

Vorträge zur FIS Bildung-Herbsttagung vom 07.12.2004 zum Thema
**„Elektronische Volltexte - Konzepte des überregionalen Zugriffs und
Verfahren zur Langzeitverfügbarkeit“**

1. Hans Liegmann, Die Deutsche Bibliothek (DDB): Langzeiterhaltung digitaler Ressourcen
2. Projekt NESTOR - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland
3. Reinhard Altenhöner, Die Deutsche Bibliothek (DDB): Projekt KOPAL - Kooperativer Aufbau eines Langzeitarchivs Digitaler Informationen



Langzeiterhaltung Digitaler Ressourcen

Deutsches Institut für
Internationale Pädagogische Forschung
Frankfurt, 7.12.2004

Hans Liegmann, Die Deutsche Bibliothek

Überblick

- Begriffsklärung
- Die Probleme und Wege zur Lösung
- Das Referenzmodell OAIS
- Metadaten zur Langzeiterhaltung
- Wege in das Depotsystem
- Metadaten

Begriffsklärung

Begriffsklärung - 1

Langzeit

Langzeitarchivierung

Langzeiterhaltung

Langzeitverfügbarkeit

Begriffsklärung - 2

„Digitale Ressource“

- CD-ROM, DVD, Diskette
- elektronische Zeitschrift (eJournal)
- Online-Hochschulschrift
- Homepage, Website, Blog
- eBook
- Datenbank, Datenpools von Suchmaschinen
- eMail-Newsletter, Mailingliste
- digitalisierte, originär analoge Ressourcen
- Instrumentalsoftware, Betriebssoftware
- Computerspieletc.....etc.....

Begriffsklärung - 3

bibliographisch

„monographisch“ - „fortlaufend“

technisch

„statisch“ - „dynamisch“

Begriffsklärung - 4

Depotsystem – Archivsystem

vs.

Dokumentserver

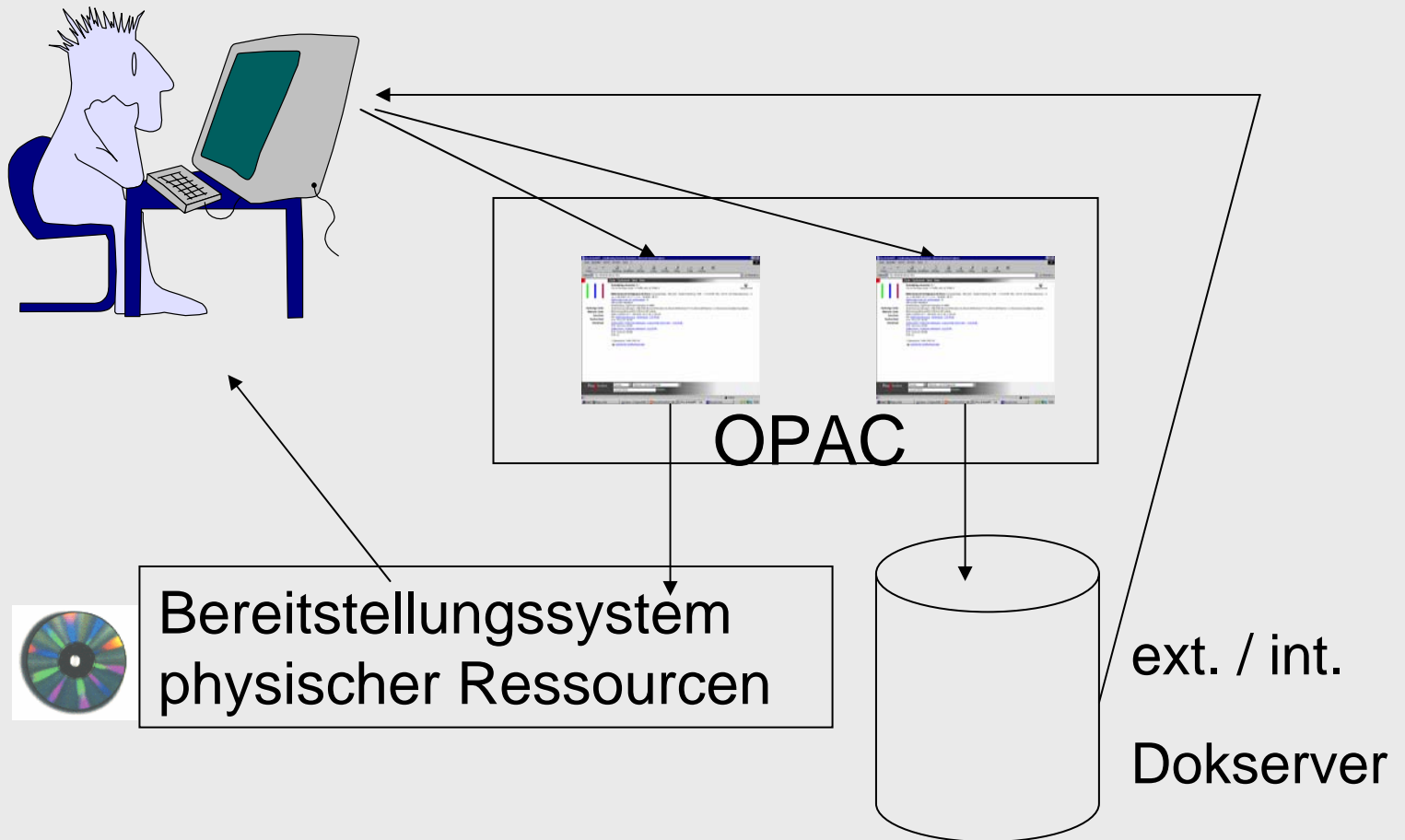
vertrauenswürdigen [digitales] Archiv



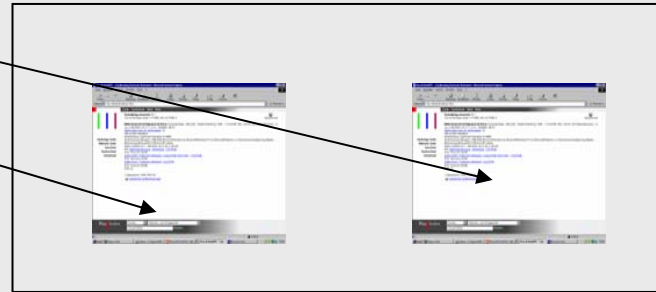
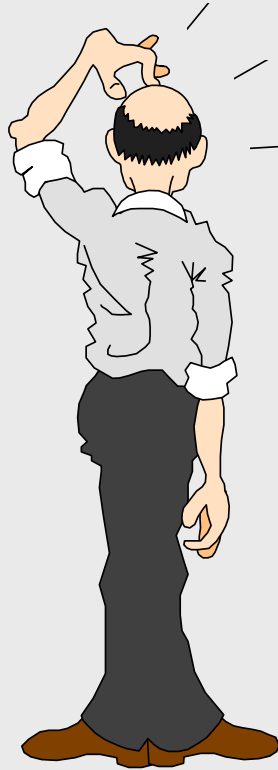
Die Probleme

Wege zu ihrer Lösung

Gegenwart

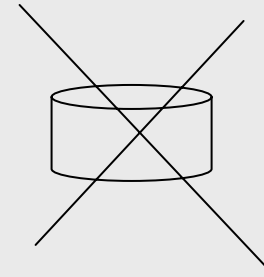
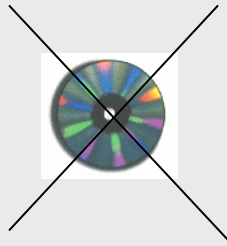


Zukunft



WINDOWS95?
intelPentium?
ISO9660?

PDF? HTML?
XML? JAVA?





Die Probleme

- Substanz ist gefährdet
- Nutzungsumgebungen verschwinden
- Bindung an proprietäre Umgebungen

Substanzerhaltung (1)

- Grundlage für weitere Bemühungen
- Datenträgerstabilität ?
- Kopieren (refreshing)
- Datenträgermigration
z.B. magnetisch --> optisch
- Katastrophenmanagement

Substanzerhaltung (2)

Störfaktoren:

Mengenproblem

➔ Automatisierung

unerwartetes Materialverhalten

➔ laufende Materialbeobachtung

Fehler bei Kopiervorgängen

➔ strikte Prozesskontrolle

Kopierschutzverfahren

➔ bislang keine Lösung!!



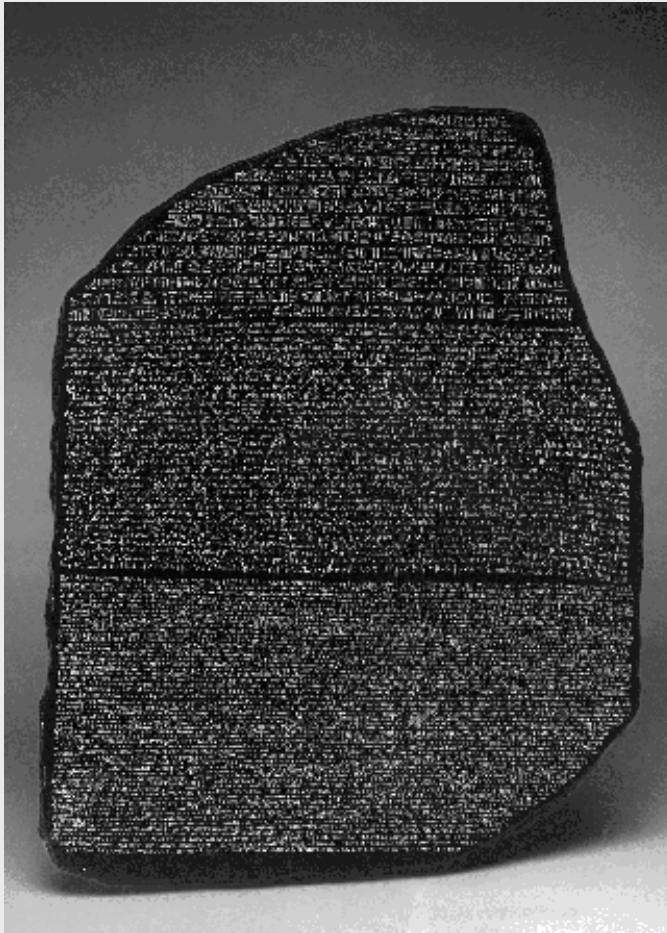
Erhaltung der Benutzbarkeit (1)

Die Substanz, die wir erhalten, muss interpretiert werden können. Was müssen wir dafür tun?

11011101110011110011010010101011
10101110010101011100011010101010
10001101010101010101010100010101
01010101010101010101010100010101
01010101010100111000001011111100



Erhaltung der Benutzbarkeit (2)



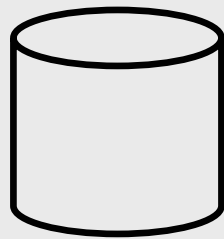
Rosetta Stone –
Der Stein von Rosette
- Raschid

* 196 v.Chr.
gefunden 1799

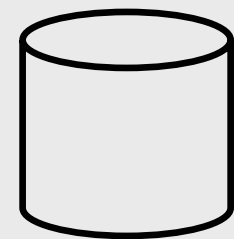
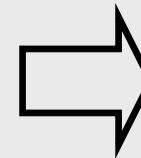
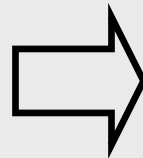
British Museum,
London

Technische Voraussetzung

TRANSFERIERBARKEIT



Originalserver



Langzeitarchiv



Wege zur Langzeitverfügbarkeit

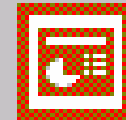
Migration

Emulation

STANDARDISIERUNG

Migration (1)

Diese Präsentation wird aus einer früheren PowerPoint-Version aktualisiert.



Datei: C:\TEMP\MMB-Präsentation.ppt

Folien fertig konvertiert: 7 von 15

Abbrechen

Migration (2)

Vorteile:

- bewährte Strategie
- „handwerkliche Tätigkeit“

Nachteile:

- Fass ohne Boden
- Anwendbarkeit hat Grenzen
- hohes Risiko der Datenverfälschung



Emulation (1)



Emulation (2)

Vorteile:

- EP bleiben unversehrt erhalten
- Aufwendungen sind kalkulierbar
- für alle EP anwendbar

Nachteile:

- hohe Investitionen in F&E erforderlich
- Interesse des Marktes ist begrenzt

Standardisierung

z.B. OpenEbook

z.B. Strukturvorgaben für
Dissertationen

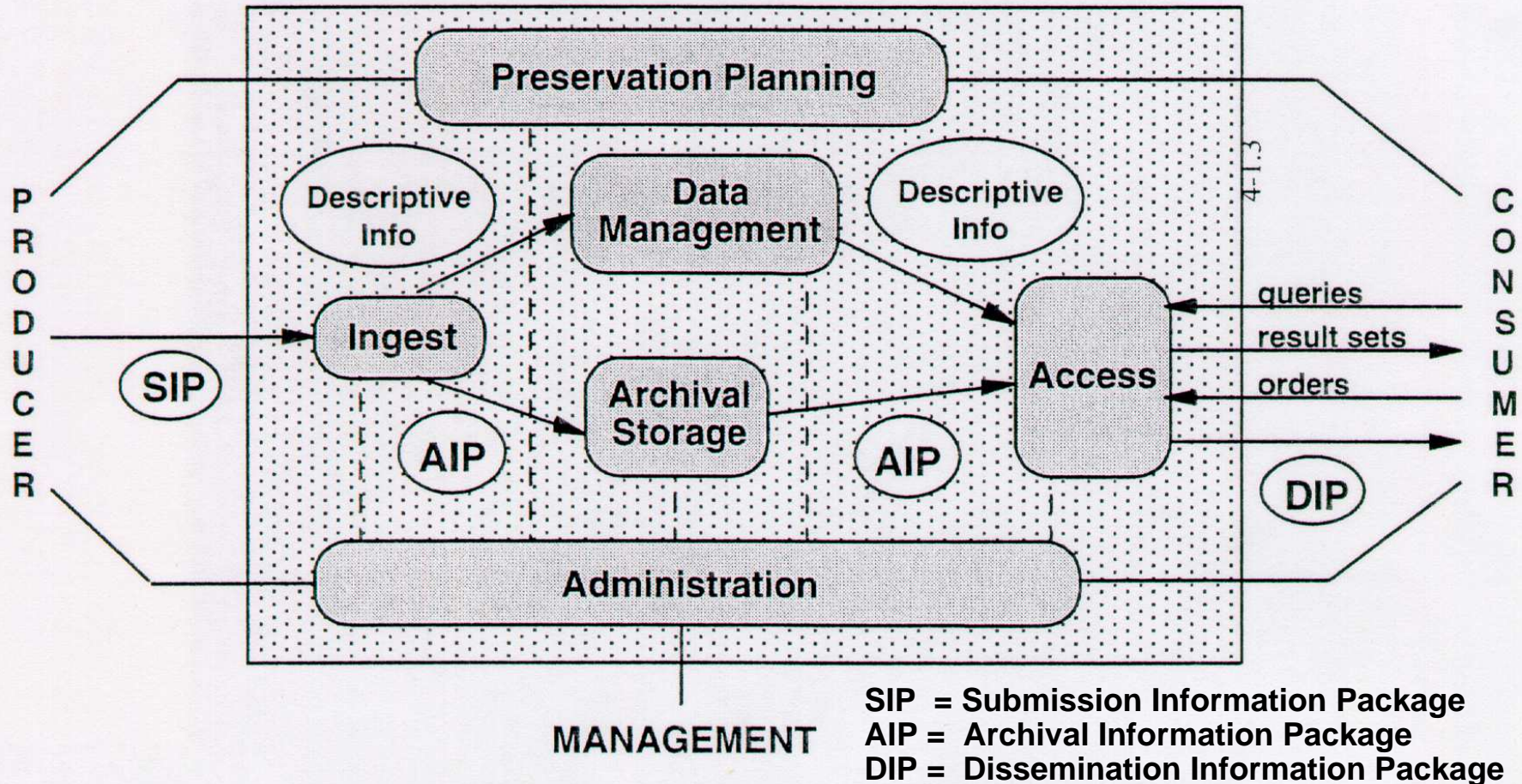
z.B. Container für Mehrdateien-
Dissertationen

z.B. OpenEjournal ?

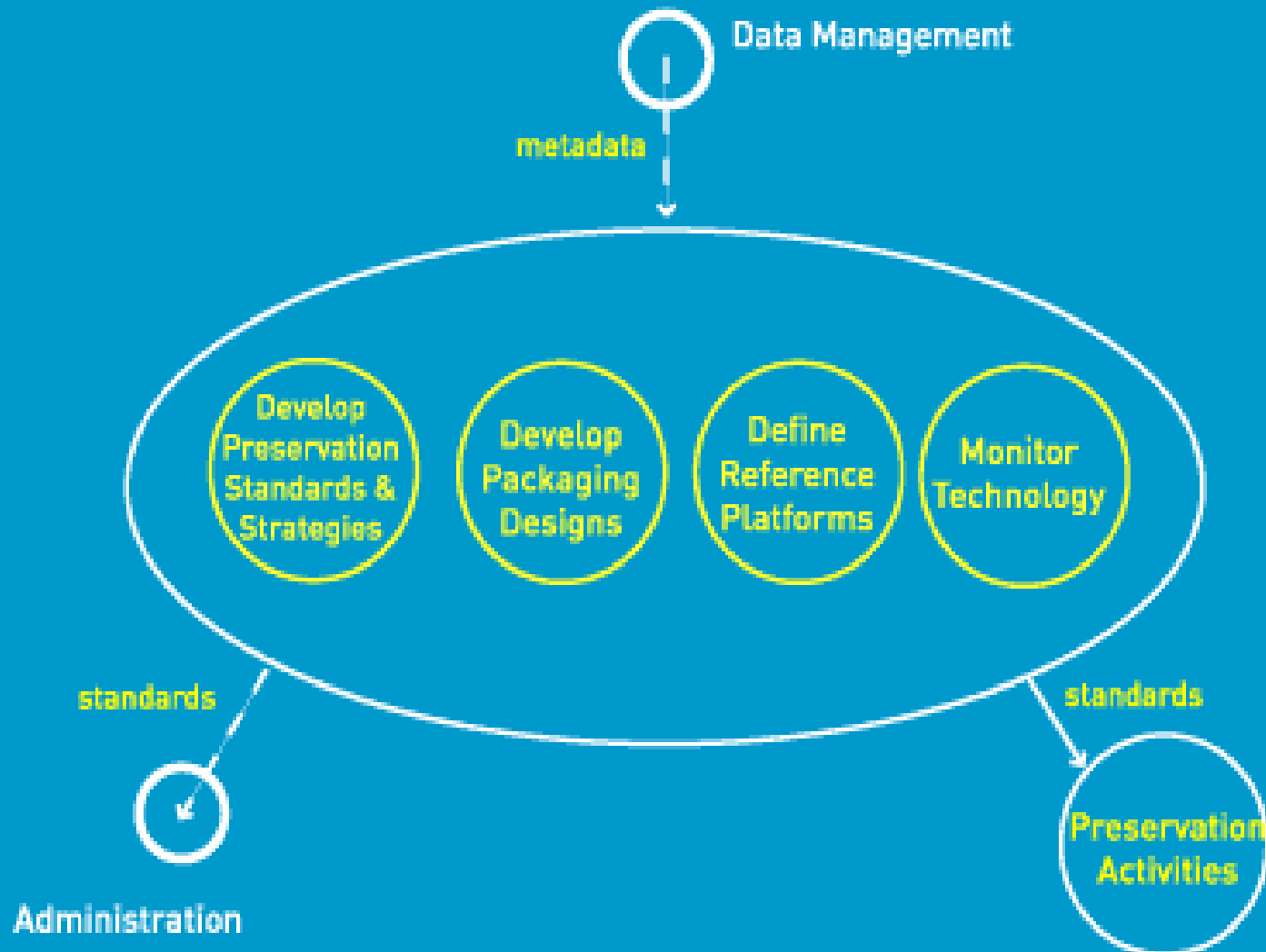
Das Referenzmodell OAIS



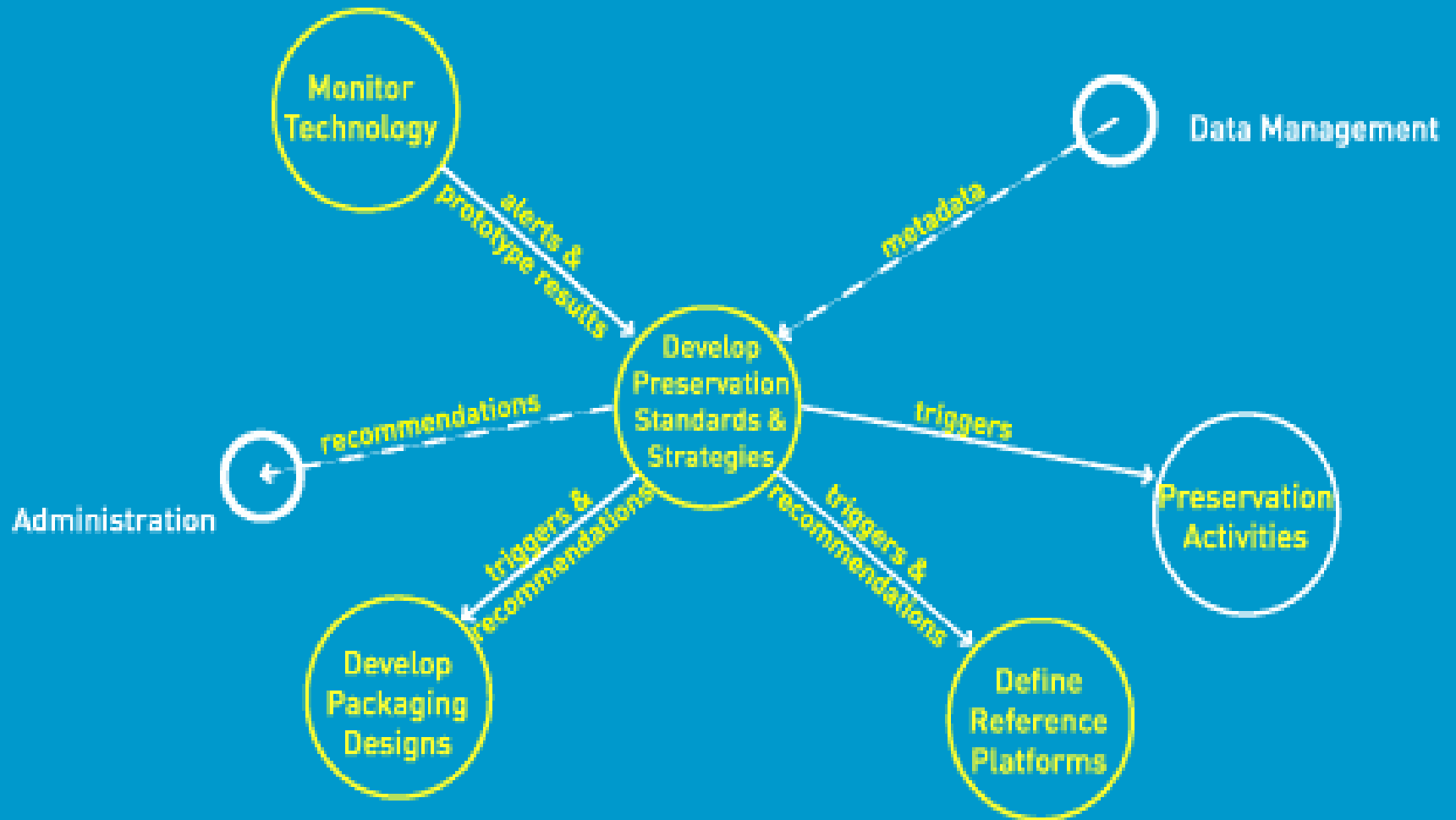
Open Archival Information System - OAIS



Preservation: Sub-process: **Preservation Planning**



Preservation: Sub-process: **Develop Preservation Standards & Strategies**



Metadaten zur Langzeiterhaltung

Metadaten - Kategorien

administrative M.

prozesssteuernde M.

beschreibende M.

technische M. (Langzeiterh.)

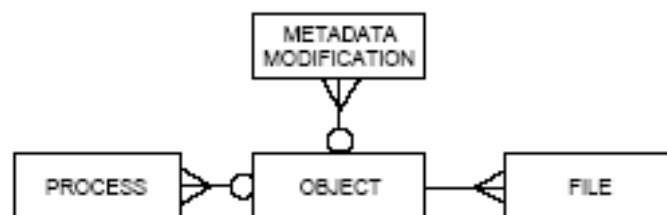
HTML bzw. XML

Appendix 1 – Preservation Metadata ModelENTITY 1 - OBJECT

- 1.1 Name of object
- 1.2 Reference number
- 1.3 Identifier - Object IID
- 1.4 Persistent identifier - PID
- 1.5 UNIX location
- 1.6 Date of creation of preservation master
- 1.7 Technical composition
- 1.8 Structural type
- 1.9 Hardware environment
- 1.10 Software environment
- 1.11 Installation requirements
- 1.12 Access inhibitors
- 1.13 Access facilitators
- 1.14 Quirks
- 1.15 Authentication
- 1.16 Metadata record creator
- 1.17 Date of metadata record creation
- 1.18 Comments

ENTITY 2 - PROCESS

- 2.1 Object IID
- 2.2 Process
- 2.3 Purpose
- 2.4 Business unit or agency
- 2.5 Permission
- 2.6 Date of permission
- 2.7 Hardware used
- 2.8 Software used
- 2.9 Steps
- 2.10 Result
- 2.11 Guidelines
- 2.12 Completion date and time

ENTITY 3 - FILECommon elements:

- 3.1 Object IID
- 3.2 File IID
- 3.3 Structural context
- 3.4 Filename and extension
- 3.5 File size
- 3.6 File date and time
- 3.7 MIME type/format
- 3.8 Version
- 3.9 Target indicator

3.10 Image:

- 3.10.1 Resolution
- 3.10.2 Dimensions
- 3.10.3 Tonal resolution
- 3.10.4 Colour space
- 3.10.5 Colour management
- 3.10.6 Colour lookup table
- 3.10.7 Orientation
- 3.10.8 Compression

3.11 Audio:

- 3.11.1 Resolution
- 3.11.2 Duration
- 3.11.3 Bit rate
- 3.11.4 Compression
- 3.11.5 Encapsulation
- 3.11.6 Track number and type

ENTITY 3 – FILE (cont.)3.12 Video:

- 3.12.1 Frame dimensions
- 3.12.2 Duration
- 3.12.3 Frame rate
- 3.12.4 Compression
- 3.12.5 Encoding structure
- 3.12.6 Sound

3.13 Text:

- 3.13.1 Compression
- 3.13.2 Character set
- 3.13.3 Associated DTD
- 3.13.4 Structural divisions

3.14 Datasets:

Uses common elements only

3.15 System Files:

Uses common elements only

ENTITY 4 - METADATA MODIFICATION

- 4.1 Object IID
- 4.2 Modifier
- 4.3 Date and time
- 4.4 Field modified
- 4.5 Data modified




METS Metadata Encoding & Transmission Standard

Official Web Site



The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the [XML schema language](#) of the [World Wide Web Consortium](#). The standard is maintained in the [Network Development and MARC Standards Office](#) of the Library of Congress, and is being developed as an initiative of the [Digital Library Federation](#).



Sammlung digitaler Ressourcen

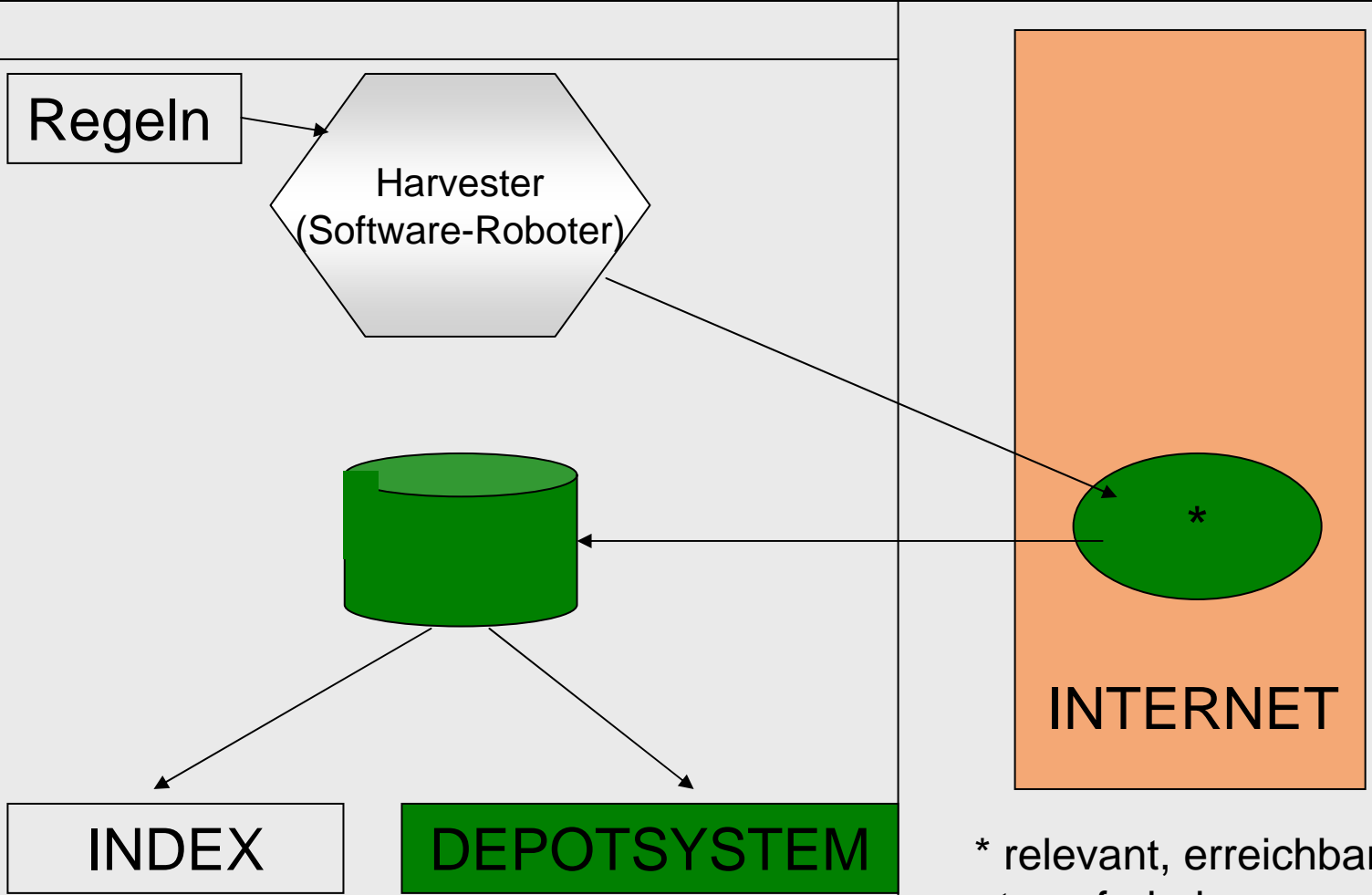
selektiv - objektbezogen

Einrichtung von Lieferwegen zwischen
Produzenten und dem Archiv

„vollständig“ – protokollabhängig

Web-Harvesting

Web-Harvesting



* relevant, erreichbar, transferierbar

dbib

Vor- und Nachteile der Sammelmethoden

objektbezogene Sammlung

- V: hohe Qualität
- V: Objekt "Publikation"
- N: schwieriger automatisierbar
- N: Mitwirkung der Ablieferer

Harvesting

- V: vollständig automatisierbar
- V: beeindruckende Mengen
- N: begrenzte Abdeckung, Qualität?
- N: Objekt "Datei"

nun zu nestor ...

nestor

Kompetenznetzwerk Langzeitarchivierung
und Langzeitverfügbarkeit Digitaler
Ressourcen für Deutschland

**Network of Expertise in Long-Term
Storage of Online Resources**

Hans Liegmann, Projekt nestor

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

nestor 2003 - 2006

Fördervolumen ca. EUR 800.000

- Die Deutsche Bibliothek
- Bayerische Staatsbibliothek, München
- Niedersächsische Staats- und Universitätsbibliothek Göttingen
- Computer- und Medienservice der Humboldt-Universität, Berlin
- Generaldirektion der Bayerischen Staatlichen Archive
- Institut für Museumskunde, Berlin



S M
B Institut für Museumskunde

Ziele (1)

- Informations- und Kommunikationsplattform für alle Aktivitäten auf dem Gebiet der Langzeiterhaltung in Deutschland
- Entwicklung eines eines kooperativen Konzepts zur bereichsübergreifenden Zusammenarbeit
- Förderung von Synergien zwischen deutschen und internationalen Vorhaben zur Langzeiterhaltung digitaler Ressourcen

Ziele (2)

- Entwurf einer dauerhaften Organisationsform für das „Kompetenznetzwerk Langzeitarchivierung“



nestor

www.langzeitarchivierung.de

Aufgabe 1

Informationen
verfügbar machen

Informationen verfügbar machen

- www.langzeitarchivierung.de - Bündelung deutschsprachiger Informationsquellen zum Aufgabenfeld Langzeitarchivierung
- Brücke zu Informationsquellen in anderen Sprachen (PADI)
- Newsletter
- Reviews
- Kalender

Aufgabe 2

Kompetenzen sichtbar
und nachnutzbar machen

Kompetenzen sichtbar und nachnutzbar machen

- Maßnahmen, um verteilte Kompetenzen (Organisationen, Institutionen, kommerzielle Unternehmen, Individuen) zusammenzuführen
- Diskussionslisten
- Diskussionsforen
- Workshops
- Übersichten, Statistiken

Aufgabe 3

Diskussion über
Langzeitarchivierung, sowie
deren Strategien und technische
Methoden anregen

Diskussion anregen

- einschlägige deutsche Projekte bzw. Projekte mit deutscher Beteiligung miteinander in Verbindung bringen
- Wissenstransfer mit dem Ausland
→ KOPAL, ARCHISIG, SUNCoE, ...
← DPC, NDIIP, PANDORA, PRONOM, IIPC ...
- „Kontaktpunkt“ für Aktivitäten im Ausland

Ergebnisse von nestor

- Kriterien für vertrauenswürdige digitale Archive und deren Zertifizierung
- Analyse von Sammelrichtlinien, Konzepte zur Arbeitsteilung
- Grundsätze der Langzeitarchivierung
- Rechtliche Aspekte der Langzeitarchivierung
- Perspektiven der Langzeitarchivierung multimedialer Objekte
- Langzeitarchivierung von eJournals

... weitere Ergebnisse ...

- Vergleich von Archivierungssystemen
- Beschreibungsprofil für eine nationale Strategie zur Langzeiterhaltung digitaler Ressourcen
- Digitalisierung und Erhalt von Digitalisaten in deutschen Museen
- Erfordernisse zur Archivierung von Forschungs- und Rohdaten

Vision und Ziel

- www.langzeitarchivierung.de als deutschsprachiges subject gateway im Verbund mit internationalen Partnern
- ein Kompetenznetzwerk als stabile Interessenvertretung für die Belange der Langzeitarchivierung in Deutschland
- Entwicklungsprojekte werden nicht isoliert voneinander betrieben, sondern innerhalb einer "community", die nestor aufgebaut hat



"Who cares to praise his enemy ?" [Nestor. Ovid, *Metamorphoses* 12.548]

"I seek no other vengeance than to ignore Heracles' mighty deeds" [Nestor. Ovid, *Metamorphoses* 12.575]

"A lie will Nestor not utter, for he is wise indeed." [[Athena](#). Homer, *Odyssey* 3.20]

[Achilles](#): The prize I am giving you has no relation to the sports.
Nestor: Now, I must leave this sort of thing to younger men and take the painful lessons of [Old Age](#) to heart." [Homer, *Odyssey* 23.615ff.]

<http://homepage.mac.com/cparada/GML/Nestor.html>



nestor

www.langzeitarchivierung.de

Vielen Dank für Ihre
Aufmerksamkeit



Projekt KOPAL – Kooperativer Aufbau eines Langzeitarchivs Digitaler Informationen

Deutsches Institut für internationale
pädagogische Forschung
Frankfurt, den 7. 12. 2004

Reinhard Altenhöner, Die Deutsche Bibliothek

Agenda

- I. Ausgangssituation: Die Deutsche Bibliothek und ihre Aktivitäten im Kontext Elektronischer Publikationen
- II. Das Projekt KOPAL
- III. Ausblick

Die Deutschen Bibliothek und ihre 3 Standorte

Deutsche Bibliothek Frankfurt am Main



Deutsche Bücherei Leipzig



Deutsches Musikarchiv Berlin



Sammeln und Archivieren

- In Deutschland veröffentlichte Publikationen seit 1913 (auf physischen Trägern)
- Deutschsprachige Publikationen generell
- Übersetzungen deutschsprachiger Publikationen in andere Sprachen seit 1913
- Fremdsprachige Publikationen über Deutschland veröffentlicht seit 1913
- Werke und Dokumente von und über deutschsprachige Emigranten zwischen 1933 und 1945

Und immer weiter so?



Aktivitäten der DDB 1998ff.

- Exilpresse digital
- E-Zeitschriften (z.B. Springer)
- Netzpublikationen
 - auf freiwilliger Basis, in Kooperation mit Verlegern/Ablieferern
- Online-Dissertationen
 - in enger Kooperation mit Universitäten
- Transfer der Daten von Diskette und Altsystemen
 - Amiga-, Atari-, Commodore Applikationen
- URN-Resolver
- LZA-RegBib

Die Deutsche Bibliothek

Suche

Zeitschrift

Acht-Uhr-Abendblatt

Jahrgang 2

Ausgabe 6

Seite

1

Ansicht 50%

Anzeigen

+ Schlagwortkatalog "Aufbau"

+ Copyright-Hinweis

+ Informationen zur Realisierung

Renate Seib

Copyright: Die Deutsche Bibliothek Annette Siegenthaler / 12.12.2003



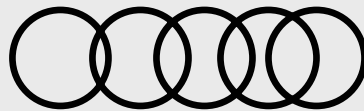
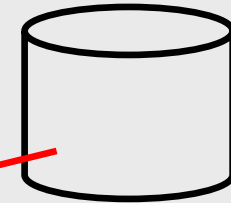
Exilpresse digital
Deutsche Exilzeitschriften 1933-1945

Renate Seib

Copyright: Die Deutsche Bibliothek Guido Pelzer, Annette Siegenthaler / 12.10.2000

Archivierung SpringerLINK

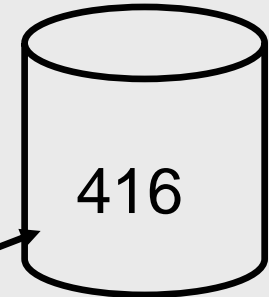
<http://link.springer.de>
(~500 journals)



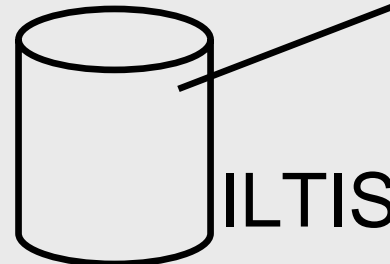
updates auf Träger



aktualisierte Navigation



deposit.ddb.de



ddb


[SPRINGER](#) [FAQ](#)

SEARCH FOR

RETURN

Articles



GO

[ABOUT](#)[BROWSE](#)[FAVORITES](#)[ALERT](#)[ORDERS](#)[LOG OFF](#)[> Home /](#)

What's New on SpringerLink?

No Boundaries. No Waiting.

Pay-per-view now available on SpringerLink.

Due to an overwhelming demand from the research community, SpringerLink now offers a pay-per-view option for journal and some bookseries articles. This option opens the door to expanded research and removes the boundaries non-subscribers often face. Pay-per-view is available to registered users of SpringerLink. Customers use secure or standard purchasing and will then be able to view purchased articles anytime through their Order History.

Browse and Explore

With a collection of journals and book series that account for over 300,000 documents in SpringerLink, the browse and explore functions help users to get quickly to the information and titles they need.

[Browse Publications](#)
[Browse by Online Libraries](#)
[Browse Publishers](#)

New Publications in SpringerLink

SpringerLink welcomes the following new titles in full-text that have recently been added to our collection of online publications.

[The Chinese-German Journal of Clinical Oncology](#)
[Chromatographia](#)

You are not logged in.

To use the personalized features of this site, such as Favorites or Table Of Contents Alerting, you should login or register as an individual user.

Login

[Login via Athens](#)

If you are already registered, please login. If you are not sure, we can

[help.](#)

Username:

Password:

 [Remember my login.](#)

Guest

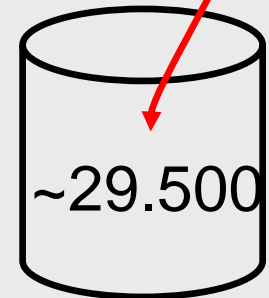
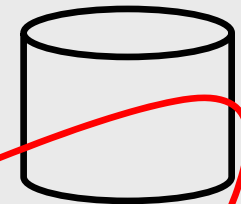
If you do not have a login and do not wish to register, enter as a guest.

Online-Hochschulschriften bei DDB

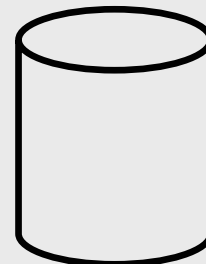
Dt. Universitäten (~ 80)

Orig.-Server

Metadaten
Anlieferung



deposit.ddb.de




Bibliothekssystem

Online-Dissertation via OPAC, Zugang über URN

Deutsche Bibliothek Frankfurt am Main

Suchen UND Standardnummer <num> ?

Sortiert nach Erscheinungsjahr

 Titel: Nitric oxide in the olfactory epithelium [Elektronische Ressource] / von Oliver Schmachtenberg

Verfasser: [Schmachtenberg, Oliver](#)

Erscheinungsjahr: 2001

Hochschulschrift: Hannover, Univ., Diss., 2001

Persistent Identifier: [urn:nbn:de:gbv:089-3321752945](http://nbn-resolving.org/urn:nbn:de:gbv:089-3321752945)

URL: [Archivserver Der Deutschen Bibliothek](#)

Sachgruppe: 33 Medizin ; 32 Biologie



<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:gbv:089-3321752945>

Datentransfer von instabil gewordenen Trägern

z.B. C64 / AMIGA



Original System 1988

migration

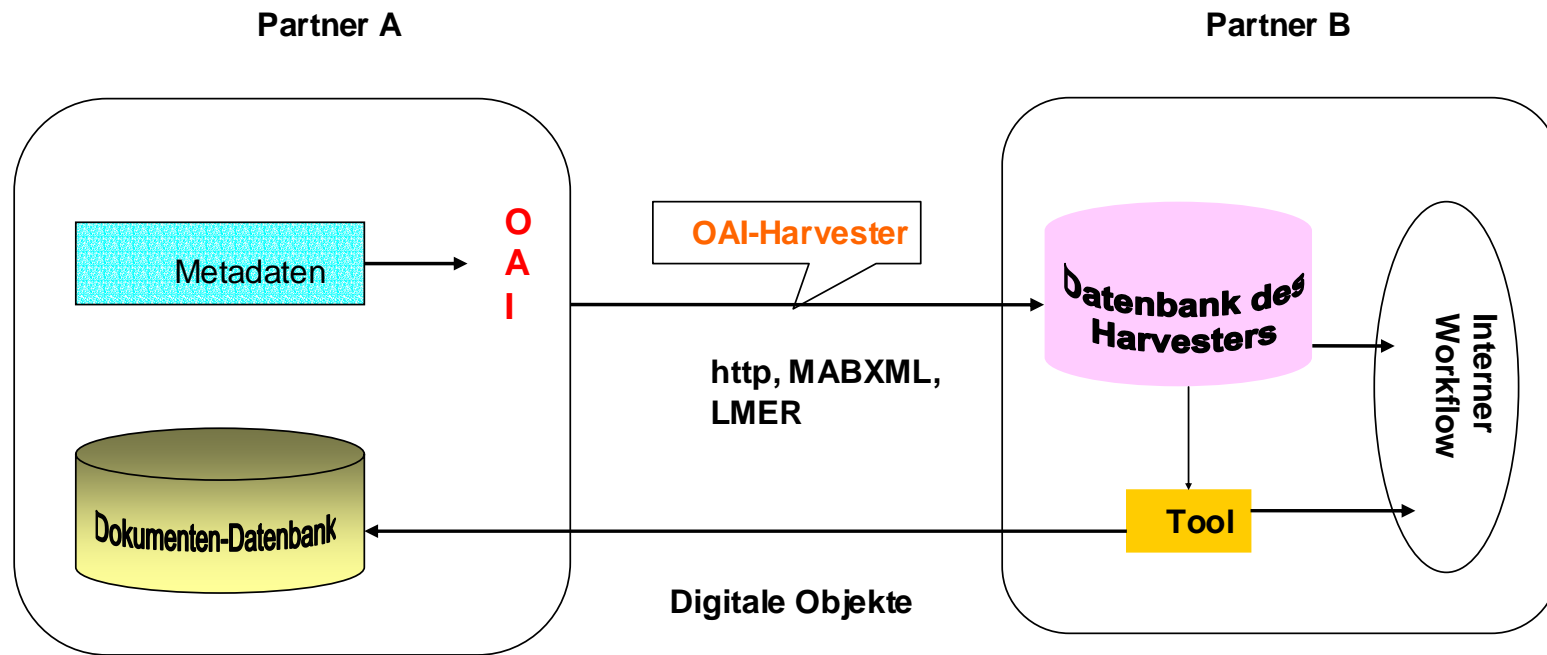
Ziel System 2002

Emulation OS

img



Technisches Modell LZA-RegBib



Austausch von Metadaten und Objekten via OAI in der Kooperation LZARegBib

Ausgangssituation / Rahmen

- Rasch wachsender elektronischer Online-Publikationssektor (E-only, born digital)
 - Digitales Publizieren bekommt zunehmend eine eigene Qualität (Multimedia, Hyperlinking, lebende Dokumente...)
 - Zunehmender Teil kultureller Aktivitäten findet im Netz statt
 - Der Erhalt dieses „Nachlass“ ist – demnächst auch gesetzlich fixiert – Aufgabe von Nationalbibliotheken und vergleichbaren Einrichtungen („Gedächtnisorganisationen“)
 - Aktivitäten dazu haben begonnen, Sammelverfahren werden etabliert
- ⚡ Generelle Anmerkung: Bislang aufgebaute digitale Archive entsprechen überwiegend nicht den Anforderungen, die an „trusted digital repositories“ für die Sicherung der Langzeitverfügbarkeit gerichtet werden



II. Das Projekt KOPAL

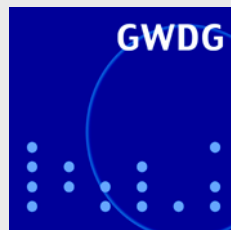
KOoperativer Aufbau eines Langzeitarchivs Digitaler Informationen

GEFÖRDERT VON



Bundesministerium
für Bildung
und Forschung

ddb



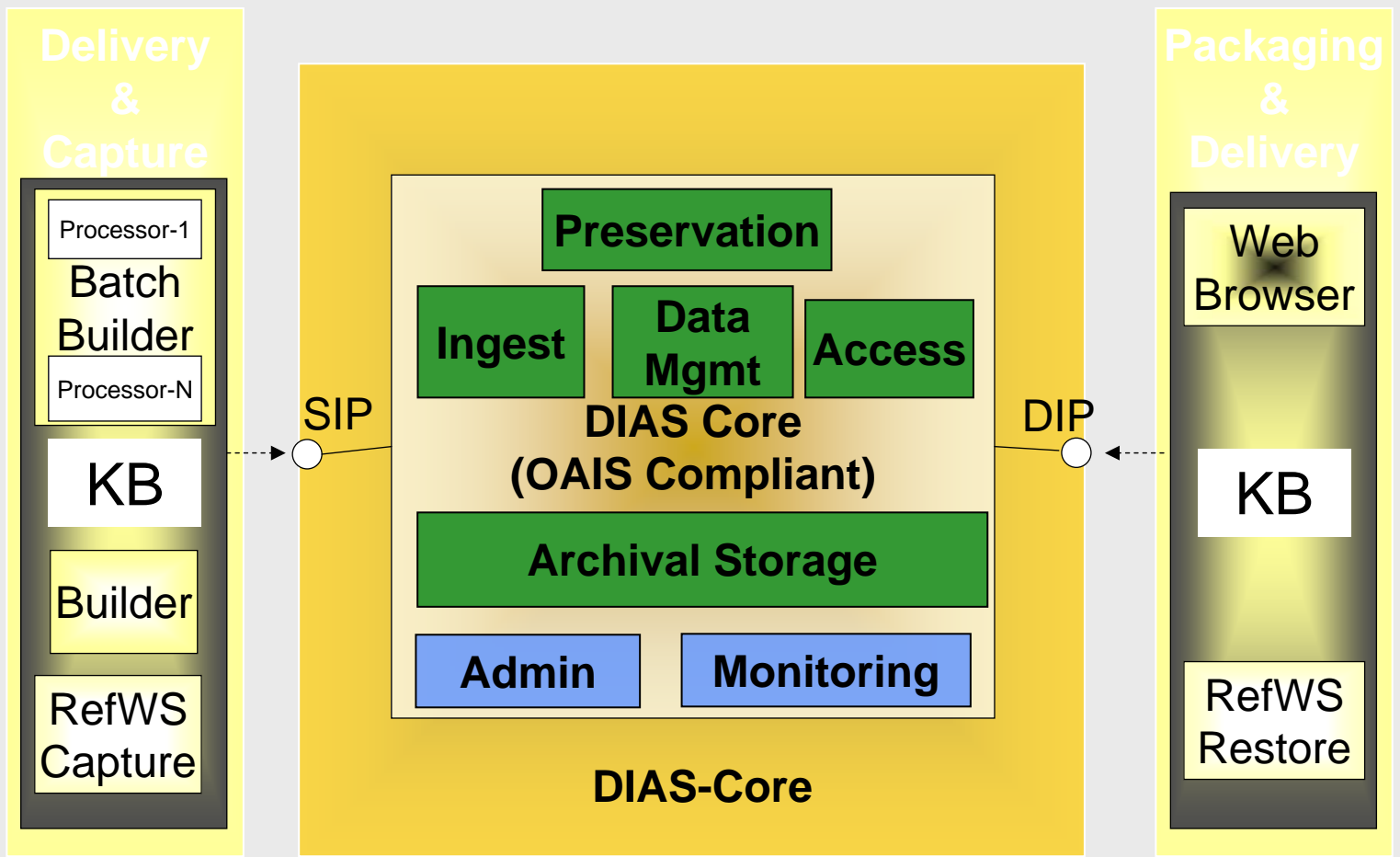
KOPAL: Einige Eckdaten

- Förderprojekt des BMBF
- Charakter: Forschungs- und Innovationsprojekt
- Fördervolumen: 4,2 Mio €, Laufzeit: 1.7.2004 – 30.6.2007
- Projektziel: Aufbau einer von Gedächtnisorganisationen nachnutzbaren technischen und organisatorischen Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen
 - Basis: **DIAS** (Digital Information Archiving System) der Koninklijke Bibliotheek, The Hague
 - Vermeidung von Einzelaktionen und Insellösungen
 - Kooperative Bewältigung der Aufgaben
 - Verfahrenslösungen für heterogene Daten und Anbieter

KOPAL konkret:

- Prototypische Implementierung eines skalierbaren Depotsystems für digitale Dokumente
- Weitestgehende Nachnutzung existierender Ansätze (technische Metadaten, Konzepte)
- Kooperation mit Nutzern (z.B. Dokumentservern)
- Kooperation mit Produzenten zur Entwicklung von Transferwegen
- Internationale Ausrichtung

DIAS II



Aufgabenstellung KOPAL, Projektplanung

- Kooperativer Betrieb und nachnutzbarer Transfer des DIAS-Core
- Entwicklung neuer Features: remote access, Mandantenfähigkeit, erweitertes SIP-Format, Datenmodell (technische Metadaten)
- Entwicklung und Implementierung von Tools für dedizierte Klassen digitaler Objekte
- Automatisierung der Einlieferrouinen, Optimierung für Massenverfahren
- Einführung von „preservation planning“-Verfahren
- Programmierung von Zugangs-Tools (Liefer- / Nutzungsschnittstellen)
- Workflow-Integration
- Pilotphase (07 – 10/2004)
- Entwicklungsphase (11/2004 – 06/2007)



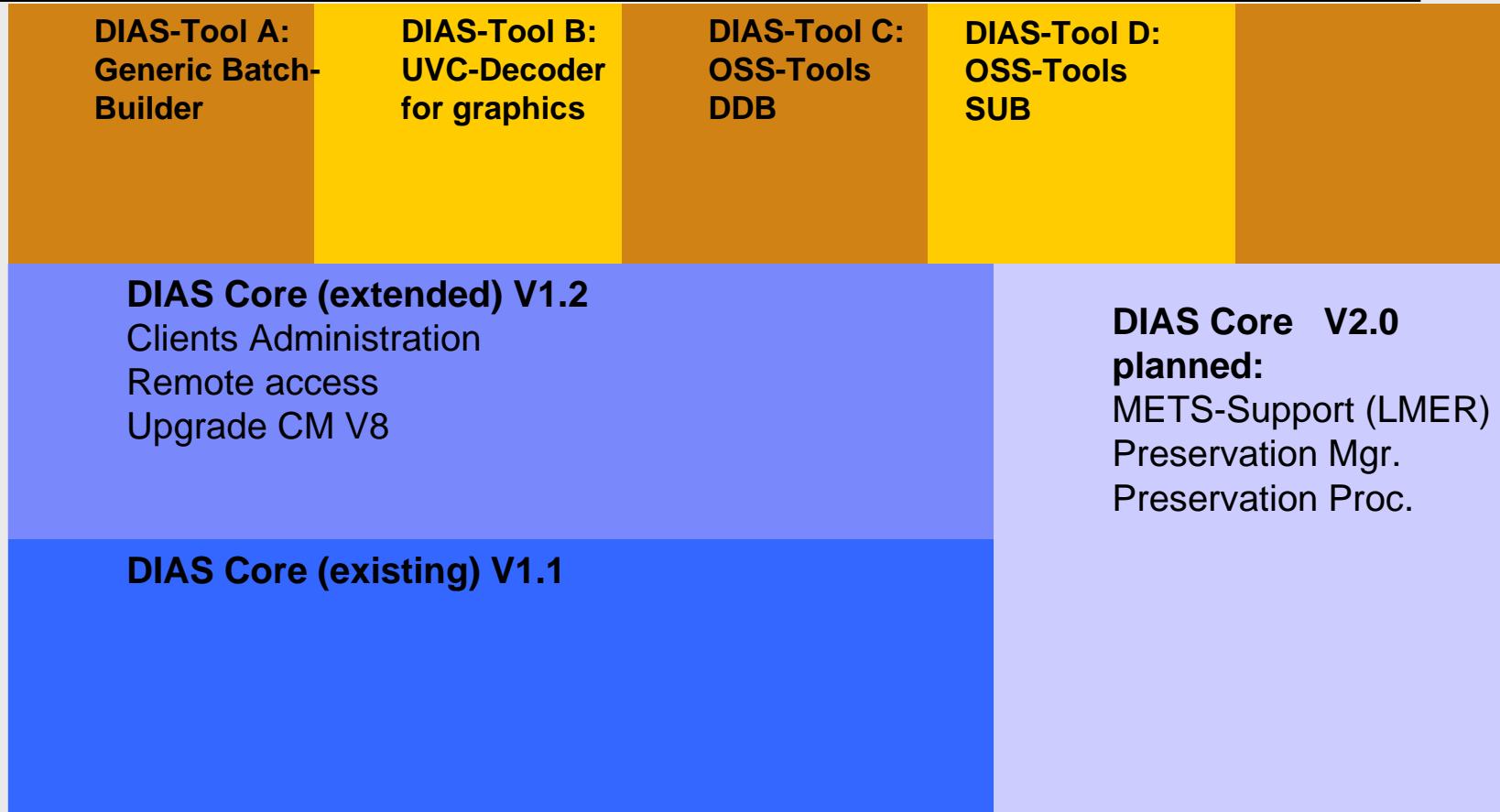
Das KOPAL-Projekt: Eckdaten 2

Projektpartner (Projektleitung: Die Deutsche Bibliothek)

- Die Deutsche Bibliothek (DDB)
- Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)
- Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG)
- IBM Deutschland GmbH (IBM)
- User Group mit niederländischer Nationalbibliothek (und weiteren Partnern?)

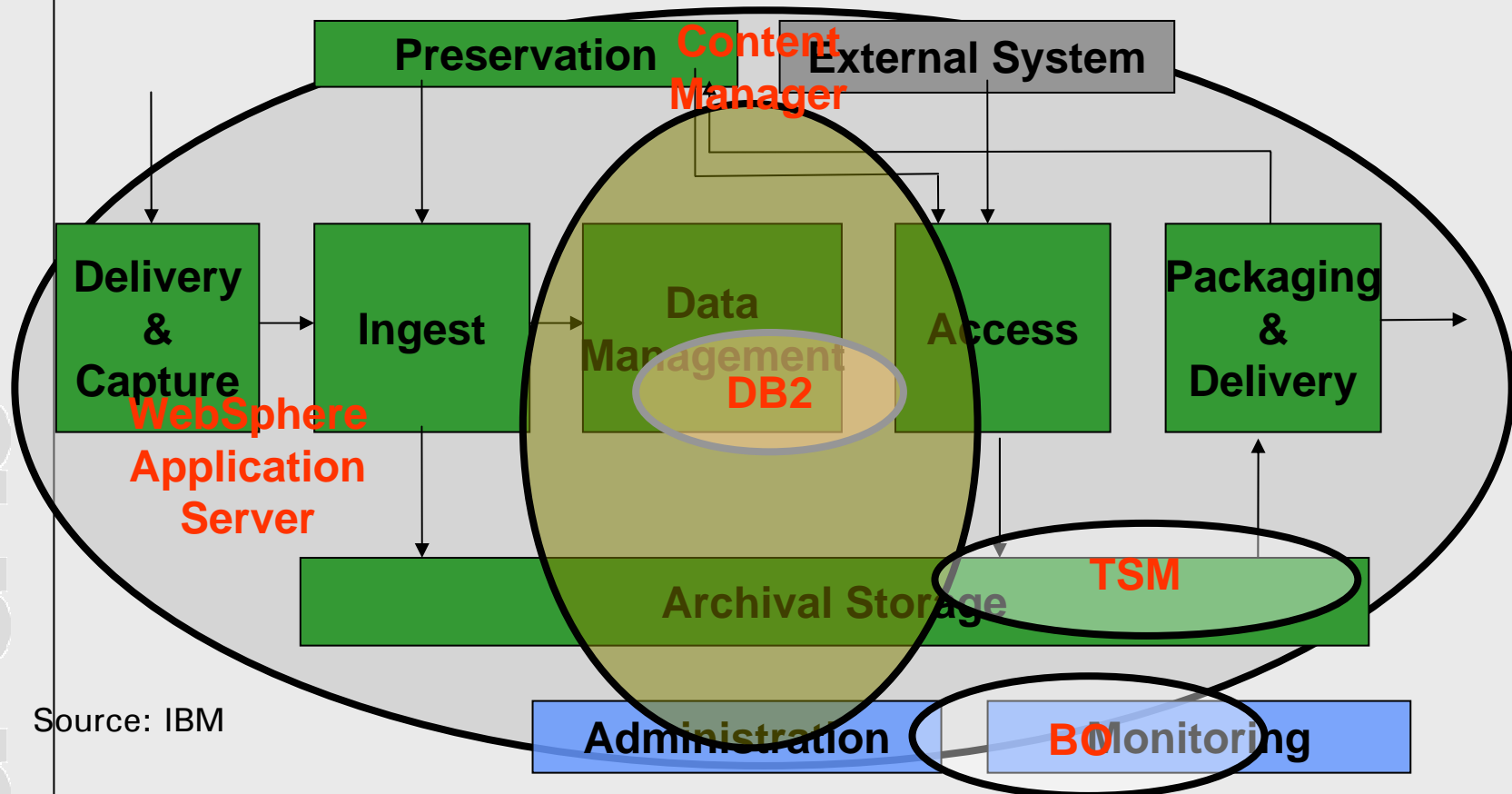


Software Entwicklung & Teilung





Software – Architektur



Source: IBM

Testergebnisse der Pilotphase: Übersicht

- Ziele der Tests in der Pilotphase: Erkenntnisgewinn über das bestehende DIAS-System und dessen Potenzial, als Basis für das Vorhaben KOPAL zu dienen
- Testsystem: Kernkomponente des Systems der Niederländischen Nationalbibliothek installiert bei der GWDG
- Testmaterial: Eigenes in passender Form und holländisches
- Testfälle: Von IBM konzipierte und eigene
- Testblöcke:
 - DDB: Von IBM konzipierte Tests mit eigenem Material unter Nutzung eines holländischen Tools („Builder“)
 - DDB: Eigene Tests mit dem holländischen Material
 - GWDG: Ausfalltests auf Basis der von IBM konzipierten Tests
 - SUB: Eigene Tests mit eigenem Material, teilweise unter Nutzung des Tools „Builder“

Zielsetzung und Struktur der Detailplanung

- Dauer der gesamten Laufzeit der Entwicklungsphase 32 Monate (Nov. 04 – Juni 07)
- Detailliertere Planung des KOPAL Projektes zeitlich und ressourcenseitig
- Strukturierung des Projektverlaufs anhand von vier Arbeitspaketen:
 - E1: DIAS-Core Basisentwicklung und Bereitstellung
 - E2: DIAS-Core-Anbindung
 - E3: Preservation Planning und Bitstream-Preservation
 - E4: Exemplarische Übernahme bestehender Datenbestände
- Strukturierung einer umfassenden Projektbegleitung durch drei weitere Arbeitspakete:
 - A1: Projektkoordination
 - A2: Projektunterstützung (fachlich und IT-seitig)
 - A3: Öffentlichkeitsarbeit
- Unterteilung der Arbeitspakete in Arbeitseinheiten, Benennung von Zuständigkeiten sowie Aufwandschätzungen

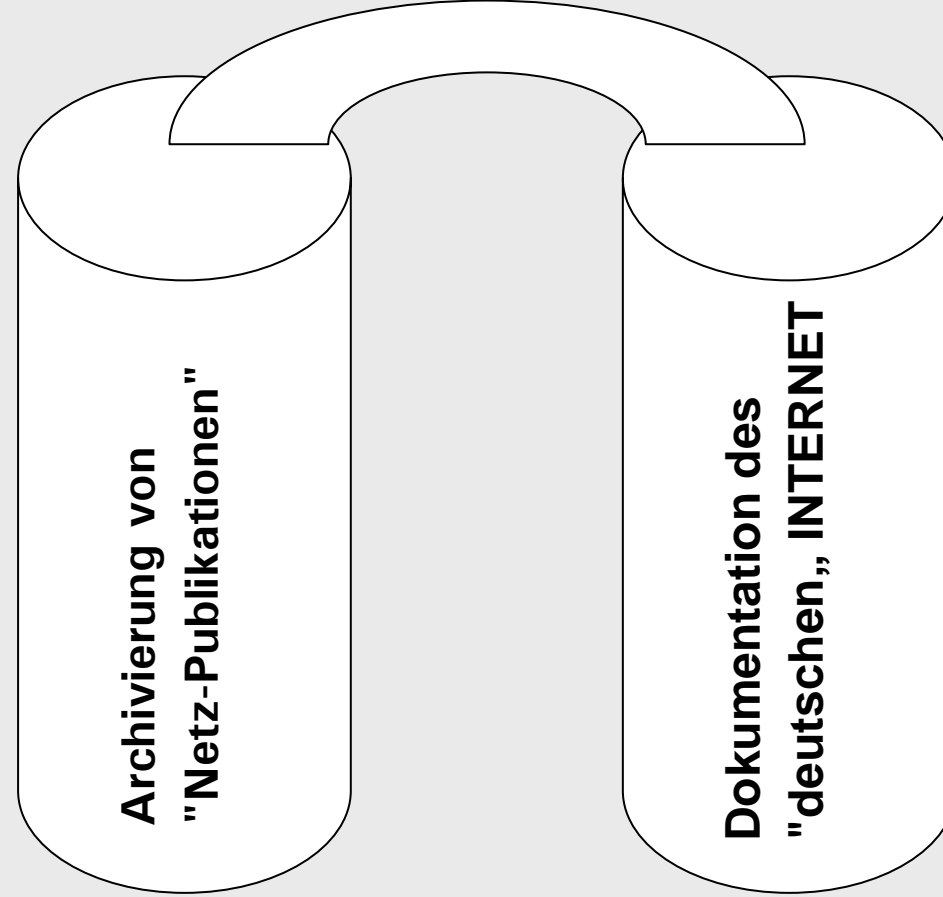
Objektspezifikation von DDB und SUB

- Definition eines universellen Objektformat für den Im- und Export
- Wichtiger Baustein zur Entwicklung eines nachnutzbaren Systems
- Aufbau auf vorhandenen Standards: METS, LMER
- Erfahrungen und Anforderungen von DDB und SUB
- Universeller Einsatz durch flexible Struktur
- Klar benannte Felder, die für eine effektive Verwaltung in einer Datenbank zu halten sind
- Anforderungen zur Administration der Objekte und zur Integration in eine Preservation-Strategie

III. Ausblick



Die Deutsche Bibliothek - künftige Sammelstrategie





Danke schön!

<http://www.langzeitarchivierung.de>

altenhoener@dbf.ddb.de