarbeit auf internationaler, nationaler und regionaler Ebene



# Weiterführende Quellen & Hinweise

---

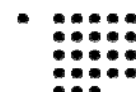
- ↗ Borghoff, Uwe M. et al.: Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg 2003 (auch als eBook)
- ↗ nestor-Website: <http://www.langzeitarchivierung.de>
- ↗ nestor-Ratgeber Media <http://www.archaeobooks-kalkriese.de/nestor/>
- ↗ Liegmann/Neuroth (2010). Einführung. In: Neuroth, H. et al.: [Ed.] (2010): nestor Handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Version 2.3), Boizenburg urn:nbn:de:0008-2010030508 bzw. [http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_390.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_390.pdf)
- ↗ Neuroth, H. et al. [Ed.] (2010). nestor Handbuch - Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Version 2.3), Boizenburg <http://nestor.sub.uni-goettingen.de/handbuch/index.php>
- ↗ Informationen der DNB zu Netzpublikationen <http://www.d-nb.de/netzpub/index.htm>
- ↗ PADI – Preserving Access to Digital Information (1997-6/2010) <http://nla.gov.au/nla.arc-10691>



# Weiterführende Quellen & Hinweise

---

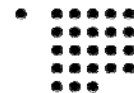
- ↗ Kirchhoff, Amy et al.: BECOMING A CERTIFIED TRUSTWORTHY DIGITAL REPOSITORY: THE PORTICO EXPERIENCE. iPRES 2010, Wien <http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/Kirchhoff-35.pdf>
- ↗ **UNESCO Charta zur Bewahrung des digitalen Kulturerbes** (32. UNESCO-Generalkonferenz; 17. Oktober 2003); <http://www.unesco.de/444.html?&L=0> (aufgerufen: 22.11.10)  
Pflichtablieferungsverordnung vom 17. Oktober 2008 (BGBl. I S. 2013) <http://www.gesetze-im-internet.de/pflav/BJNR201300008.html> (22.11.10)
- ↗ *Sammelrichtlinien / Deutsche Nationalbibliothek.*  
[Red.: Angela Matthias ; Brigitte Wiechmann]  
Deutsche Nationalbibliothek <Leipzig; Frankfurt, Main>. - Leipzig : Dt. Nationalbibliothek, 2009, Stand: 1. Juni 2009  
<http://files.d-nb.de/pdf/sammelrichtlinien.pdf> ; (22.11.10)
- ↗ Lifecycle Information for E-literature, A summary from the LIFE project; Report Produced for the LIFE conference 20 April 2006, <http://eprints.ucl.ac.uk/1855/1/LifeProjSummary.pdf> (22.11.10)



# Bildnachweise

---

- **3,5“ Diskette**  
<http://upload.wikimedia.org/wikipedia/commons/thumb/3/36/3%2C5%22-Diskette.jpg/275px-3%2C5%22-Diskette.jpg> (24.10.09)
- **5 1/4“ Diskette**  
<http://upload.wikimedia.org/wikipedia/commons/thumb/1/1e/5.25%22-Diskette.jpg/707px-5.25%22-Diskette.jpg> (24.10.09)
- **Festplatte**  
[http://de.wikipedia.org/w/index.php?title=Datei:Samsung\\_HD753LJ\\_03-Opened.jpg&filetimestamp=20091013081316](http://de.wikipedia.org/w/index.php?title=Datei:Samsung_HD753LJ_03-Opened.jpg&filetimestamp=20091013081316) (24.10.09)
- **Compact disc**  
[http://de.wikipedia.org/w/index.php?title=Datei:Compact\\_Disc.jpg&filetimestamp=20070410122954](http://de.wikipedia.org/w/index.php?title=Datei:Compact_Disc.jpg&filetimestamp=20070410122954) (24.10.09)



## Beispiele für Integritätsprobleme

---

### Zum Erhalt von Integrität

rn -> rn -> r n (eingeschobenes Leerzeichen)

1365,89 -> -1365,89

1365,89 -> 1365,89

1365,89 + 1365,89 = 0,00 -> 1365,89 + 1365,89 = 2731,78

Quelle: Borghoff, Uwe M.: Vertrauenswürdige digitale Langzeitarchive, vom 11. 5.2007,  
[www.babs-muenchen.de/content/workshop/1\\_vwdlza\\_roedig.pdf](http://www.babs-muenchen.de/content/workshop/1_vwdlza_roedig.pdf) (aufgerufen: 24.10.09)



# Was wird / könnte LZA kosten? Ergebnisse des LIFE-Projektes

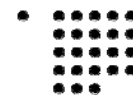
**Lifecycle Information for E-literature**, A summary from the LIFE project  
Report Produced for the LIFE conference 20 April 2006, p. 4

LIFE further predicts that in the tenth year of the same digital assets' existence;

- The total lifecycle cost for a hand-held e-monograph is £48
- The total lifecycle cost for a hand-held serial is £14 per issue
- The total lifecycle cost for a non hand-held e-monograph is £30
- The total lifecycle cost for a non hand-held e-serial is £8 per issue
- The total lifecycle cost for a new website is £6,800
- The total lifecycle cost for an e-journal is £3,000

It is in this predictive work that further research is required. For example, by year ten significant rises are measured for both Web Archiving and e-journals yet e-serials reduce per issue in cost. These figures come from a small sample of the collections and must be tested further to see if these costs are constant.

<http://eprints.ucl.ac.uk/1855/1/LifeProjSummary.pdf> (aufgerufen 24.10.09)



# Was wird / könnte LZA kosten? Kostenvergleich Migration - Emulation

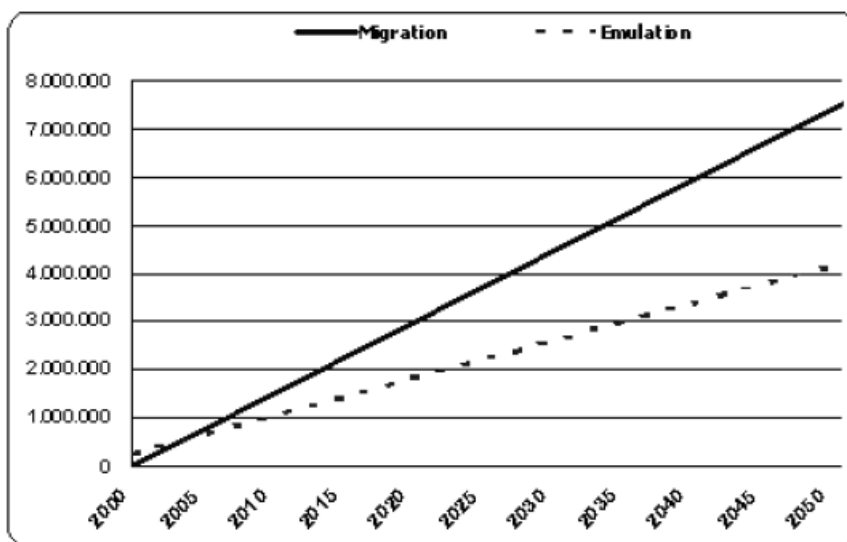


Figure 2: Costs for migration (blue line) and emulation (red dotted line) for maintaining an archive of 1,000,000 digital objects over a period of 50 years. The initial investments of setting up an emulation tool yields high costs in the first five years. But soon after that, the migration costs are higher than the emulation costs and the difference increases every year. In 50 years, the migration costs are 45% higher than the emulation costs.

Aus:

Erik Oltmans / Nanda Kol:  
**A Comparison Between  
Migration and Emulation  
in Terms of Costs.** –

In: RLG DigiNews April 15,  
2005, Volume 9, Number 2  
[http://www.worldcat.org/arcviewer/  
1/OCC/2007/08/08/0000070519/  
viewer/file959.html#article0](http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html#article0)  
(24.10.09)

