

Archivierung digitaler Dokumente

Problemstellung - kooperative
Lösungsansätze - Aktivitäten der
Österreichischen Nationalbibliothek
Teil 1 (Max Kaiser)

Max Kaiser / Bettina Kann
max.kaiser@onb.ac.at
bettina.kann@onb.ac.at

Aufgabe von Literaturarchiven

Quellen zur Literatur

- sammeln
- bewahren
- erschließen
- zugänglich machen

Nachlaß

Regeln zur Erschließung von Nachlässen und Autographen (RNA) – einleitende Richtlinien:
schriftlicher Nachlass: Summe aller Unterlagen –

- Werkmanuskripte und Arbeitspapiere,
- Korrespondenzen,
- Lebensdokumente,
- Sammlungen

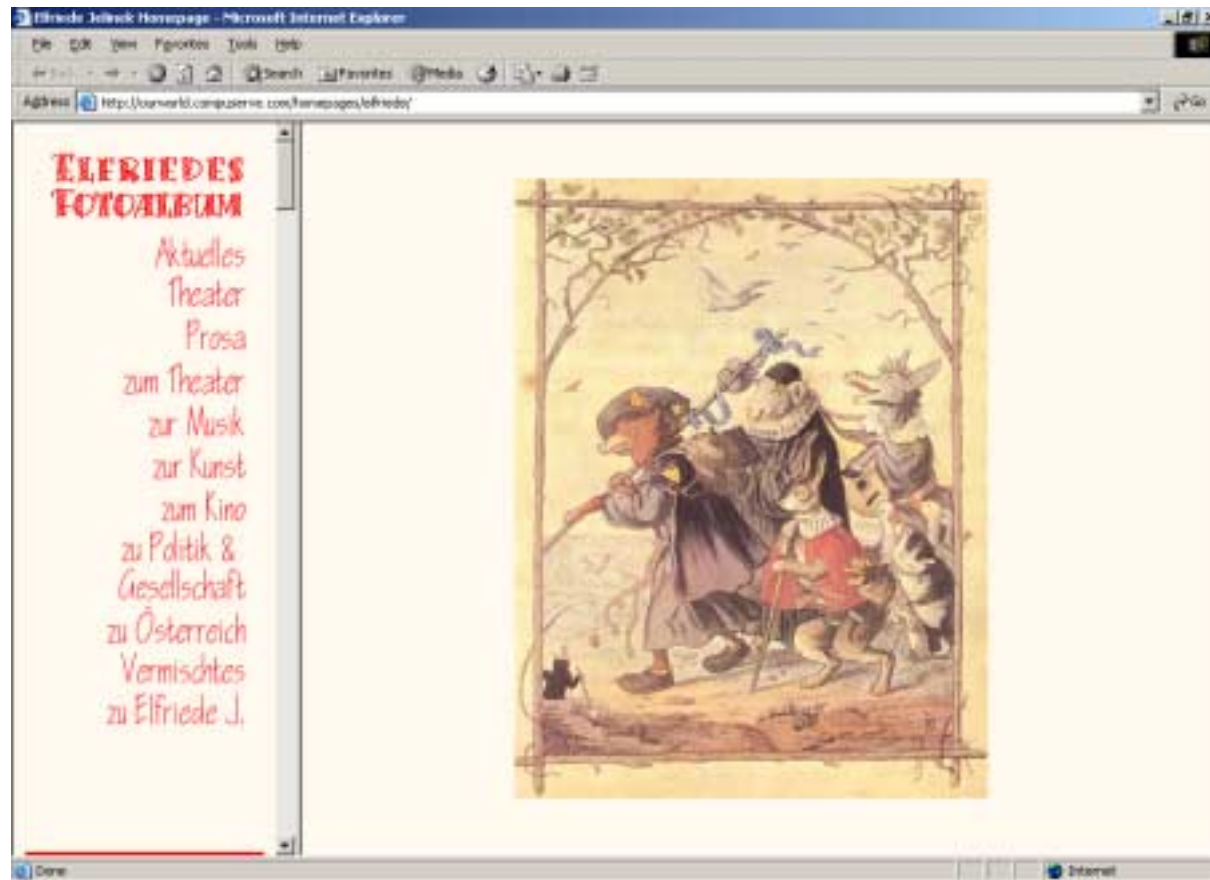


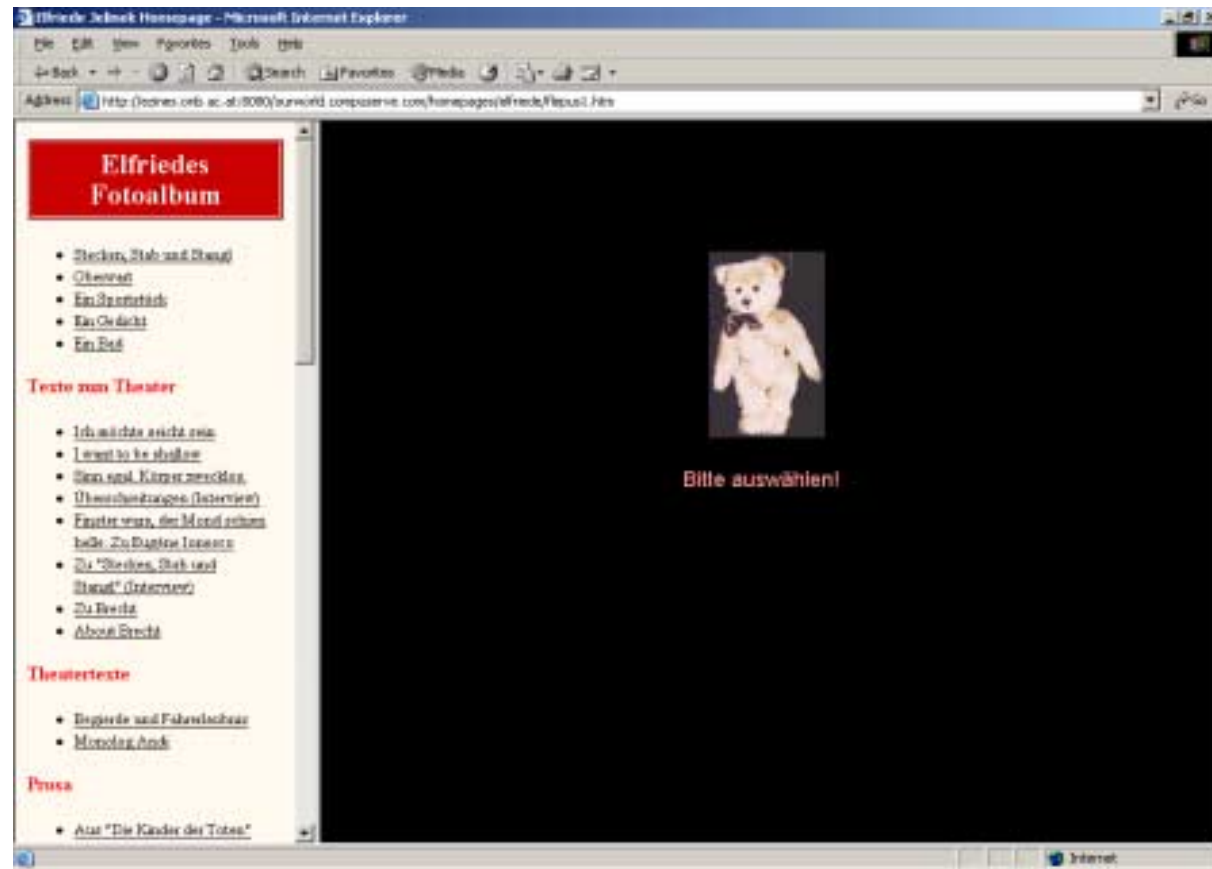






Österreichische Nationalbibliothek





Nachlaß

- Analoge Dokumente
- Digitale Dokumente
 - Auf unterschiedlichen Datenträgern (Festplatten, Floppydisks, Bänder, CD-Rs, DC-RWs, DVD-Rs, Syquest, ZIP, JAZZ, ...)
 - Für unterschiedliche Computerplattformen (IBM, Apple, Atari, ...) und Betriebssysteme (DOS, Windows 95, XP, Linux, Apple OS X, ...)
 - Für unterschiedliche Softwareapplikationen (MS Word, Wordperfect, MS Access, ...)
 - Online Dokumente

„Digitaler Nachlaß“

- Werkmanuskripte und Arbeitspapiere:
 - Dateien unterschiedlicher Textverarbeitungsprogramme auf unterschiedlichen Datenträgern
 - Multimedia-Objekte (Flash-Animationen, interaktive Programme etc.)
 - Homepages, etc.
- Korrespondenzen:
 - Emails in unterschiedlichen Email-Applikationen sowie in Datenbanken:
 - Plain-Text-Emails;
 - HTML-Emails mit Textformatierungen;
 - Emails mit Attachments in unterschiedlichen Datenformaten
- Lebensdokumente / Sachakten:
 - Unterschiedliche digitale Objekte (z.B. digitale Fotosammlungen, Datenbanken etc.)
- Sammlungen
 - Unterschiedliche digitale Objekte

Digitale Surrogate

- Literaturarchive erstellen in Digitalisierungsprojekten auch selbst zunehmend digitale Dokumente
 - meist mit hohem Ressourcenaufwand verbunden
 - es muß sichergestellt werden, daß die Ergebnisse auch in einigen Jahren noch zugänglich sind

Digitale Langzeitarchivierung

- Langfristige Speicherung, Bewahrung und Zugänglichmachung digitaler Dokumente als Konsequenz einer Archivierungsstrategie
- Prozeß, der die langfristige Zugänglichkeit und Benutzbarkeit digitaler Dokumente unter Bewahrung ihrer Integrität, Authentizität und Funktionalität gewährleistet.
- Hardware und Software müssen so auf das digitale Dokument angewandt werden, daß eine integrale, authentische und funktionale Repräsentation generiert wird

Digitales Dokument

1. Physikalische Ebene: Abfolge von physikalischen Zuständen auf einem Datenträger („0 / 1“)
2. Logische Ebene: Kodierung des Dokuments, die für einen Computer interpretierbar ist – abhängig von Computerplattform, Hard- und Software
3. Konzeptionelle Ebene: „Inhalt“ des Dokuments
4. Signifikante Eigenschaften: jene Bestandteile der konzeptionellen Ebene, die für Langzeitarchivierung relevant sind

Vgl.: Guidelines for the Preservation of Digital Heritage. Prepared by the National Library of Australia. Paris: UNESCO – Information Society Division 2003

Risiken für digitale Dokumente

- Dokument geht durch Zerstörung oder Beschädigung des Datenträgers verloren
- Dokument geht verloren, weil die für seine Interpretation notwendige Software oder Hardware nicht mehr vorhanden ist
- Dokument geht verloren, weil die falsche Archivierungsstrategie gewählt wurde

Strategien für digitale Langzeitarchivierung

- Ausarbeitung einer institutionellen Archivierungspolitik für digitale Dokumente
- Proaktive Herangehensweise: Handeln, bevor die Zugänglichkeit der Dokumente ernsthaft gefährdet ist
- Zusammenarbeit mit Autoren
- Informierte Selektion dessen, was archiviert wird
- Anwendung von Standards und „Best Practices“
- Beschreibung der Dokumente durch ausreichende Metadaten
- Schutz des Datenstroms, aus dem das digitale Dokument besteht
- Anwendung geeigneter Verfahren zur Sicherstellung der Zugänglichkeit des digitalen Dokuments (Konversion, Emulation u.a.)
- Archivierung der Objekte in einer sicheren Umgebung
- Zusammenarbeit mit anderen Institutionen

Bewertung / Selektion | 1

- Digitaler Nachlaß: meist große Anzahl disparaten Materials → Bewertung aufwendig
- Bewertung: zeitkritisch – muß um den Erwerbungszeitpunkt erfolgen, sonst deselektiert sich das Material selbst...
- 2 Alternativen:
 1. Alles wird archiviert, weil Selektion zu aufwendig und personalintensiv wäre
 2. Es wird selektiert, um Kosten für Datenspeicherung und –verwaltung zu senken

Bewertung / Selektion | 2

- Selektion für Langzeitarchivierung: muß sich an den signifikanten Eigenschaften des Objekts orientieren
→ oft schwierig zu bestimmen
- Bewertung digitaler Nachlässe: bestandsspezifische und fachwissenschaftlichen Expertise UND spezielle Kenntnisse in digitaler Archivierung
- Digitale Dokumente: oft nicht nach den üblichen Klassifikationsschemata einzuordnen
- Bewertung / Selektion: sollte sich an einer festgelegten institutionellen Archivierungspolitik für digitale Objekte orientieren können

Standards und „Best Practice“

Überblick:

- PADI-Portal der Australischen Nationalbibliothek, <http://www.nla.gov.au/padi/>
- Handbook der Digital Preservation Coalition in Großbritannien, <http://www.dpconline.org/>
- Digital Preservation Testbed, <http://www.digitaleduurzaamheid.nl/>
- NEDLIB-Projekt der EU, <http://www.kb.nl/coop/nedlib/>
- CEDARS-Projekt, <http://www.leeds.ac.uk/cedars/>

PADI *Preserving Access to Digital Information*

PADI is a subject gateway to digital preservation resources

[Home](#)

[About PADI](#)

[Search](#)

[Browse Topics](#)

[Feedback](#)

[Contributions](#)

[What's New?](#)

[Partners](#)

[Working Groups](#)

RESOURCE TYPES

- ▶ [Events](#)
- ▶ [Policies, Strategies & Guidelines](#)
- ▶ [Projects](#)
- ▶ [Organisations & Websites](#)
- ▶ [Bibliographies](#)
- ▶ [Discussion Lists](#)
- ▶ [Glossaries](#)
- ▶ [Journals & Newsletters](#)
- ▶ [News & Discussion padiforum-I](#)
- ▶ [Quarterly Digest](#)

[Current issue](#)

▶ [Safekeeping](#)

Selected resources in PADI are safekept

DIGITAL PRESERVATION TOPICS

- ▶ [General Resources](#)
- ▶ [Issues](#)
- ▶ [Strategies](#)
- ▶ [Rights Management](#)
- ▶ [Data Documentation & Standards](#)
- ▶ [Formats & Media](#)
- ▶ [National Approaches](#)
- ▶ [Digitisation](#)
- ▶ [Digital Records](#)
- ▶ [Digital Libraries](#)
- ▶ [Management](#)

▶ [PADIUpdate](#)

Use PADIUpdate to add resources

PADI is managed by the National Library of Australia with the support of [CLIR](#) and [DPC](#)

Council on Library and Information Resources 

 Digital Preservation Coalition

PADI is an international resource. [Contributions](#) are encouraged.

The PADI site contains links to a wide range of resources which do not belong to the National Library


Digital Preservation Coalition - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Copy Paste

Address <http://www.dpconline.org/graphics/index.html> Go Google

Text Version Members Section



Welcome to the Digital Preservation Coalition website

The Digital Preservation Coalition (DPC) was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.

The contents of this site include:

- About the Coalition:** an introduction to our mission, programme of work, members and how to participate.
- Events:** details of past and forthcoming DPC members forums and events.
- Handbook:** the online edition of 'Preservation Management of Digital Materials - A Handbook by Maggie Jones and Neil Beagrie, now maintained and updated by the DPC.
- What's New** in digital preservation: a regular summary of selected recent activity compiled as a joint service between the DPC and Preserving Access to Digital Information (PADI) gateway.
- Links** to selected external resources and partners of the DPC.
- Contact:** how to contact the DPC

Further information and news on the Coalition will be disseminated via the digital-preservation email list on JISCmail (to subscribe to the list or view its message archive see the Web pages at <http://www.jiscmail.ac.uk/lists/digital-preservation.html>).

Home
About
Events
Handbook
What's New
Reports
Links
Contact



TESTBED DIGITALE BEWARING

The Challenge



≡
Het programma Testbed ▶
Actueel ▶
Kennissbank ▶
Testbed Project ▶
Knowledge Bank ▶

The digital government: it seemed to be so far away in the previous century. Now, in the 21st century, the government is working more and more with digital documents. E-mail communication has become part of the daily routine and databases are used everywhere. The government has an obligation to treat digital government information in a responsible manner.

Digital documents must be preserved and remain accessible for coming generations. This principle also applies to paper-based information that is managed and preserved. Building the digital government means that the appropriate digital infrastructure needs to be in place as soon as possible. Records not only have to be found quickly, they also have to be authentic and readable (regardless of the current technology) and remain so in the future.

Commissioned by the Ministry of the Interior and Kingdom Relations and the States Archives' Service, the Digital Preservation Testbed was established in 2001. The goal of the Testbed is to ensure the lasting accessibility and reliability of government information in the digital era, now and in the future. The constant technological changes present the biggest problems when it comes to preserving government information: new computer systems, new versions of software, etc. The Digital Preservation Testbed is researching approaches to overcome these problems and helping to build the digital government.



[FAQ](#)

[News](#)

[Results](#)

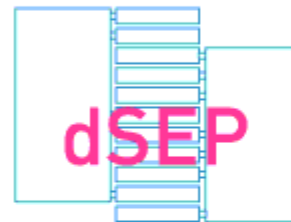
[Workshop](#)

[Links](#)

[Discussion](#)

[Calendar](#)

Preserving Our Digital Heritage



[click here to view the Interactive dSEP Workflow](#)

**Networked
European
Deposit
Library**

NEDLIB is a collaborative project of European national libraries. It aims to construct the basic infrastructure upon which a networked European deposit library can be built. The objectives of Nedlib concur with the mission of national deposit libraries to ensure that electronic publications of the present can be used now and in the future.

Cedars

curl exemplars in digital archives

[Cedars Home](#)

[Project Overview](#)

[Cedars Administration](#)

[Publications & Conferences](#)

[Collection Management](#)

[Technical Strategies](#)

[Distributed Archive Prototype](#)

Publications and Conferences

Papers and articles prepared by the Cedars and CAMiLEON project teams and associates. Conferences, workshops and seminars hosted by Cedars.

Guidance Documents

The Cedars Guidance documents provide a concise, accessible synthesis of the relevant work of Cedars and related effort in five discrete areas. Each Guide includes a section on costs. They are aimed primarily at librarians working in research institutions.

[IPR](#)

[Preservation Metadata](#)

[Collection Management](#)

[Digi Preservation Strategies](#)

[Digital Archiving Prototype](#)

Cedars Publications

[Cedars Working Papers](#)

[Articles and Papers by Cedars and CAMiLEON](#)

[Cedars Conferences, Seminars and Workshops](#)

Standards - OAIS

- CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. January 2002. - ISO 14721:2002
http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html
- Generisches Modell für die Organisation eines digitalen Archivs
- Identifiziert die für die Langzeitarchivierung erforderlichen Komponenten und Prozesse eines Archivs
- Entwickelt eine Terminologie für digitale Objekte und die mit ihnen assoziierten Metadaten

Metadaten

- Technische und administrative Metadaten: Auch hier kein einheitlicher Standard verfügbar
- Wichtige Entwürfe für Metadatenschemata:
 - Australische Nationalbibliothek,
<http://www.nla.gov.au/preserve/pmeta.html>
 - CEDARS-Projekt,
<http://www.leeds.ac.uk/cedars/guideto/metadata/>
 - NEDLIB-Projekt,
<http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>
 - OCLC/RLG Working Group on Preservation Metadata,
<http://www.oclc.org/research/pmwg/>



Home
About OCLC
News

Support
Services and Databases
Contacts

What's New
Librarian's Toolbox
Site Map

Search OCLC
 GO

◀ OCLC Research home ▶ Projects ▶ Preservation Metadata Working Group

OCLC Research

OCLC Research home About People **Projects** Programs Publications Software

Preservation Metadata Working Groups

- Background
- Current Membership
- Working Group I: Preservation Metadata Framework
- Working Group II: Implementation Strategies
- Related Links
- Contact

Preservation Metadata for Digital Materials

What is preservation metadata?

Preservation metadata:

- is the information necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital materials.
- includes information ranging from a description of the hardware/software environment needed to render a particular class of digital object to a record of the migration of a digital object through successive formats over time.
- is an essential component of most digital archiving strategies.

Why is it important?

Information in digital form is notoriously fragile, suffering from the twin perils of bit decay and technological obsolescence.

The threshold in a digital object's lifecycle beyond which the imperatives of preservation must be confronted arrives relatively early. In the extreme it can be concurrent with the object's creation.

Archiving practice has long recognized the importance of maintaining careful records of the preservation process.



Home
About OCLC
News

Support
Services and Databases
Contacts

What's New
Librarian's Toolbox
Site Map

Search OCLC
 GO

◀ OCLC Research home ▶ Projects ▶ Preservation Metadata Working Group

OCLC Research

OCLC Research home About People **Projects** Programs Publications Software

Preservation Metadata Working Groups

- Background
- Current Membership
- Working Group I:
Preservation
Metadata
Framework
- Working Group II:
Implementation
Strategies
- Related Links
- Contact

Working Group II: Implementation Strategies

PREMIS (PREservation Metadata: Implementation Strategies)

The [OCLC Digital & Preservation Cooperative](#), the [OCLC Office of Research](#), and [RLG](#) are pleased to sponsor the formation of a working group aimed at the development of recommendations and best practices for implementing [preservation metadata](#).

Working Group Charge

In 2001-02, OCLC and RLG convened [an expert working group](#), reflecting a variety of institutional and geographical backgrounds, to consider the types of information falling within the scope of preservation metadata. The group's efforts culminated in the public release of [A Metadata Framework to Support the Preservation of Digital Objects](#) (PDF:696K/54pp.), a comprehensive metadata framework describing the information necessary to carry out, document, and evaluate digital preservation processes.

Using the metadata framework as a conceptual foundation and starting point for its work, the PREMIS working group will carry forward the activities of the first working group. The focus of the PREMIS group will be on the practical aspects of implementing preservation metadata in digital preservation systems.

Objectives

The working group's objectives include:

- An implementable set of "core" preservation metadata elements, with broad applicability within the digital preservation community

Schutz des Datenstroms

- Ziel der digitalen Langzeitarchivierung: Erhaltung der konzeptionellen Ebene des digitalen Dokuments
Voraussetzung: Schutz der physischen Ebene („Bits und Bytes“)
→ etablierte Strategien im IT-Bereich
- Backup
- Redundanz (Datenredundanz / Datenträgerredundanz / Ortsredundanz)
- „Media Refreshing“
- Datenmigration

CD-Rs

- Benutzbarkeit einer leeren CD-R: 5–10 Jahre
- Lebenswartung einer beschriebenen CD-R: 50–100 Jahre?
→ in der Praxis oft bereits nach 5–10 Jahren Probleme
- CD-Rs sollten nicht mit einfacher Geschwindigkeit (1x) gebrannt werden
- CD-Rs sollten nicht beschriftet werden: Farbstoff kann Reflexions- und Farbschicht (dye) der CD-R beschädigen
- Etiketten sollten nicht verwendet werden: können die Oberfläche der CD-R beschädigen
- CD-Rs sollten ausschließlich in Kunststoffboxen aufbewahrt werden
- CD-Rs sollten nicht direktem Sonnenlicht, hoher Temperatur oder hoher Luftfeuchtigkeit ausgesetzt werden.
- Nähere Details: ISO 18925:2002 Imaging materials - Optical disc media - Storage practices

Fortsetzung:
Teil 2 (Bettina Kann)