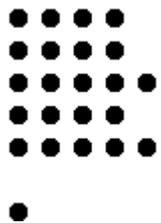


Vorträge zur gemeinsamen Herbsttagung von FIS Bildung und GIB
am 5.12.2001 zum Thema

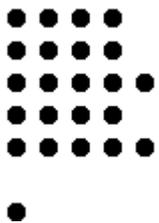
„WWW-Volltexte: Erschließung, Speicherung, Dienste“

1. Klaus Lepsky: Wörter - Texte - Information. Möglichkeiten und Grenzen automatischer Erschließungsverfahren. Volltextsuche am Beispiel MILOS/Kascade
2. Diann Rusch-Feja: Metadaten für den Bildungsbereich
3. Beate Träger: Informationsverbund Bildung-Sozialwissenschaften-Psychologie
4. Berthold Weiss: Dauerhafte Archivierung und Authentifizierung von WWW-Volltexten



Wörter - Texte - Information

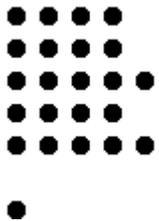
Möglichkeiten und Grenzen automatischer Erschließungsverfahren



1. Sprache und Informationserschließung

Informationssuche ist in der Regel Suche nach oder in Text/Sprache. Informationserschließung und Information Retrieval kommen deshalb nur unter erheblichem Qualitätsverzicht ohne sprachverarbeitende Verfahren aus.





Problem 1: Mismatching

Verschiedenartigkeit von Dokument- und Suchsprache

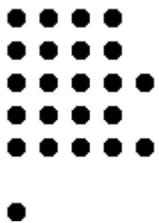
Sprache

**Sprachen
sprachlich
Sprachphilosophie
Fremdsprachen**

Lösung:

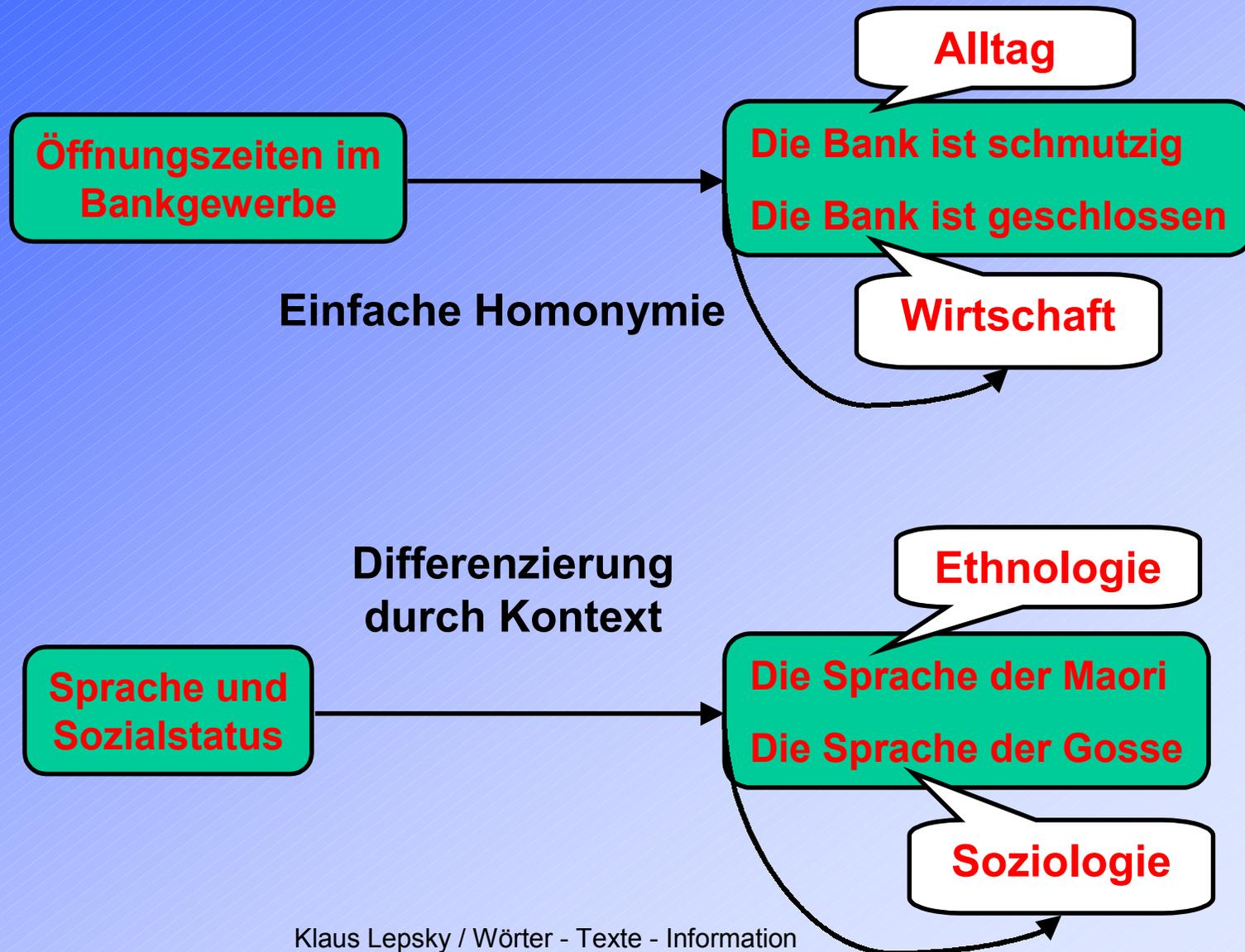
Einsatz einer morphologischen Komponente

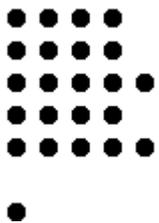




Problem 2: Bedeutungs-differenzierung

Trennung von sprachlicher Form und Bedeutung





Lösung: Einsatz einer kontextsensitiven Sprachanalyse

1. Partielles oder vollständiges Parsing

Identifizierung von
Satzzusammenhängen

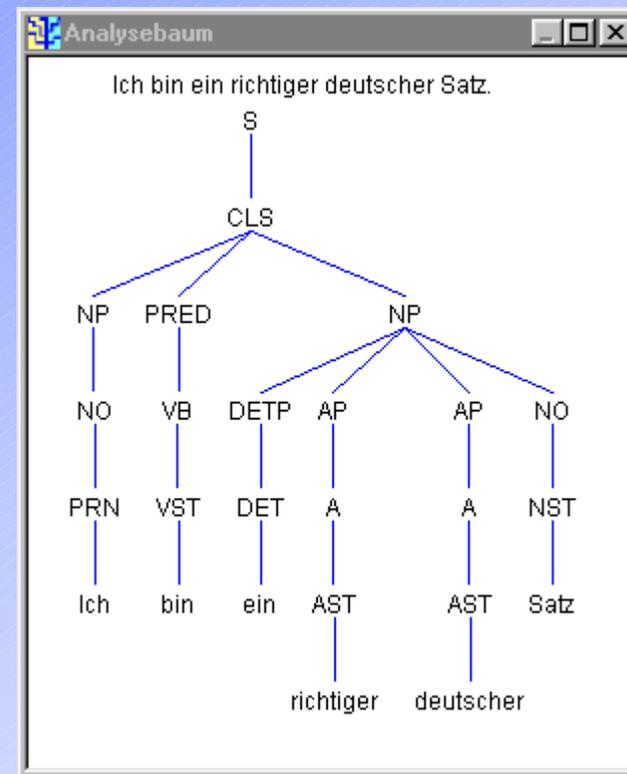
**Die Bank betrieb seit langem
schmutzige Geschäfte**

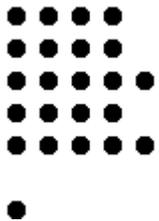
Identifizierung von
"Mehrwortgruppen" bzw.
"Themen-Aspekt-Beziehungen"

progressives Steuermodell

Lehrbuch der Psychologie

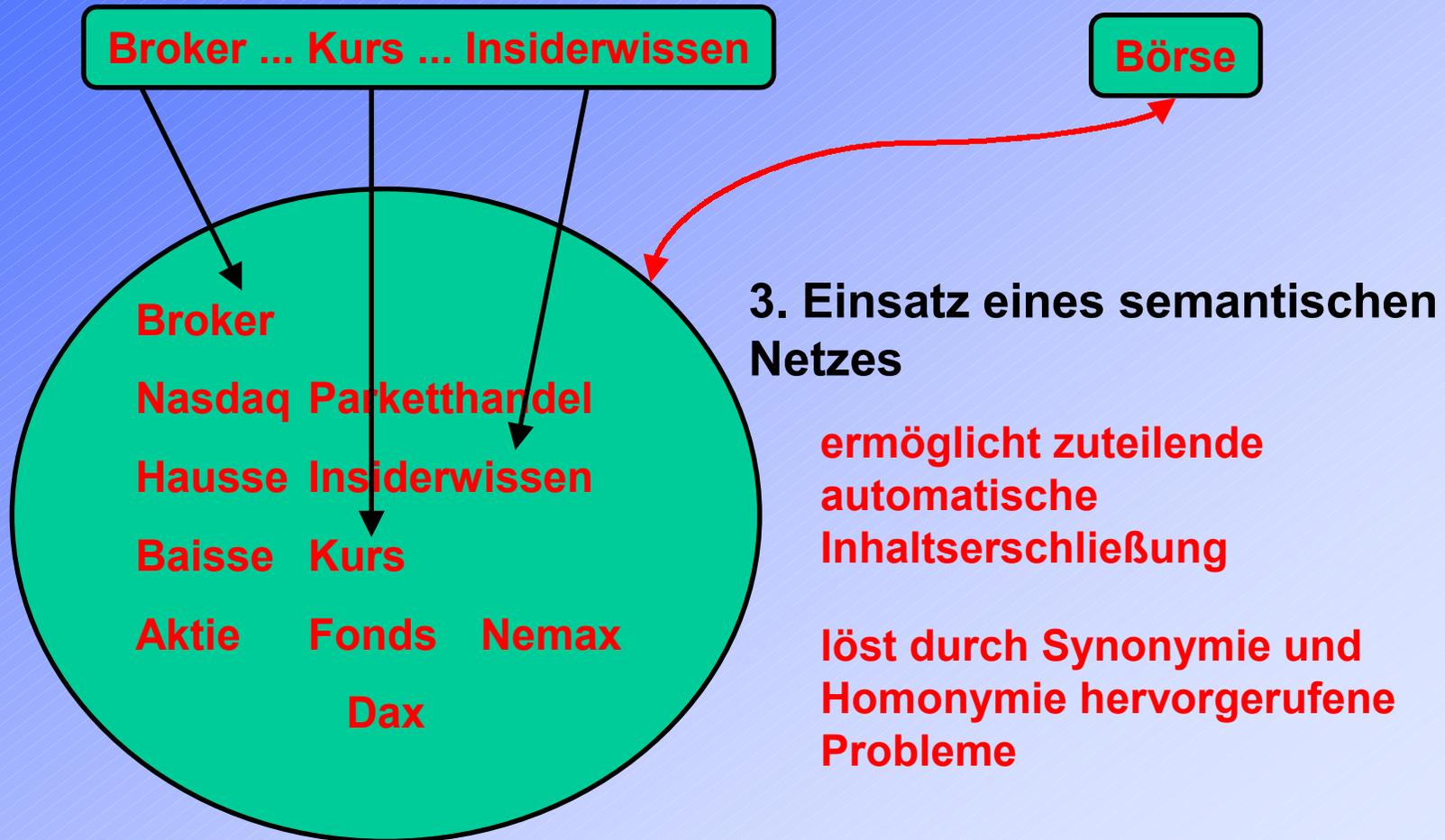
Information und Kommunikation

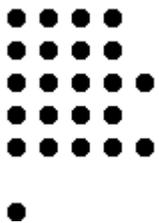




2. Einsatz einer statistischen Komponente zur Gewichtung

Identifizierung von thematischen Beziehungen





Bedingungen und Grenzen für den Einsatz angewandter Computerlinguistik in der Informationserschließung

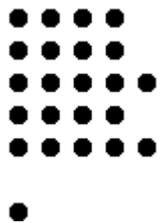
1. Die Textbasis der Dokumente muss ausreichend sein

Volltexte, Inhaltsverzeichnisse, Abstracts, Titel

2. Der für die Verarbeitung verfügbare Text muss ausreichende Aussagen über den Inhalt der Dokumente machen

3. Die Datenbasis für statistische und/oder zuteilende Verfahren muss hinreichend homogen sein

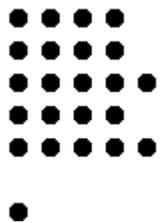
Allgemein sind die Grenzen automatischer Sprachverarbeitung dort erreicht, wo die Intelligenz beginnt.



2. Die MILOS/KASCADE-Indexierung

1. Anreicherung herkömmlicher textschwacher Titelaufnahmen durch Inhaltsverzeichnisse
2. Linguistisch basierte automatische Indexierung im reduzierten Umfang, d.h. nur Grundformerzeugung und Dekomposition
3. Statistische Gewichtung aller Begriffe
4. Ausschluss aller nicht relevanten Begriffe über einen Cut-Off-Wert
5. Linguistisch basierte Indexierung relevanter Begriffe mit semantischen Relationen
6. Durchführung eines Retrievaltests





Erfahrungen

Dokumentanreicherung
(Scanning)

technisch problematisch
und nicht effizient

Indexierung I

problemlos, vgl. MILOS I

$G = F1 * KG + F2 * DG + F3 * LG$
mit
KG = Kollektionsgewicht
DG = Dokumentgewicht
LG = Längengewicht
F1-F3 = Konstanten

Gewichtung

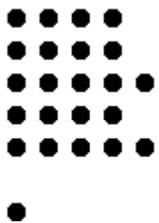
zuverlässig
aber
Performance-kritisch

Indexierung II

problemlos,
vgl. MILOS II

Kritischer Punkt!
Theoretisch eingängig,
in der Praxis nicht effektiv bis
kontraproduktiv

Cut-Off-Verfahren



Ergebnisse und Empfehlungen

- ➡ **Gewichtung verbessert die Ergebnisse einer *guten* linguistischen Indexierung *nicht***
- ➡ **Anreicherung + Automatische Indexierung ist in den meisten Fällen intellektueller Erschließung überlegen**
- ➡ **Gewichtung sollte für Relevance Ranking genutzt werden**
- ➡ **Erschließungsmodell**

Anreicherung und automatische Indexierung deutschsprachiger Quellen

Verbale Erschließung und/oder Übersetzung fremdsprachiger Quellen

Klassifikatorische Erschließung aller Quellen

Metadaten für den Bildungsbereich

*Dr. Diann Rusch-Feja
Max-Planck-Institut für Bildungsforschung
Berlin*

Welche Lehr- und Lern- ressourcen brauchen Metadaten?

*Kursmaterialien, Lerneinheiten,
Curricula, Lehrpläne, Syllabi, Texte,
Aufgaben, multimediale Werke,
Videos, Audiofiles, Bilder, CGI-Forms*

Metadaten

Informationen über Information

- Gezielter Zugang zu verteilten Informationsressourcen durch definierte Indexierung
- Bibliographische, inhaltliche, klassifikatorische, kontextuelle u. identifizierende Eigenschaften
- Ergänzung der Meta-Tags im HTML-Header
- XML Metadaten, RDF-Strukturen, SGML
- Analog zum Bibliothekskatalog / Index, aber mehr - (Mehrdimensional, neue Formate ...)

Besondere Anforderungen im Bildungsbereich an Metadaten

- Einstufung erforderlicher Vorkenntnisse
- Zuordnung zur Zielgruppe - Definition der Zielgruppe
- Integration: Orientierung an einer Reihenfolge (Beziehung zu anderen Ressourcen)
- Bewertung für die Praxis - Validitätssiegel
- Harmonisierung von Schulstufen, Abschlüssen, fachlichen Inhalten etc. (linguistisches / multilinguales Mapping)

Besondere Schwierigkeiten bei Metadaten im Bildungsbereich

- Bildungsinstitutionen sind in die Sozialstrukturen e. Landes eingebettet & widerstreben Vereinheitlichung/Internationalisierung
- Sprachbezeichnungen sind nicht gleich „eins-zu-eins“ zu korrelieren
- Nutzeroberfläche für die Suche in mehreren Servern, Datenbanken ermöglichen
- Geeignete Zielgruppenhilfen: Online Hilfsseiten, HelpDesk etc. bereitstellen

Metadaten - Ersteller

- Dublin Core Educational Metadaten AG (DC-ED)
- Instructional Management System (IMS) - Global Learning Consortium
- Alliance of Remote Instructional Authoring and Distribution Networks for Europe (Ariadne)
- Learning Technology Standardisation Committee, Learning Objects Metadata (LTSC-LOM) - Arbeitsgruppe des Institute of Electrical and Electronics Engineering (IEEE)
- European Committee for Standardization-Informat Society Standardization System (CEN-ISSS) - Learning Technologies Workshop

Dublin Core Educational Metadata

- Arbeitsgruppe seit 1998, >90 Mitglieder
- Prof. Stuart Sutton (U of Washington) und Jon Mason (EdNA, Australien) sind Ko-Leiter
- „Audience“ (Zielgruppe“) als neues Element mit Qualifier(n)
- Dauer der Lerneinheit
- 2001 DCMI-DC9 Tagungsergebnisse

DC-9 Ergebnisse für DC-ED

- Usage Board hat „Audience“ bestätigt.
- „Mediator“ wurde als Qualifier akzeptiert.
- „Level“ als Qualifier wird noch überlegt.
- „Conforms To“ Qualifier für DC.Relation, um die Zuordnung im Rahmen externer Standards, Benchmarks und Richtlinien zu belegen.
- Bildungsspezifische Ressourcenarten („Type“) werden durch ein Standardvokabular noch diskutiert.

MOU zwischen DCMI / LTSC

- Memorandum of Understanding (MOU) zwischen DCMI und IEEE Learning Technology Standards Committee (LTSC) (2001).
- Einheitliche Architektur, die wenige Probleme für die beiden Gruppen bereitet.
- Allgemeine Metadaten von DCMI, bildungsspezifische von LTSC.
- DC-ED Chairperson(s) können in LTSC mit abstimmen.
- HTML und XML Metadaten (auch RDF) werden akzeptiert.

Ottawa Communiqué

- Treffen August 2001 in Ottawa (DCMI, IMS, IEEE LTSC LOM AG).
- Grundlegende Prinzipien für ein gemeinsames „application profile“.
- Beispiele und Richtlinien bis Ende 2001.
- Beispiele eines Application Profile bis Ende Dezember 2001 in maschineller Form.
- Entwicklung & Pflege von Registries.

DC Metadaten für OAI Anwendung

- OAI (Open Archives Initiative) Metadata Harvesting Protocol beruht auf ein Kernsatz von DC Metadaten („unqualified“)
- OAI Protokoll für Suche in verschiedenen, verteilten „Archiven“ (Datensammlungen)
- http-basiert, Metadaten in XML (oder mit Hilfe eines Konversionstools)
- Eingesetzt für institutionelle und fachbezogene Server/Portale aller Art

<http://www.openarchives.org>

Instructional Management System (IMS) Global Learning Consortium

- Ursprünglich Konsortium von 30 Unis in den USA, die die ISM Infrastruktur übernahmen
- 2001 ca. 200 Mitglieder, die IMS Standards und Anwendungsfähigkeit prüfen
- IEEE 1484 Learning Objects Metadata (IEEE-LOM) zugrundegelegt
- “Learning objective”, “learning method”
- IMS “Module Set” - Kurseinheiten, Lernangebote

Alliance of Remote Instructional Authoring and Distribution Networks for Europe

- EU-Projekt „Telematics Applications“ Programme“ - Metakategorien für pädagogische Ressourcen
- Multilingualer Auftrag
- Neben üblichen Metadaten auch
 - Pädagogische Attribute
 - Technische Voraussetzungen
 - Benutzungsbestimmungen & Preisgestaltung

Learning Technology Standards Committee - Learning Objects Metadata

1996 vom Institute of Electrical and Electronics Engineering (IEEE) gegründet

- Rechnerunterstützte Trainingssysteme - Multimediale Lernumgebungen etc.
- Mitglieder aus der ganzen Welt (USA, Japan, Europa, Australien ...)
- „Glossar für kollaboratives Lernen“
- Regelmäßige Treffen & Arbeitssitzungen

LTSC - LOM Arbeitsgruppen

- 20 Arbeitsgruppen - Lernende, Daten und Metadaten, Inhaltlichbezogenheit etc.
- Überwindung der „Grenzen“ und „Schranken“ der individuellen Sozialstrukturen und Bildungsstrukturen
- LOM, auch kommerzielle Mitglieder
- Entwürfe zur Standardisierung und Strategieplanung
- LOM Entwurf Version 6 (Februar 2001)

Europäische Kommission für Standardisierung (CEN)

Information Society Standardisation System
- Learning Technologies Workshop (CEN-
ISSS) - 1997

- *Offene Teilnahme,
Offene Konsensbildungsverfahren*
- *Fokussiert auf Metadatenstandards für
Multimediale Lernobjekte inkl. Vertreter der
Bildungsministerien, Software-Produzenten,
Universitäten, Bildungseinrichtungen (auch
nichttraditionelle)*

Kooperationen

- Die 5 Gruppen kooperieren und tauschen sich aus über die Standardisierung von bildungsbezogenen Metadaten
- Gemeinsame Sitzungen und Erprobung der Standards in den einzelnen Gruppen
- Kollaborationen in Training, Seminaren und Terminologie

... Fazit: gemeinsam ungelöste Probleme angehen und lösen ...

Dr. Diann Rusch-Feja
Max Planck Institute for Human Development
Berlin, Germany

email: ruschfeja@mpib-berlin.mpg.de

www.mpib-berlin.mpg.de/dok/ehomee.htm



Der Informationsverbund Bildung

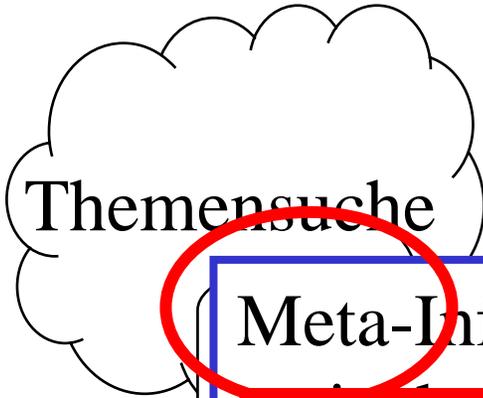
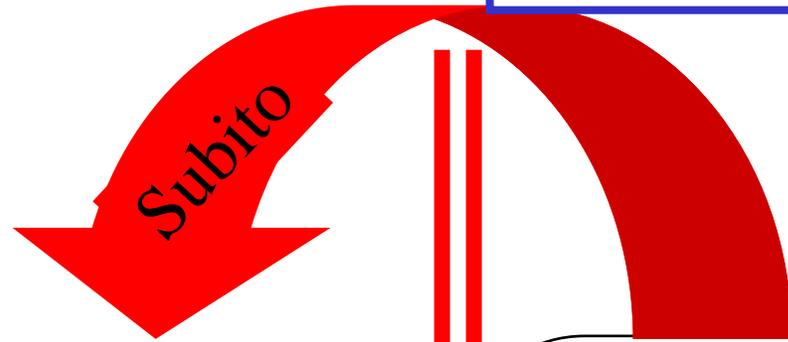
Der direkte Weg von der Fachinformation zum Text

Dr. Beate Tröger

Deutsches Institut für

Internationale Pädagogische Forschung

Fachinformation: Die Zukunft



Meta-Informationen zu

FIS Bildung

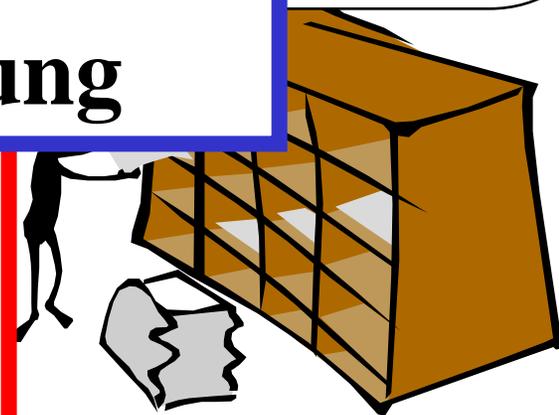
Bibliothek

är-Info
(-) Text

**Informations-
Verbund
Bildung**

Meta-Informat

Dt. Bildungsserver
zu Internet-Ressourcen,
zu Multimedia etc.





Informationsverbund Bildung

Zugleich Vernetzung mit Bibliothekskatalogen und Dokumentlieferdiensten zum Angebot von Volltexten = Primärinformation unabhängig von ihrer physikalischen Form

liche, örtliche und mediale Information (= Volltext) Information über den

... eines Komplet-Systems ein integrierter Zugang zu allen relevanten Informationen ihres Interessengebietes angeboten:

von den differenzierten und qualitativ hochwertigen Metadaten auf Einzeldokument Ebene aus den Fachdatenbanken bis hin zum online abrufbaren Volltext



Informationsverbund Bildung

Die bislang existierende zeitliche, örtliche und mediale Mauer zwischen der Primärinformation (= Volltext) und der Meta-Information (= Information über den Volltext) wird überwunden:

Komplettsystem

Pay-per-View-Basis

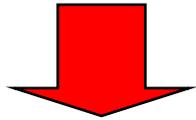
Vorhandene
vh. Medien
Information
ebenso wie
Ressourcen
werden aufgehoben.

Integration heterogener
Datenarten und Quellen:
Struktur- und Faktendaten,
Multimedia etc.

erfolgt zeit- (24h-
auch orts-
Temporäre, aber
ch lokale
Nutzungsgrenzen
werden drastisch erweitert.

Inhalte:

Fokus: Lizenzpflichtige Zeitschriften der einschlägigen deutschen und internationalen wissenschaftlichen und praxisorientierten



Verträge mit
stellung der

Basis für Nu
- die Preisge

Darüber hinaus integriert werden kostenfreie Dokumente / ‚Graue Literatur‘: E-Journals, Preprints, Reports, Dissertationen etc.; Lehr- und Lernmaterialien

- deutsche und internationale Zeitschriften / Ressourcen

- wissenschaftsorientierte und praxisorientierte Ressourcen

- lizenzpflichtige und lizenzfreie Ressourcen

Multimedia



Informationsverbund Bildung

Funktionalitäten:

Zuverlässige Verfügbarkeit des Zugriffs - auch für ‚nicht-aktuelle‘ Literatur - durch Volltextspeicher und durch Langzeit-Archivierung durch DDB und SSG-Bibliotheken

Ortsunabhängiger (= campusferner) Zugriff -

- aber zugleich Nachweis der Campuslizenzen über EZB

Elektronische und konventionelle Zahlungsverfahren

Push-/Alertingdienste sowie qualifizierte Help-Desks

Integration nicht elektronisch verfügbarer Literatur durch bibliothekarische Kataloge und Dokumentlieferdienste

Offene Architektur - (DC-) Metadaten für URNs / DOIs



Informationsverbund Bildung

Die Projektpartner:

Bildungsinformation des DIPF Frankfurt / Main

IZ Sozialwissenschaften Bonn, ZPID Psychologie Trier

SSG-UBs Erlangen-Nürnberg, Köln, Saarbrücken

(Die Deutsche Bibliothek)

(EZB Regensburg)

**Bibliotheken und dokumentarische
Fachinformationseinrichtungen
arbeiten unmittelbar zusammen**

BMBF- / DFG-Projekt

Laufzeit 3 (BMBF) bzw. 2 (DFG) Jahre ab 1.6.2001



Informationsverbund Bildung

Die Aufgabenverteilung:

Bildungsinformation DIPF

IZ Sozialwissenschaften Bonn

ZPID Psychologie Trier

SSG-UB Erlangen-Nürnberg

SSG-UB Köln

SSG-UB Saarbrücken

(Die Deutsche Bibliothek)

(EZB Regensburg)

Inhaltliche Artikel-
Erschließung,
technische
Realisierung des
Info-Verbundes

Erwerbung der
Content-Lizenzen,
Archivierung

Archivierung

Lizenznachweise



Informationsverbund Bildung

Aktueller Stand des Projektes:

Die Stellenbesetzungen bei den Projektpartnern sind erfolgt

Die technischen Arbeiten des ersten Meilensteins laufen

Eine Marketing-Studie / -Entwicklung wird in Auftrag gegeben

Die Namensfindung und Logo-Entwicklung werden in Auftrag gegeben

Die Vorbereitungen (Vorstudie) für ein Ergänzungsmodul „Wissenschaftsportal Bildung“ laufen an

Das „Wissenschaftsportal Bildung“

Ziel: Informations- und (bei Bedarf) Kommunikationsplattform speziell für den Wissenschaftsbereich

Inhalt: Informationsbündelung der sehr verteilten Ressourcen auf Fachbereichs- und Hochschulschriftenservern etc.

D.h. ein zentraler Zugang zu vielen dezentralen Inhalten - z.B. zu Projekten (dabei enge Vernetzung mit Deutschem Bildungsserver)

Einbindung von professional homepages

Verlinkung bildungspolitischer Bereiche

Kommunikationsforum für den wissenschaftlichen Nachwuchs

Das „Wissenschaftsportal Bildung“

Methode:

Dezentral „gefülltes“ und zentral (und ggf. in Teilen auch dezentral) redaktionell bearbeitetes System

Basis dabei sind standardisierte Verfahren zur Qualitätssicherung bei Metadaten, Thesauri und Systematiken entsprechend der fachwissenschaftlichen Erfordernisse

Die entsprechende technische Infrastruktur wird zentral eingerichtet und gepflegt: Searching und Browsing im einem sehr nutzerfreundlichen „niedrigschwelligen“ System

Gewährleistung von Nachhaltigkeit



Eigentlich: Informationsverbund Bildung-Sozialwiss.-Psychologie

Hoher inhaltlicher Berührungs- / Überschneidungsgrad:

Erprobung neuer Finanzierungsmöglichkeiten für elektronische Dienstleistungen in den Geistes- und den Sozialwissenschaften
transparente Kosten- und Qualitätskriterien

Basis: Definition der Zielgruppen je Infoverbund und je Infoverbund-Partner - d.h. z.B. für die Bildung:

Kunden sind vor allem auch (campusferne) Interessierte aus der Bildungspraxis - LehrerInnen, WeiterbildnerInnen, Eltern - und der Bildungspolitik neben Kunden aus der Bildungsforschung und -lehre



... und die anderen Informationsverbände ...

Informationsverbund Bildung-Sozialwiss.-Psychologie

GetInfo - Informationsverbund Naturwissenschaft /
Technik

EconDoc - Informationsverbund Wirtschaft

Informationsverbund Medizin

Internationale Ausrichtung

Verlags- und Graue Literatur

Kooperation mit internationalen Preprint-Servern,
Universitätsverlagen und Forschungseinrichtungen
(z.B. MPG)

Angebot an Verlage: Digitalisierung on Demand

zusätzl. Angebot an NutzerInnen: Print on Demand

Subskriptionsunterstützung



www.getinfo-doc.de

Ihr Zugang zu Volltext in Wissenschaft und Technik - Netscape

File Edit View Go Communicator Help

GetInfo

Ihr Zugang zu Volltext in Wissenschaft und Technik

[Home](#) [Neukunde](#) [Kundenverwaltung](#) [Hilfe](#) [Kontakt](#) [Übersicht](#) [English](#)

Über GetInfo

- [Was ist GetInfo?](#)
- [Wer steht hinter GetInfo?](#)
- [Angebot an Verlage](#)
- [Angebot an Autoren](#)

Was ist neu?

- [News](#)
- [Pressekontakt](#)

Weitere Dienste

- [FIZ AutoDoc](#)
- [TIB Order](#)

GetInfo Partner

- [FIZ Karlsruhe](#)
- [TIB Hannover](#)

GetInfo ist ein gemeinsames Unternehmen von FIZ Karlsruhe und TIB Hannover zur Volltextversorgung in Wissenschaft und Technik. Durch die Integration der führenden Dokumentliefer- und -vermittlungsdienste TIBORDER (TIB Hannover) und FIZ AutoDoc (FIZ Karlsruhe) ist GetInfo Ihr zentraler Ansprechpartner für alle Fragen der Lieferung elektronischer und gedruckter Dokumente in Wissenschaft und Technik.

Copyright © 2001 by GetInfo, Germany - Last Update: Thursday, January 01, 1970 01:00:00

Dokument: Übermittelt

Start | Ihr Zugang zu Volltext... | Inbox - Netscape-Ordner | Microsoft PowerPoint - [Pr... | 10:45

Realisierung eines FIZ Wirtschaft

Fokus: deutschsprachige Zeitschriften

Keine graue Literatur (= Teil der Virtuellen Fachbibliothek Wirtschaft)

Fokus: Literatur des laufenden Jahres; i.d.R. erfolgt keine retrospektive Erfassung

Die Recherche selbst ist kostenfrei; es wird ein Serviceaufschlag für den Zugang zu den Fachdatenbanken bei der eigentlichen Volltextauswahl erhoben

Suche_FIZ_W - Netscape

Datei Bearbeiten Ansicht Gehe Communicator Hilfe

Suche in den **EconDoc** Themenpools nach:

<input type="radio"/> Wissenschaft	<input type="radio"/> Presse	<input checked="" type="radio"/> Gesamte Zeit
<input type="radio"/> Wirtschaftswissenschaften	<input type="radio"/> Tages- / Wochenpresse	<input type="radio"/> Letzter Monat
<input type="radio"/> Betriebswirtschaft	<input type="radio"/> Fachpresse	<input type="radio"/> Letzte Woche
<input type="radio"/> Volkswirtschaft	<input type="radio"/> Wirtschaft	
<input type="radio"/> Politik	<input type="radio"/> Politik	

Suche in einzelnen Datenbanken - EconDoc

BLISS Betriebswirtschaft	HWWA Betriebs- u. Volkswirtschaft / Wirtschaftspraxis
Kölnener Betriebswirt. Katalog	ECONIS Wirtschaftswissenschaften
Volkswirtschaft / ifolit	Volkswirtschaft / ifokat
MMKA - Management and Marketing Abstracts	MIND Kreditwirtschaft

Dokument: Übermittelt

Start Suche_FIZ_W - Nets... Inbox - Netscape-Ordner Microsoft PowerPoint - [Pr... 10:56



Informationsverbund Medizin

Internationale Ausrichtung

Fokus: Verlagspublikationen; fachentsprechend
kaum graue Literatur

Zugriff über die Homepages der beiden Partner,
kein Aufbau eines separaten Portals

Die Recherche in den Verlagsdaten bis hin zu den
Abstracts erfolgt kostenfrei



Die AG der Informationsverbände

Ziele:

- Absprachen über den Content der einzelnen Verbände und eine wo sinnvoll gemeinsame Erwerbungspolitik
- Aufstellung von Musterverträgen (u.a.: Preissplitting in kommerzielle/nicht-kommerzielle Nutzergruppen und Berücksichtigung einer Konsortialzugehörigkeit)
- Gemeinsames Marketing
- Pflege der fachspezifischen Systeme / Dienstleistungen / Kundenbindungen, aber zugleich Möglichkeit eines technisch einfachen integrierten Zugriffs auf die Datenbanken / Informationen aller Informationsverbände



Die AG der Informationsverbände

Vorteile für Verlage sind u.a.:

Zielgruppenerweiterung = Kundenkreiserweiterung

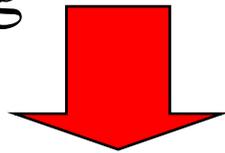
Kostenneutraler Kompetenzgewinn

Kostenneutraler Dienstleistungsgewinn -

- bei eigenständiger Preisgestaltung

Einheitliche, sehr tiefe und qualitative Erschließung

Marketing



Nach anfänglichem Zögern sind Verlage mittlerweile zunehmend an einer Teilnahme interessiert und schließen entsprechende Verträge ab

Vorteile für NutzerInnen:

Themensuche



Meta-Informationen zu

meist konvent
publizierter L

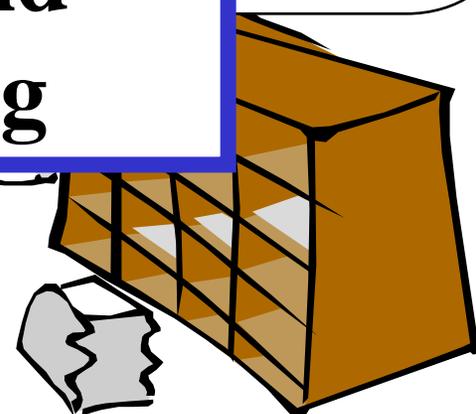
Bibliothek

är-Info
(-) Text

**Informations-
Verbund
Bildung**

Meta-Informa

Dt. Bildungsserver
zu Internet-Ressourcen,
zu Multimedia etc.





Gemeinsame Fachtagung
von FIS Bildung und Gesellschaft Information Bildung (GIB)
DIPF, 5. Dezember 2001

**Dauerhafte Archivierung und
Authentifizierung von WWW-Volltexten**
Die Vorgehensweise Der Deutschen Bibliothek
bei der Sammlung von Netzpublikationen

Berthold Weiß, M.A.

Die Deutsche Bibliothek Frankfurt am Main



Gliederung

- Sammelauftrag Der Deutschen Bibliothek
- Sammlung von Netzpublikationen
- Langzeitverfügbarkeit von Netzpublikationen
- Persistent Identifier



Die Deutsche Bibliothek Sammelauftrag

- Gesetz über die Deutsche Bibliothek vom 31. März 1969
 - Vollständige Sammlung, Erschließung und dauerhafte Sicherung deutschsprachiger und in Deutschland erschienener Publikationen
 - Gilt auch: für digitale Publikationen auf physischen Trägern
 - Gilt (noch) nicht: für Netzpublikationen



Die Deutsche Bibliothek Sammelauftrag - Netzpublikationen

Grundlegende Aspekte (Verabschiedung durch den Verlegerausschuss des Börsenvereins des Deutschen Buchhandels, Juni 1997)

- Ablieferung über Datennetze, ggf. auf Datenträgern
- Netzpublikationen in verschiedenen Formaten sind in dem von der Bibliothek gewünschten Format abzuliefern
- Werden inhaltlich identische Medienwerke gleichzeitig in physischer Form und als Netzpublikation verbreitet, sind beide Formen abzuliefern





Die Deutsche Bibliothek Sammelauftrag - Netzpublikationen

- Inhaltlich identische Netzpublikationen von mehreren Anbietern sind nur einmal abzuliefern
- Kopie der digitalen Publikation für Zwecke der Langzeitarchivierung (Sicherung der Authentizität des Inhalts)

Siehe auch: Conference of European National Librarians / Federation of European Publishers (CENL/FEP): International declaration on the deposit of electronic publications

-><http://www.ddb.de/news/epubstat.htm>

Sammlung von Netzpublikationen Richtlinien

- Anwendung des derzeit geltenden Gesetzes mit Anpassungen an die Materialart Netzpublikationen
- Formen der Behandlung
 - Archivierung
 - Verlinkung und Erschließung
- Ansätze zur Sammlung
 - „Flächiger Ansatz“ (Nordic Web Archive)
 - „Selektiver Ansatz“ (DDB)



Sammlung von Netzpublikationen Richtlinien

- Nicht sammelpflichtig
 - E-Mails
 - Allgemeine Informationen über Personen, Institutionen, Projekte, Veranstaltungen etc.
 - Servicedienste für Konsumenten
 - Suchdienste, Bestandskataloge
 - Nur kurzzeitig zur Verfügung stehende Netzpublikationen
 - ...



Sammlung von Netzpublikationen AG Elektronische Depotbibliothek

- Ziel: Vorbereitung eines Gesetzes zu Pflichtabliefereung von Netzpublikationen
 - Fragestellungen: technische Abrufmöglichkeiten, Zugriffsmodalitäten, Formatfragen, Metadaten
- Teilnehmer
 - Die Deutsche Bibliothek, Springer, DuMont, Wiley/VCH, Saur, Buchhändler-Vereinigung
- Die Deutsche Bibliothek archiviert ausgewählte Netzpublikationen
 - Z.B.: Zeitschriften von Springer LINK



Sammlung von Netzpublikationen Abgabeverfahren

Ausweitung des Verfahrens zur Abgabe von Netzpublikationen auf weitere Ablieferer (9/2001)

- Kontaktaufnahme
 - Übermittlung von Adressdaten
 - Zuteilung einer „Anmelder-Identifikation“
- Abgabe einer Netzpublikation
 - Übermittlung von Metadaten über das interaktive Anmeldeformular



Abgabe von Netzpublikationen - Netscape

Datei Bearbeiten Ansicht Gehe Communicator Hilfe

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

Lesezeichen Adresse: http://deposit.ddb.de/cgi-bin/epmeld_form.pl

Abgabe von Netzpublikationen

Die Deutsche Bibliothek

(*) Anmelderkennung:

(*) Passwort:

Weiter

! Die Abgabe von Online-Hochschulschriften erfolgt über das [Formular zur Anmeldung von Online-Hochschulschriften](#). Für CIP-Meldungen steht ein separates [Online-Meldeformular](#) zur Verfügung.
Für Testzwecke steht ein Gastzugang zur Verfügung: Anmelderkennung *F1111-1111*, Passwort *Gast*. Unter dieser Kennung kann das Formular ausgefüllt, aber nicht abgesandt werden.

[↑](#) [|||](#) [@](#) [Informationen zur Abgabe von Netzpublikationen](#)

Copyright: Die Deutsche Bibliothek
Marcus Klein / 29.10.2001

Dokument: Übermittelt



Abgabe von Netzpublikationen - Netscape

Datei Bearbeiten Ansicht Gehe Communicator Hilfe

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

Lesezeichen Adresse: http://deposit.ddb.de/cgi-bin/epmeld_form.pl

Abgabe von Netzpublikationen

 Die Deutsche Bibliothek

Anmelder-ID: F1111-1111
Name: Musterverlag
Hauptsitz: Musterstadt
Weitere Sitze des Verlages:
Anschrift:
Bemerkung: Test user fuer Formularnutzung, fuer FTP gesperrt

Ansprechpartner

Name: Mustermann
Funktion:
E-Mail: mustermann@musterverlag.de
Telefon:
Fax:

 **Art der Netzpublikation**

abgeschlossene Publikation : Monographie
 fortlaufende Publikation : Zeitschrift / Datenbank

Dokument: Übermittelt

5. Dezember 2001



Abgabe von Netzpublikationen - Netscape

Datei Bearbeiten Ansicht Gehe Communicator Hilfe

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

Lesezeichen Adresse: http://deposit.ddb.de/cgi-bin/epmeld_form.pl

Abgabe von Netzpublikationen

Die Deutsche Bibliothek

(2/3) Bibliographische Metadaten : Monographie

(*) markiert Pflichtfelder
+ erweitert das Formular um weitere Eingabefelder

Verlag / herausgebende Stelle

(*) Funktion: Verlag Vertrieb/Host
Anmelder-ID: F1111-1111
(*) Name: Musterverlag
Ort(e): Musterstadt +

Autor

Nachname:
Vorname: +

Dokument: Übermittelt



Abgabe von Netzpublikationen - Netscape

Datei Bearbeiten Ansicht Gehe Communicator Hilfe

Zurück Vor Neu laden Anfang Suchen Guide Drucken Sicherheit Shop Stop

Lesezeichen Adresse:

Abgabe von Netzpublikationen  Die Deutsche Bibliothek

(3/3) Angaben zur Archiv-Version der Netzpublikation : Monographie

Einstiegsdatei (Startseite)

Dateiname
(ggf. mit Pfadangabe):

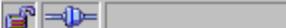
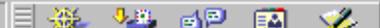
Datenstand

(?) (JJJ-MM-TT):

(?) **Zugriffsrechte gemäß Vorgabe der abgebenden Stelle**

beschränkt unbeschränkt

(?) **Systemvoraussetzungen**

 Dokument: Übermittelt 

5. Dezember 2001



Sammlung von Netzpublikationen Abgabeverfahren

- Quittierung der Anmeldung
 - Rücklieferung der Metadaten
 - Zuteilung einer Lieferungs-Identifikation
- Übermittlung der Netzpublikation
 - Erstellung des Lieferformats
 - Ablage auf dem ftp-Server Der Deutschen Bibliothek

Langzeitverfügbarkeit von Netzpublikationen

- Erhaltung der Substanz
 - Probleme
 - Datenträgerfragilität
 - Technologiewandel
 - Maßnahmen
 - Media refreshment
 - Datenträgermigration

Langzeitverfügbarkeit von Netzpublikationen

- Erhaltung der Benutzbarkeit
 - Migration der Inhalte / Formate
 - Emulation
 - Prävention
 - Verringerung der Vielfalt
 - Bewertung der Formate im Sinne der Langzeitverfügbarkeit
 - Separierung von Inhalt, Struktur und Layout

Langzeitverfügbarkeit von Netzpublikationen

- Erhaltung der Authentizität
 - Standards
 - Digitale Zertifikate
 - Signaturen
 - Zeitstempel
 - Beständige Identifikatoren

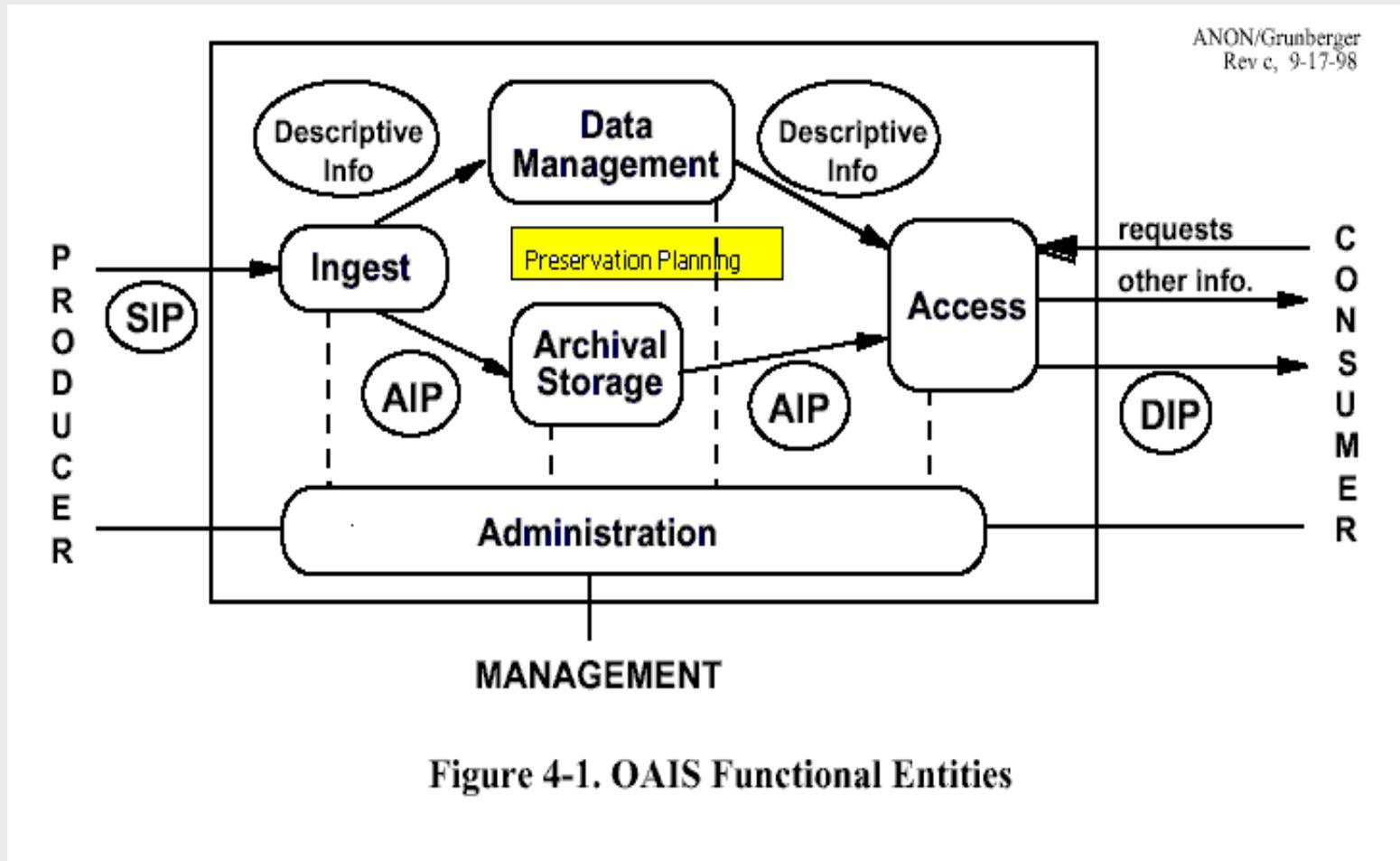
EU-Projekt (1998-2000) – Ergebnisse:

- **An Experiment in Using Emulation to preserve Digital Publications** / by Jeff Rothenberg, RAND-Europe
- **Metadata for Long Term Preservation** / by Catherine Lupovici and Julien Masanès, Bibliothèque Nationale de France.
- **Standards for Electronic Publishing: an overview** / by Mark Bide & Associates

- **Standards for a DSEP** Standards for the Implementation of a Deposit System for Electronic Publications (DSEP) / by Bendert Feenstra, IBM
- **The Nedlib Guidelines** : Setting up a Deposit System for Electronic Publications / by Johan Steenbakkers
- **A Process Model** : The Deposit System for Electronic Publications / by Titia van de Werf.

- **DEPOT 2000: Functional Design for a digital depot** / by Nico van Emond, Hans Hofman, Jacqueline Slats, Tamara van Zwol. State Archive's Services (Rijksarchiefdienst)
- **List of NEDLIB Terms** / by Genevieve Clavel-Merrin, Schweizerische Landesbibliothek

Depotsystem für EP - DSEP

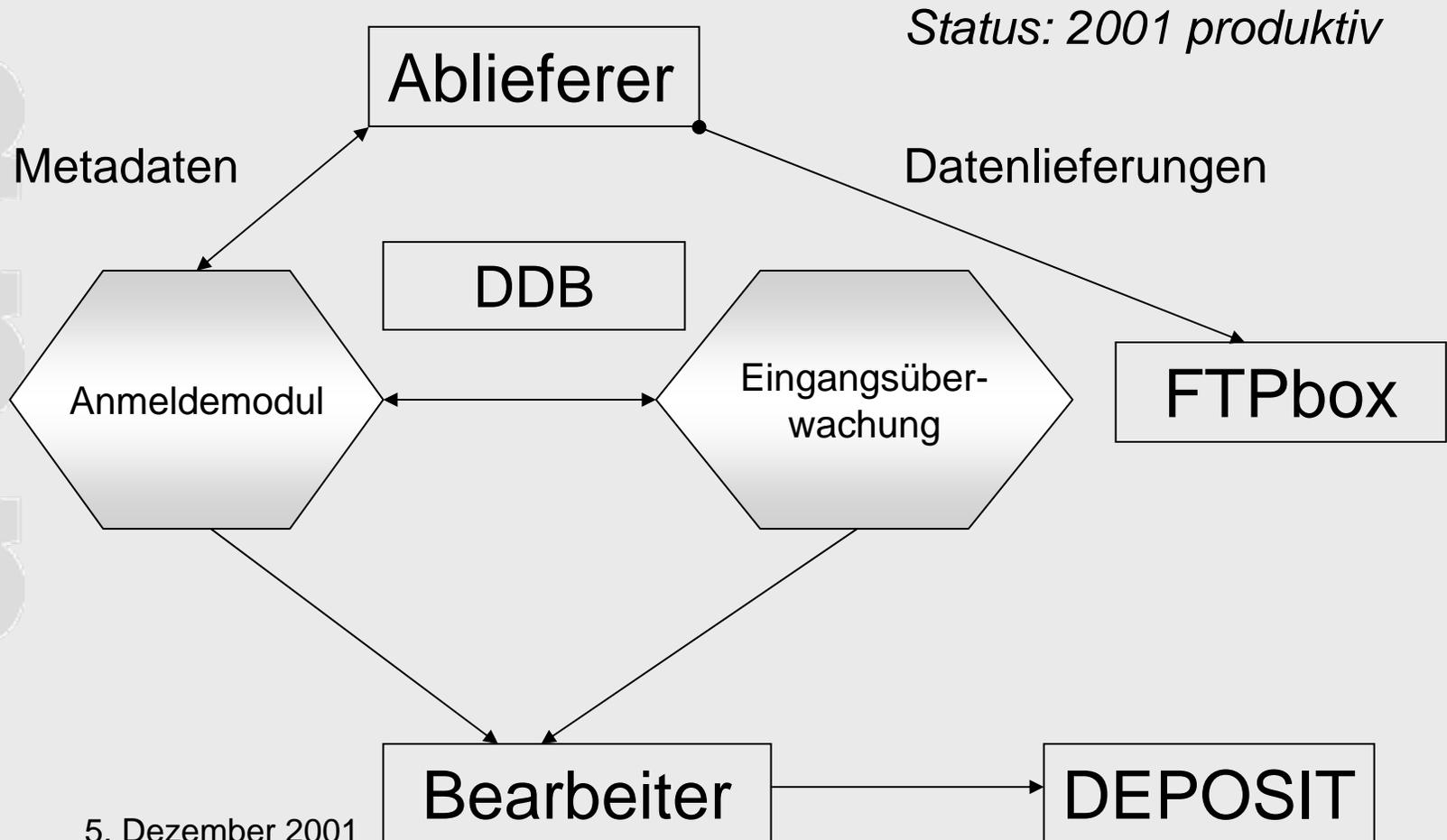


NEDLIBs Systementwurf DSEP



5. Dezember 2001

Die Deutsche Bibliothek - Aktivitäten





Die Deutsche Bibliothek - Aktivitäten Archivierungsstatus

Metadaten – der Archivierungsstatus

a = archiviert

b = Archivierungsabsicht besteht

c = Archivierungsabsicht besteht nicht

n = Archivierung nicht möglich

e = wechselnder Archivierungsstatus





Die Deutsche Bibliothek - Aktivitäten Beispiel: Springer

- Pilotprojekt mit dem Springer-Verlag
 - Sicherstellung der Übertragbarkeit und Nachnutzung

Stand der Arbeiten (9/2001)

- Lieferung von 416 Zeitschriften auf CD-ROM
- Entpackung von ca. 1.600 Archivdateien
- Speicherung von 1,8 Mill. Einzeldateien
- Ca. 70 GB RAID-Speicher



Warum Persistent Identifier?

- Zugriff auf digitale Ressourcen
 - Sicherheit, Globalität, Eindeutigkeit, Automatischer Ablauf
- Verwaltung digitaler Ressourcen
 - Sicherstellung der Dauerhaftigkeit, Archivierung
- Schutz und Verwaltung von Rechten
 - Authentifizierung der Publikation und des Urhebers, Nachweis von Rechten

Persistent Identifier URN

- Kontrolle durch „Internet Engineering Task Force“ (IETF)
- basiert auf hierarchischem Aufbau von Teilbereichen
- Struktur: **URN:NID:SNID-NIS**
 - NID: Namensraum (Namespace Identifier)
 - Beispiele: ISBN, ISSN und NBN
 - SNID: Teilbereich aus dem Namespace (Subnamespace oder Sub-Prefix)
 - NISS: Namespace Specific String



**Globale Registrierungsinstanz:
„Library of Congress“ **urn:nbn****

SNID (ISO-
Ländercode)

**National-
bibliothek 1**

**National-
bibliothek 2**

...

DDB
urn:nbn:de...

Institution 1
urn:nbn:de:xyz...

Institution 2

...

SNID-Vergabe





Global Info Sonderfördermaßnahme: CARMEN

- AP 4: Persistent Identifiers and Metadata Management in Science
 - Testlauf zur dauerhaften und eindeutigen Identifikation von Online-Publikationen durch Persistent Identifier
 - Vergabe und Verwaltung der Persistent Identifier
 - Anwendung verschiedener Resolving-Mechanismen



CARMEN AP 4 DDB-Teilprojekt

- Vergabe und Verwaltung von „NBN:DE“
(Seit 9/2001)
 - Zunächst am Beispiel von Online-Dissertationen
 - Simulation des Resolving-Mechanismus „Persistent-URL“
 - Teilnehmer: 25 UBs
 - 108 vergebene URNs für 226 URLs



CARMEN AP 4, DDB-Teilprojekt Expertentreffen Bibliotheksverbände

- Gemeinsamer Namensraum

URN:NBN:DE:[SNID1]:[SNID2]-[NISSP]

- **SNID1**: Kürzel der Bibliotheksverbände
- **SNID2**: Bibliothekssigel der UBs
- **NISS**: - fortlaufende Nummer für UBs
- alphanumerischer String für UBs des BSZ
- **P**: - Prüfziffer



DDB-OPAC Beispiel



Titel:

Klonierung und funktionelle Charakterisierung des **Aktinreorganisors p150-Spir** [Elektronische Ressource] / vorgelegt von Ines Maria Otto

Verfasser:

[Otto, Ines Maria](#)

Erscheinungsjahr:

2001

Hochschulschrift:

Würzburg, Univ., Diss., 2001

URL:

[Archivserver Der Deutschen Bibliothek](#)

<http://erl.bibliothek.uni->

[wuerzburg.de/DISS/Biologie/2001/X117840/Otto.pdf](http://erl.bibliothek.uni-wuerzburg.de/DISS/Biologie/2001/X117840/Otto.pdf)

Persistent Identifier: <urn:nbn:de:bvb:20-1178402>

Sachgruppe:

33 Medizin ; 32 Biologie



Frontpage des DDB-Archivserver Beispiel



Archivserver
deposit.ddb.de

Die Deutsche Bibliothek

Autor : Möhring, Silke

Titel : Selektion und Vermehrung selbstinkompatibler Winterrapslinien
(*Brassica napus* L.) im Hinblick auf die Entwicklung von
Hybridsorten

Dissertation : Hannover, Universität, Fachbereich Gartenbau, 2000

URN (NBN) : urn:nbn:de:gbv:089-3304534323



Dokument im Format: pdf
Dateigröße: ca. 1,0 MB
MD5-Fingerprint des Dokuments



Dateibetrachter (Viewer) zur Präsentation einiger Dokumentformate



Informationen zum MD5-Fingerprint

Endnutzer

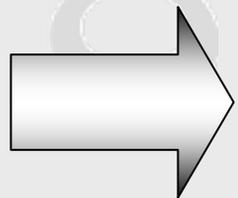
Resolving-Prozess

```
graph LR; A[Endnutzer] --> B[Resolving-Prozess]; B --> C[Online-Dokument]; C --> D[Umleitung einer PI-Anfrage an eine stabile URL];
```

PI-Anfrage

**Verwaltungs-
system**

Online-Dokument



Umleitung einer PI-Anfrage an eine **stabile URL**



CARMEN AP 4, DDB-Teilprojekt Ausblick

- Ausweitung der Verwaltung und Vergabe von Persistent-Identifizier auf
 - Normdaten
 - weitere PI-Systeme (z.B. DOI)
- Vollständige Automatisierung der PI-Verwaltung



Weitere Informationen

- Archivserver DEPOSIT.DDB.DE
 - <http://deposit.ddb.de/>
- NEDLIB
 - <http://www.kb.nl/coop/nedlib/>
- CARMEN, AP 4
 - http://www.bis.uni-oldenburg.de/carmen_ap4/



Vielen Dank für Ihre Aufmerksamkeit !

Berthold Weiß, M.A.

Deutsche Bibliothek Frankfurt am Main

Tel.: 069 / 1525 – 1404

Fax: 069 / 1525 – 1444

E-Mail: weissb@dbf.ddb.de