

Digitale Bücher ins LRZ !?

Aufgaben und Probleme der Langzeitverfügbarkeit von Information in digitaler Form



Agenda

- 1 Aufgabe: Erhaltung der Langzeitverfügbarkeit von Information in digitaler Form
- 2 Probleme: begrenzte Lebensdauer von Speichermedien, rascher technologischer Wandel
- 3 Strategie: Aufbau eines vertrauenswürdigen digitalen Langzeitarchivs
 - 3.1 Referenzmodell: Definitionen, Funktionalitäten, Informationsmodell
 - 3.2 Schnittstelle Produzent – Archiv
 - 3.3 Vertrauenswürdigkeit
4. Die Lösung der BSB in Kooperation mit dem LRZ
 - 4.1 Das Münchener Digitalisierungszentrum
 - 4.2 BABS: Bibliothekarisches Archivierungs- und Bereitstellungssystem
- 5 Ausblick
- 6 Referenzen

1 Aufgabe

- Produktion von Information in digitaler Form nimmt stetig zu:
- Kampagnen zur Digitalisierung:
 - Digitale Produktion von Lexika, Editionen, kritischen Gesamtausgaben, etc.
 - Retro-Digitalisierung
- **Erhaltung der Information in digitaler Form!**

2 Probleme

- Information wird in digitaler Form durch Daten (Folge von 0, 1) repräsentiert, die auf Speichermedien vorgehalten werden.
- Zur Interpretation dieser Daten ist eine komplexe technische Infrastruktur notwendig:
 - Speichermedium: CD
 - Lesegerät: CD-Laufwerk
 - Treiber für das CD-Laufwerk
 - Rechner mit Bildschirm
 - Betriebssystem Microsoft Windows xx
 - Für Anzeige oder Bearbeitung: Anwendungssoftware: z.B. Textverarbeitungsprogramm: Microsoft Word xx

2 Probleme

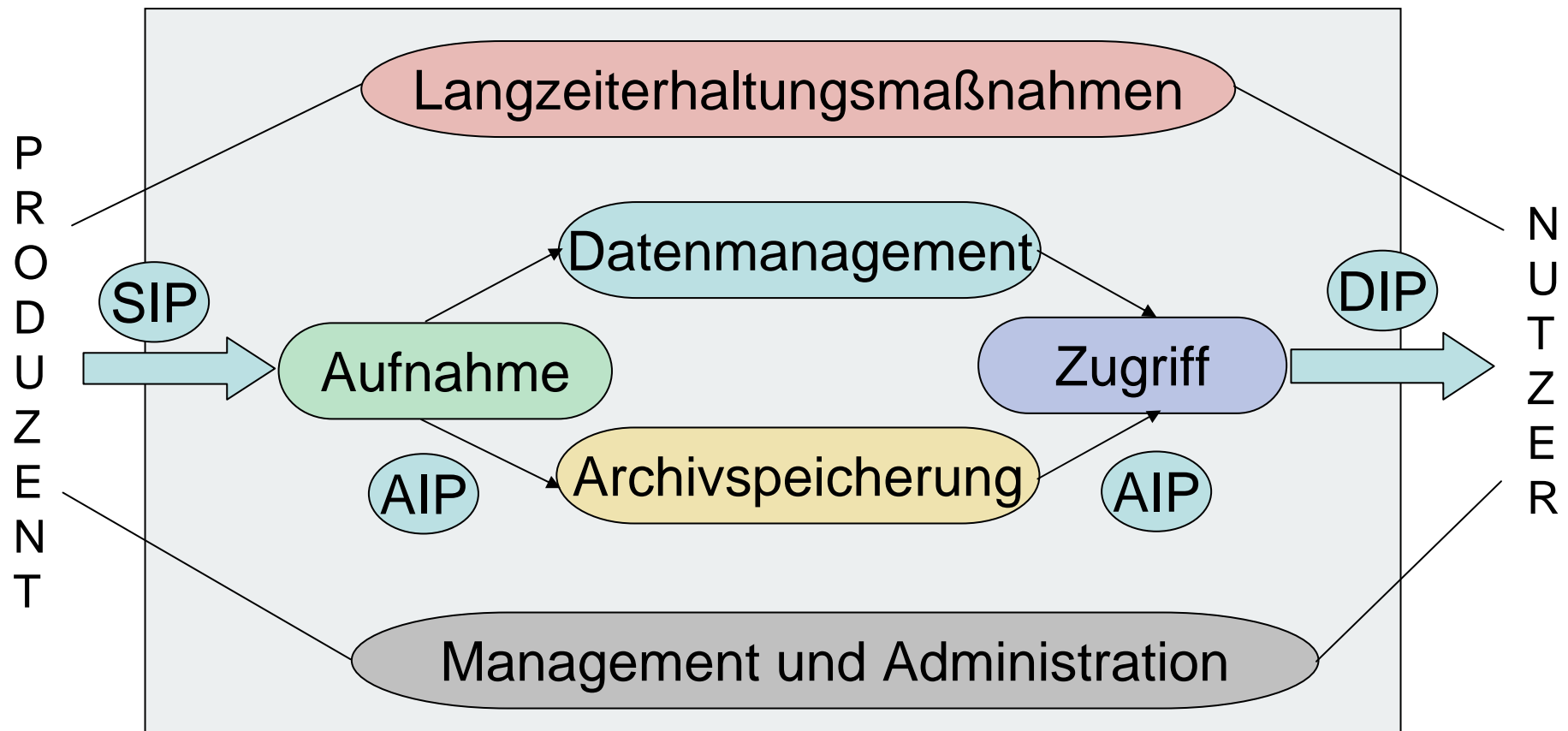
- Zur Erhaltung der Information:
 - 1. Erhaltung der Daten (Bitstream Preservation)
 - 2. Interpretierbarkeit der Daten (Content Preservation)
 - Zu 1. kurze Lebensdauer der Speichermedien: jedoch Rechenzentren, Dienstleister
 - Zu 2. rascher technologischer Wandel: eigentliche Herausforderung
- Alternative: Speicherung auf analogen Medien z.B. Mikrofilm ?

3. Aufbau eines vertrauenswürdigen digitalen Langzeitarchivs

3.1 Referenzmodell

- OAIS: Open Archival Information System, Reference Model
 - ISO-Standard ISO 14721:2003
- Definition: **digitales Langzeitarchiv**
 - eine Organisation (bestehend aus Personen und technischen Systemen)
 - die die Verantwortung für den Langzeitverfügbarkeit von Information in digitaler Form
 - zum Zwecke der Nutzung durch eine bestimmte Zielgruppe.

OAIS: Funktionalitäten



OAIS: Informationsmodell

- Lebenszyklus der Daten im digitalen Langzeitarchiv:
Informationspakete:
 - Übergabepaket: SIP
 - Archivpaket: AIP
 - Nutzungspaket: DIP
- Zum AIP gehören:
 - Inhaltsdaten,
 - Repräsentationsinformation,
 - LZA-relevanten Informationen

OAIS: Langzeiterhaltungsmaßnahmen

- Migration
 - Transformation der Daten in ein aktuelles Format
- Emulation
 - Nachbildung der alten Anwendungsumgebung in einem neuen System
- Museumsansatz
 - Erhalt der alten Hard- und Software (System muss lauffähig bleiben)

3.2 Schnittstelle Produzent - Archiv

PAIMAS: Producer Archive Interface

– ISO 20652:2006

Wege ins Archiv, nestor-Ratgeber

Aufnahme der Information ins digitale Langzeitarchiv von besonderer Bedeutung, nicht nur technisch

- Übergang der Verantwortung
- Komplexer Vorgang
- Hohe Qualitätsanforderungen
 - Kostenfaktor

Schnittstelle Produzent - Archiv

- Was soll archiviert werden? Welche Eigenschaften sollen erhalten werden?
 - Sammelrichtlinien, Selektionskriterien, Profile,
- Wie wird die Information vom Produzenten an das Archiv übertragen?
 - Transferpakete, Transferwege
 - Qualitätskontrolle
- Management des Prozesses
 - Übernahmevereinbarung, Gesetze, Erlasse..
 - Dokumentation

3.3 Vertrauenswürdigkeit

- Definition: **Vertrauenswürdigkeit**
 - Eigenschaft eines Systems, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt).
 - Es kann dieses glaubhaft machen

(RFC 2828, Internet Security Glossary)

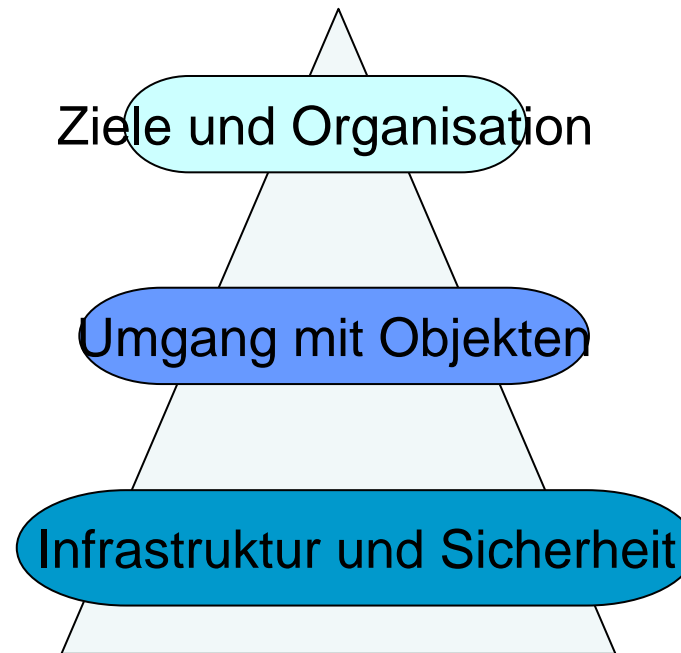
Vertrauenswürdigkeit

- Vertrauenswürdigkeit heute:
 - Kriterienkataloge, Tools zur Evaluierung
 - Orientierung
 - Dokumentation, Transparenz
 - Test-Audits
- Vertrauenswürdigkeit morgen:
 - Formaler Zertifizierungsprozess
 - › Basierend auf Standards (ISO, DIN)

- Entwicklung im internationalen Kontext, orientiert an den Bedürfnissen und Rahmenbedingungen der deutschen Anwender, insbesondere Gedächtnisorganisationen
- Zielgruppe
 - Gedächtnisorganisationen
 - Dienstleister, Softwareentwickler, u.a.
- Leitfaden für Konzeption, Planung und Umsetzung sowie Selbstevaluierung
- Normierung im Rahmen der DIN



nestor Kriterienkatalog:



Ausgewählte Kriterien

Aus dem Bereich Ziele und Organisation:

- Die Finanzierung des digitalen Langzeitarchivs ist sichergestellt.
- Es steht Personal mit angemessener Qualifikation in ausreichendem Umfang zur Verfügung.
- Es bestehen angemessene Organisationsstrukturen für das dLZA.

Ausgewählte Kriterien

Aus dem Bereich Umgang mit Objekten:

- Das dLZA identifiziert seine Objekte und deren Beziehungen eindeutig und dauerhaft.
- Das dLZA erhebt in ausreichendem Maße Metadaten:
 - für eine formale und inhaltliche Beschreibung
 - Für strukturelle Beschreibung
 - Für technische Beschreibung
 - Zur Dokumentation der Veränderungen
 - Zur Steuerung der Nutzung
- Das dLZA definiert seine Archivobjekte
 - Paketstruktur
 - Dateiformate

Mitwirkung des Produzenten

- Archivfähige Objektformate, Dateiformate
- Strukturbeschreibung
- Qualität der Objekte
- Mitwirkung bei Erschließung
- Übertragung der Verantwortung an ein vertrauenswürdigen digitales Langzeitarchiv
- Übertragung der Rechte
- Finanzierung

4. Die Lösung der BSB in Kooperation mit dem LRZ

4.1 Das Münchener Digitalisierungszentrum

Münchener Digitalisierungszentrum
Digitale Bibliothek

- **MDZ** gegründet 1997 mit Anschubfinanzierung der DFG als eines der beiden Kompetenzzentren für Digitalisierung in Deutschland
- Inzwischen verstetigt in der **Digitalen Bibliothek** (Digitalisierung, technische Unterstützung von Projekten, LZA)
- mehr als 100 Projekte, Vielzahl in Kooperation mit anderen Forschungseinrichtungen
- Bei LZA seit 2004 Kooperation mit LRZ



Digitalisierung am MDZ

- **Vier Säulen:**
- Drittmittelprojekte
- Digitalisierung zum Bestandsschutz/Bestandssicherung
- Digitisation On Demand
 - wissenschaftliche Zwecke
 - kommerzielle Nutzung
- Public Privat Partnership mit Google



Ausgewählte Kooperationsprojekte

Münchener Digitalisierungszentrum
Digitale Bibliothek

- ADB/NDB Digitales Register
 - Historischen Kommission der Bayerischen Akademie der Wissenschaften
- Regesta Imperii
 - Mainzer Akademie
- 1000 Schlüsseldokumente der russischen Geschichte
 - Russischen Akademie
- Matrikelbücher der Akademie der bildenden Künste
- dMGH u.a.
 - › www.digitale-sammlungen.de

ADB Allgemeine / Neue
Deutsche
NDB Biographie

Digitales Register

Massendigitalisierung

- VD16 II
- Google Library Copy
 - Optimierung der Workflows
 - Anpassung der Hard- / Software in Bibliothek und Rechenzentrum, Netzwerkverbindung



4.2 BABS

BABS

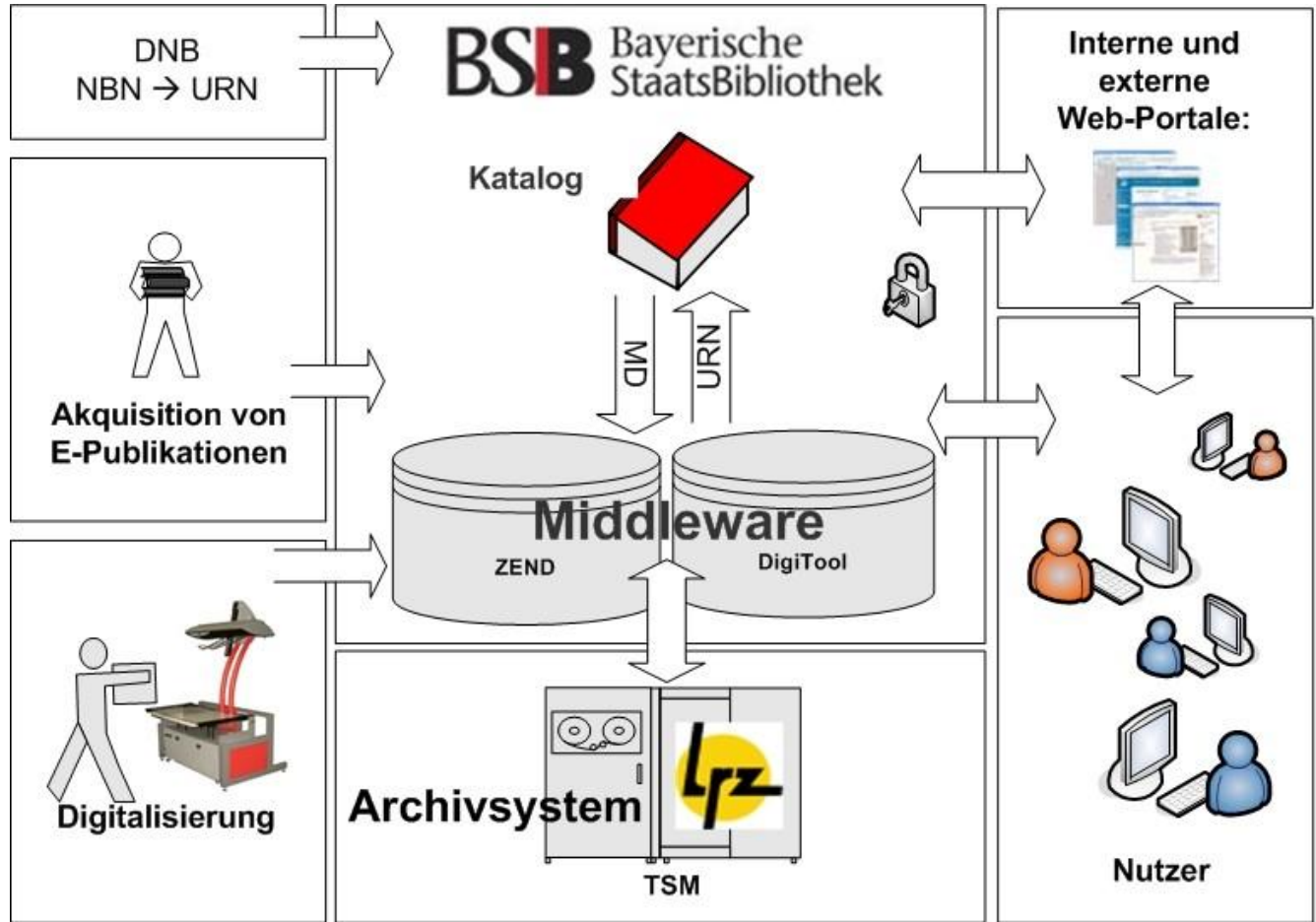
Das Bibliothekarische
Archivierungs- und
Bereitstellungssystem



- Bibliothekarisches Archivierungs- und Bereitstellungssystem
 - Partner: LRZ
 - › www.babs-muenchen.de
- DFG-Projekte
 - 1. Phase 2005-2007: Aufbau eines Pilotsystems zur LZA von Netzpublikationen
 - 2. Phase 2009-2010: Ausbau von BABS zu einem
 - › vertrauenswürdigen und
 - › skalierbaren digitalen Langzeitarchiv
 - › als Teil einer nationalen Lösung



BABS-Architektur



BABS - Fokus

- Lokale/ regionale/ nationale Verantwortung:
- Materialfokus: breites Spektrum einer Universalbibliothek von überregionaler Bedeutung
 - Eigene Produktion: Digitalisate, ZEP
 - Kooperationsprojekte mit universitären und außeruniversitären Forschungseinrichtungen
 - Bayerische AD, Bundes AD (Erlass zur Abgabe amtlicher Veröffentlichungen an Bibliotheken)
 - Bayerische „Pflicht“ (Gesetz in Novellierung)
 - Lizenzierte Materialien (lokal, regional, national)
 - SSG: Wissenschaftlich relevante open access Publikationen und Websites

LRZ als strategischer Partner

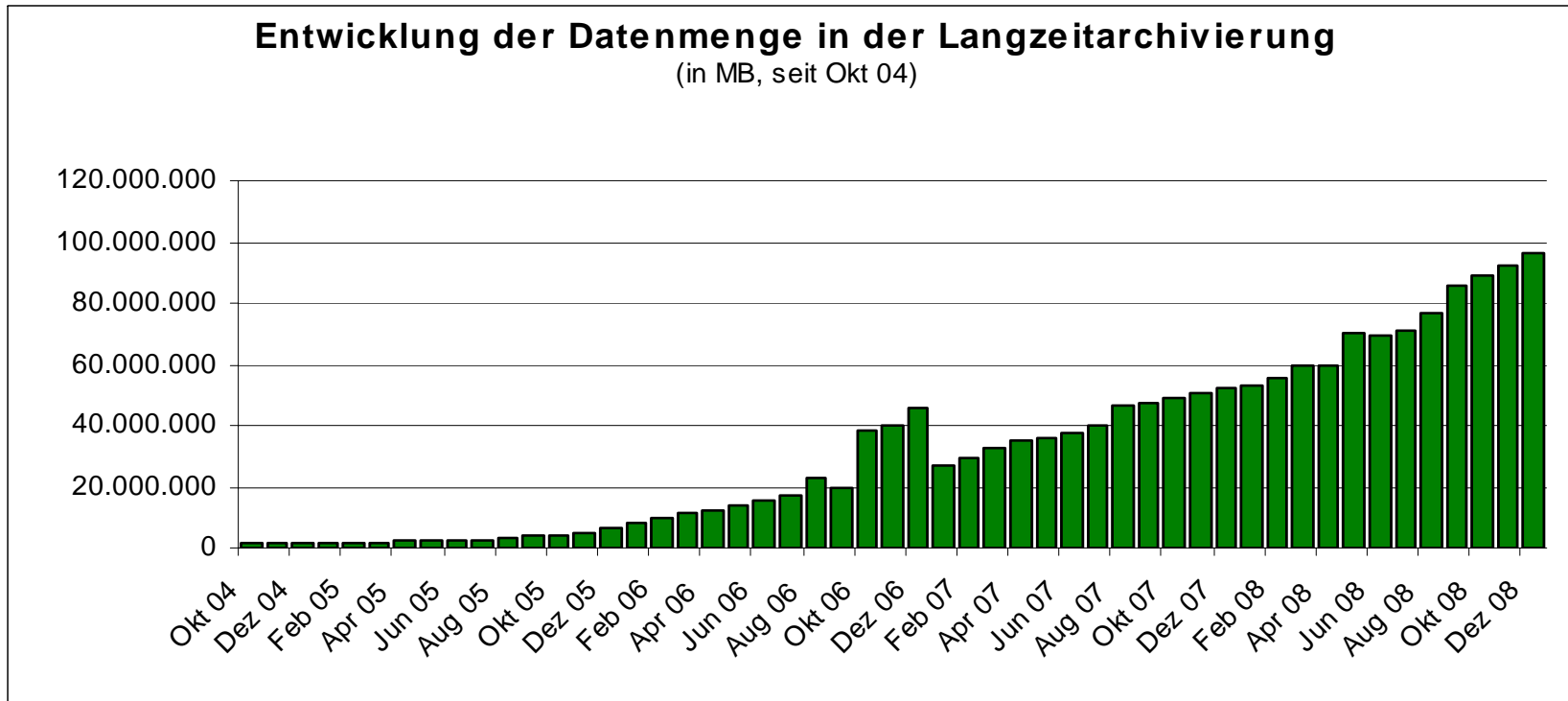


- Höchstleistungs-Rechenzentrum
- Archiv- und Backupsystem TSM mit Band-Bibliotheken (Redundante Speicherung beim RZG der MPG)
- Ausbau der Kooperation, weitere Aufgaben
 - Hosting Hard- und Software
 - Massendigitalisierungsprojekte: VD16 digital, PPP mit Google





BABS-Zahlen



Zum Jahreswechsel 2008/9: 100 TB überschritten

Jährlicher Zuwachs ab 2009: über 100 TB

Weitere Projekte

- nestor: Aufbau eines Kompetenznetzwerks für die LZA in Deutschland:
 - Informationen sammeln, erstellen und verbreiten
 - Experten und Interessierte zusammenführen
 - Knotenpunkt für internationale Kontakte
 - › www.langzeitarchiverung.de
- Auf EU-Ebene: DPE, Caspar, Planets u.a.
 - Tools: Evaluierung, Planung, Testbed
 - Schulungen
 - › www.wepreserve.eu

5 Ausblick

- *„Langzeit“ bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.*
 - › *Liegmann, Schwens*
- Ständige Aufgabe, neue Herausforderungen
 - Technology Watch
 - Community Watch
 - Preservation Planning

6 Referenzen

- Uwe M. Borghoff, Peter Rödiger, Jan Scheffczyk, Lothar Schmitz: *Langzeitarchivierung*. Dpunkt Verlag, 2003 [ISBN 3-89864245-3](#)
- OAIS:<http://public.ccsds.org/publications/archive/650x0b1.pdf>
- PAIMAS:<http://public.ccsds.org/publications/archive/651x0b1.pdf>
- METS:[http://www.**loc**.gov/standards/**mets**/](http://www.loc.gov/standards/mets/)
- PREMIS:[http://www.**loc**.gov/standards/**premis**/](http://www.loc.gov/standards/premis/)
- Nestor-Kriterien: <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>
- Nestor-Handbuch:<http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf>
- Nestor-Wege ins Archiv:<http://nbn-resolving.de/urn:nbn:de:0008-2008103009>
- BABS: www.babs-muenchen.de, www.digitale-sammlungen.de
- Nestor: www.langzeitarchivierung.de