



Die Open Archive Initiative

Dr. Bruno Klotz-Berendes
Universitätsbibliothek Dortmund

6. Inetbibtagung, Göttingen, 18. - 20.09.2002





Gliederung des Vortages

- Ziele und Entstehung von OAI
- Datenprovider - eine Zwischenbilanz
- DINI Empfehlungen für OAI Datenprovider
- Serviceprovider



Ziele von OAI

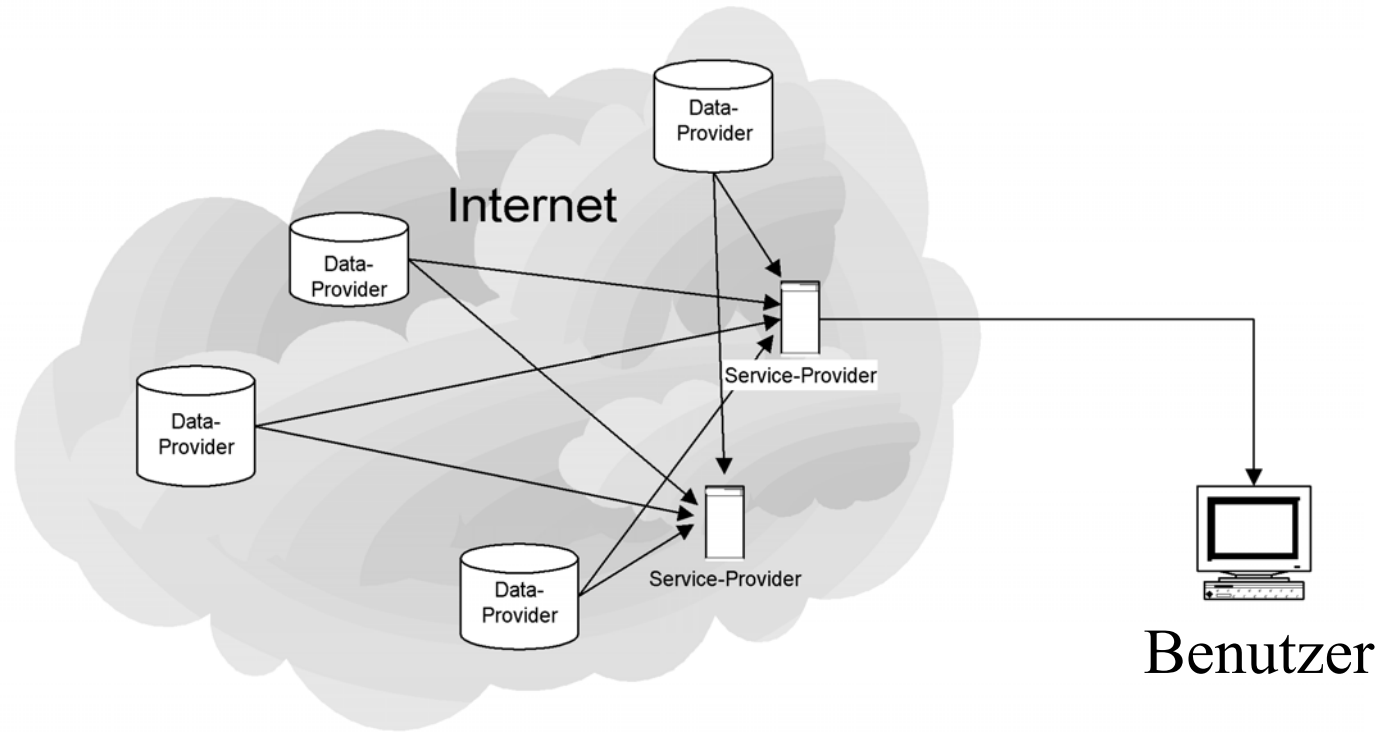
„The Open Archives Initiative Protocol for Metadata Harvesting provides an application-independent interoperability framework based on *metadata harvesting*.“

In der OAI Community gibt es zwei unterschiedliche Partner:

- Datenprovider - diese stellen die Metadaten zur Verfügung
- Serviceprovider - diese sammeln die Metadaten ein und entwickeln auf dieser Basis neue Dienstleistungen



Zusammenwirken von Daten - und Service Providern



Copyright Bernd Diekmann



Ziele von OAI

Datenprovider:

- Zugang der Metadaten
- **Nicht** notwendigerweise freier Zugriff auf Volltexte

Serviceprovider:

- Harvesting - Einsammeln der Metadaten
- Aufbereiten und Analysieren der Metadaten
- Suche, Navigation, Alerting-Dienste u.a. Informationsdienstleistungen



Ziele von OAI

- Effiziente Verbreitung von Inhalten
- Unabhängig von den angebotenen Inhalten und der technische Lösung
- freier Zugriff auf diese Archive / Metadaten
- einheitliche Schnittstellen für Archive und Serviceanbieter



Organisationsstruktur von OAI

Steering Committee - besteht aus 12 Vertretern aus verschiedenen wissenschaftlichen Institutionen

politische Weiterentwicklung, richtungsweisende Diskussion und Promotion

Executive Committee - C. Lagoze u. H. Van de Sompel
Koordination der Aktivitäten

Technical Committee - Evaluierung und Weiterentwicklung der OAI - Architektur, basierend auf Erfahrungen der Anwender

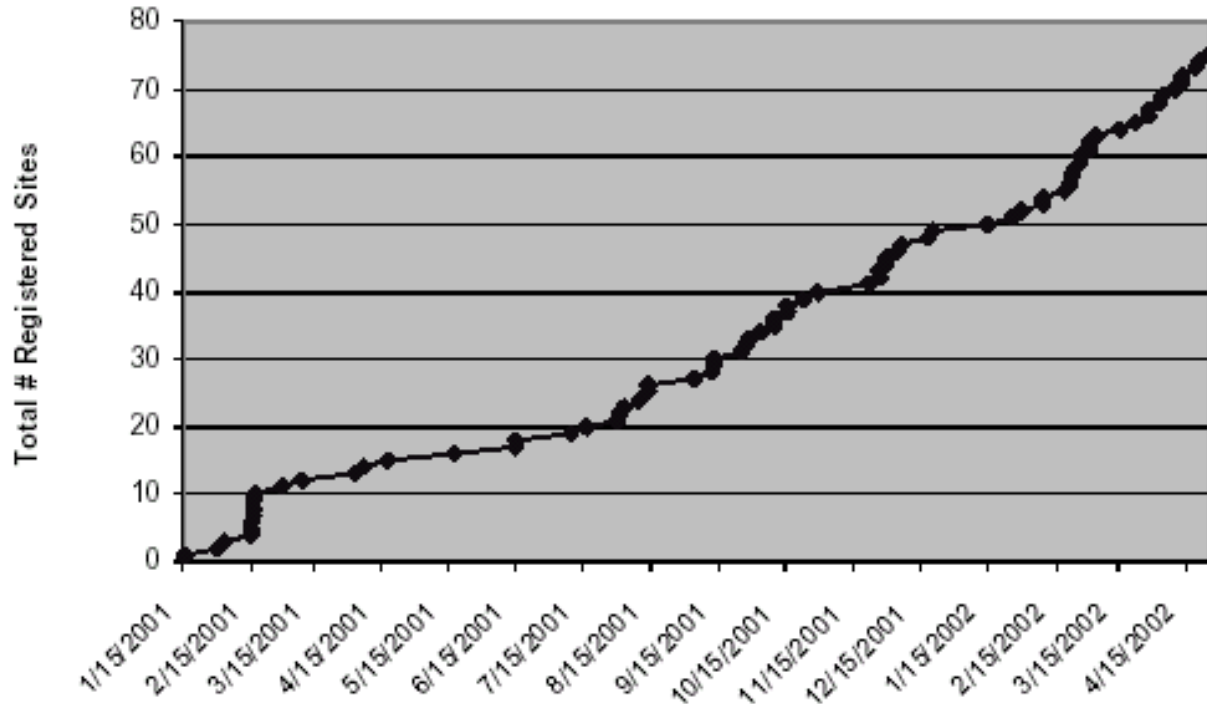


Gründe für die Akzeptanz

- leicht auf einem bestehenden Dokumentenserver zu implementieren
- geringe Mindestanforderungen an den Metadatensatz (mindestens Dublin Core)
- Nutzung akzeptierter Webtechniken (HTTP u. XML)
- frei verfügbare Tools für die Implementierung
- komplette Publikationslösungen



Zunahme der Datenprovider



Quelle: H. Van de Sompel, C. Lagoze, Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative, ECDL 2002, Rom

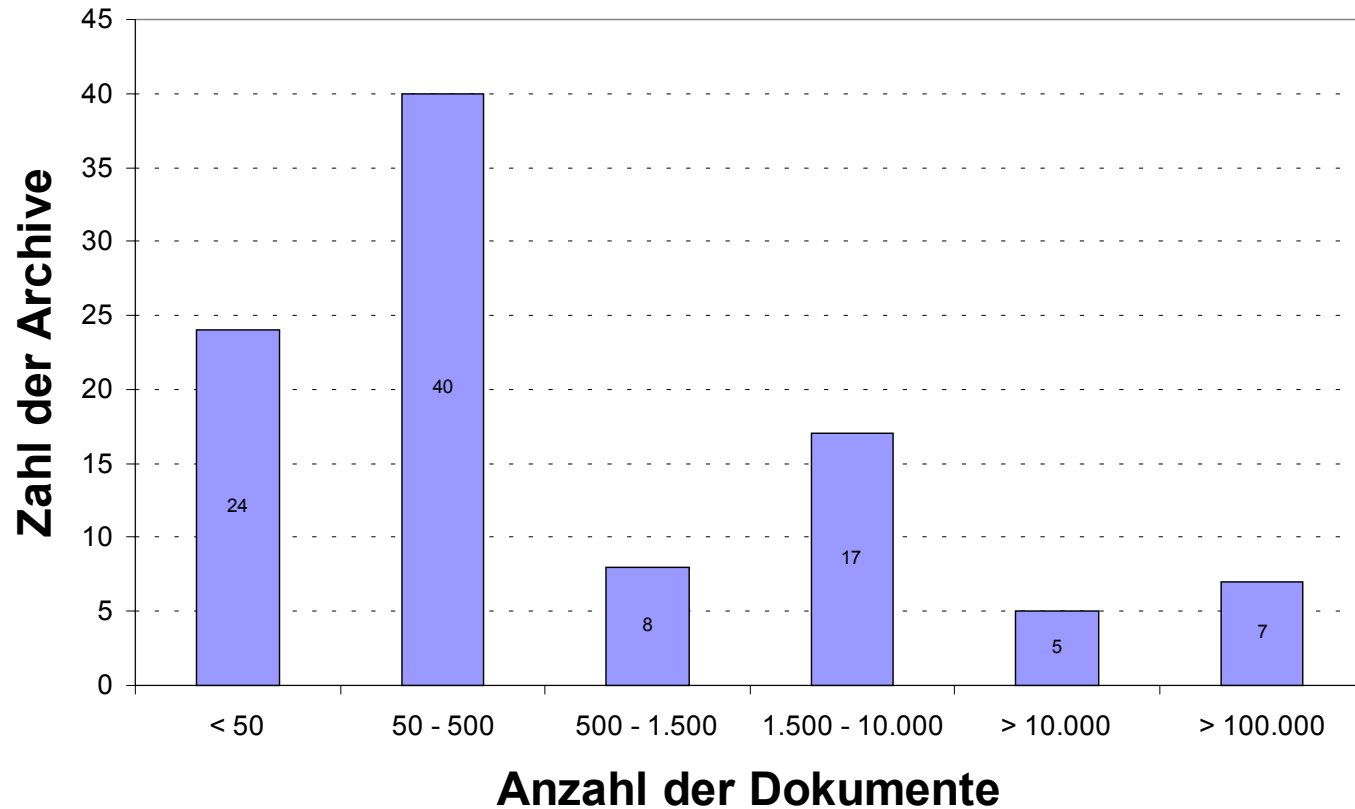


Datenprovider - ein Überblick

- Zur Zeit sind weltweit rund 100 Archive mit weit über 1 Million Dokumenten bei OAI registriert. (Stand 12.09.2002)
- Eine vollständige Liste findet man unter:
<http://www.openarchives.org/Register/BrowseSites.pl>
- Eine Liste der deutschen Datenprovider ist unter
<http://www.dini.de/dinioai/oaidp.php> zu finden.



OAI - Archive





OAI Projekte

Aus der inzwischen großen Anzahl an Projekten seien hier nur einige aufgeführt:

- National Science Digital Library
- Portallösungen gefördert durch Mellonfoundation
- Open Language Archives Community (OLAC)
- Open Archive Forum
- Cyclades
- In NRW ?



Tools

finden sich unter: <http://www.openarchives.org/tools/index.html>

Diese Tools gliedern sich in

- OAI Repository Explorer
- OAI-Specific Tools
- Related Utilities

Auch bei DINI gibt es eine kleine Sammlung:

<http://www.dini.de/dinioai/oaitechnik.php>



Minimal Server Implementation

The algorithm for [oai1.pl](#) is simply:

- read GET, POST or command line request

- check syntax of request

- if syntax correct

- return XML reply to request

- else

- return error code and message

(Warner, Simeon: Exposing and Harvesting Metadata Using the OAI Metadata Harvesting Protocol in: HEP Libraries Webzine Issue 5 / June 2001)



Requests

Insgesamt sechs Requests müssen beantwortet werden:

- Identify
- List Metadata Format
- List Sets
- List Identifiers
- List Records
- Get Records



Metadatenformat

Verpflichtend:

- unqualified Dublin Core

Weitere sind möglich:

- MARC
- OLAC (Open Language Archive Community)

Internationale Entwicklungen berücksichtigen.



Setdefinition

„A set is an optional construct for grouping items for the purpose of selective harvesting. Repositories may organize items into sets.“ (aus der Protokollbeschreibung)

- Suche nach inhaltlichen und formalen Kriterien nur über Sets möglich.

Eine inhaltliche Strukturierung ist nur mit Sets möglich.

- Aufbau von Hierarchien
- Dokumente können in mehreren Sets vorkommen



Erschließung durch Sets

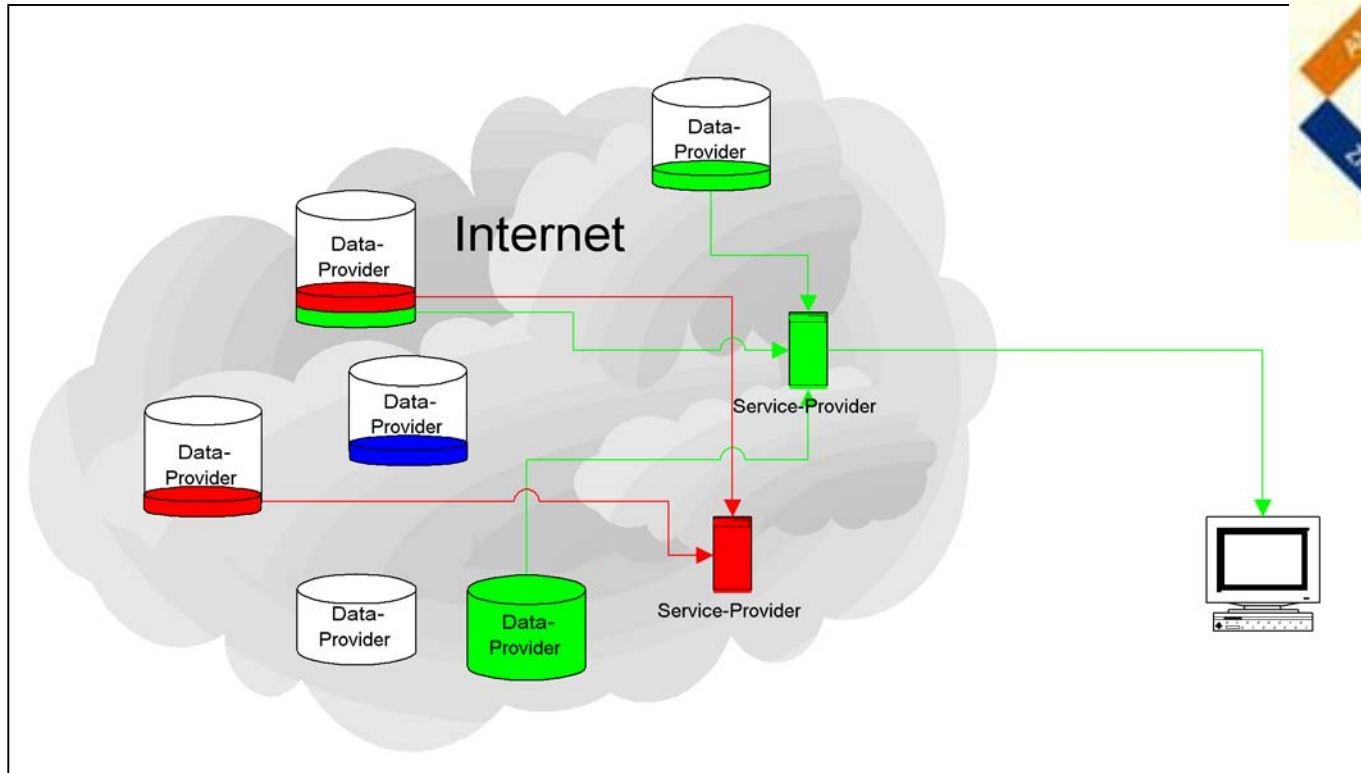
Empfehlungen der Arbeitsgruppe bei DINI

- erleichtern den Aufbau von Serviceprovidern
 - Fachliche Ausrichtung
 - Formale Ausrichtung





Benutzung von Sets



Empfehlungen der DINI Arbeitsgruppe



Die DINI Arbeitsgruppe empfiehlt die Einrichtung von drei unterschiedlichen Sets:

- Inhalt - fachliche Zuordnung
- Publikationstyp - formale Zuordnung
- Dokumentationstyp - formale Zuordnung





Set inhaltliche Beschreibung

Eine grobe inhaltliche Einordnung ermöglicht eine fachliche Strukturierung und Selektionsmöglichkeit.

➤ **DNB Sachgruppen**

Die Bezeichnung des jeweiligen Elements sollte über die Sachgruppennummer erfolgen.

Spezifikation (SetSpec): dnb:Nummer

Bezeichnung (SetName): Englisch

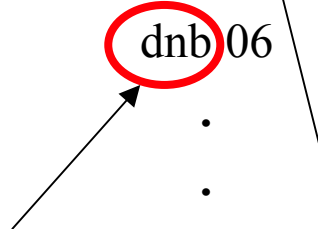




Sachgruppen der Deutschen Nationalbibliographie als inhaltliches Set

SetSpec	SetName
dnb:01	Knowledge and Culture in General
dnb:02	Books and Libraries, Information and Documentation
dnb:03	Reference Books, Bibliographies
dnb:04	Directories and Phone Books
dnb:05	Calendars
dnb:06	Journalism
.	.
.	.
.	.

Quelle



Ordnungskriterium

Bezeichnung





Formaler Set

Publikationstyp

SetSpec	SetName
pub-type	Documents having a formal publicationtype
pub-type:monograph	Books, Monographs
pub-type:article	Journal Articles
pub-type:dissertation	Dissertations and Professional Dissertations

Dokumentationstyp

SetSpec	SetName
doc-type	formal document-type
doc-type:text	Text
doc-type:image	Image
doc-type:audio	Audio

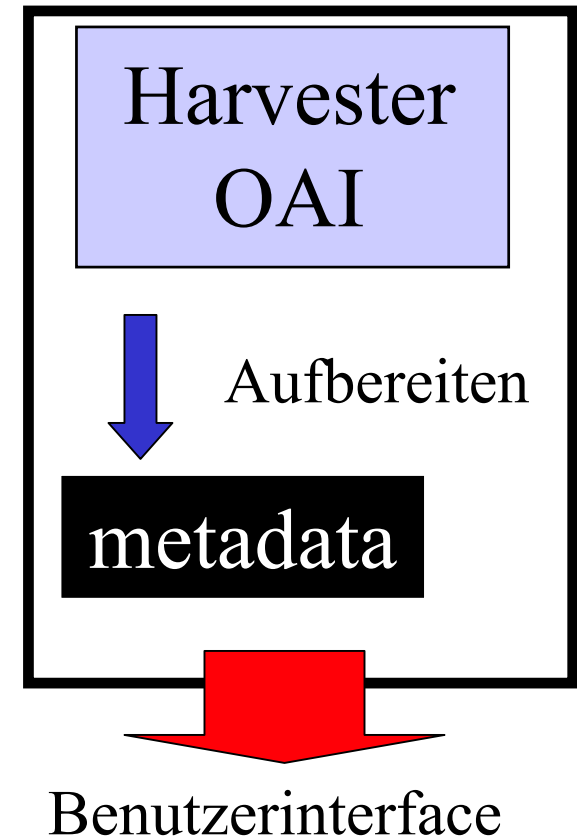




Serviceprovider

- Sammeln der Daten - OAI
- Analysieren der eingesammelten Daten
- Aufbereitung der gefundenen Dokumente
 - Zitatanalyse
 - automatisches Klassifizieren
 - Browsingfunktion
 -

Serviceprovider





Beispiele für Serviceprovider

Ein einfaches Suchinterface ohne weitere Funktionen bietet der von der OLD Dominion University entwickelte Serviceprovider ARC.

➤ ARC <http://arc.cs.odu.edu/>

– suchbare Datensätze: 1,3 Mill. Dokumente bei 100 Archiven

Der Serviceprovider, der unter MyOAI zu finden ist, hat dagegen diesen Dienst in ein ausgeweitetes Angebot eingebettet.

➤ MyOAI www.MyOAI.org

**Eine Liste der aktuell registrierten Serviceprovider findet man unter:
<http://www.openarchives.org/service/listproviders.html>**



Informationen zu OAI

- **<http://www.openarchives.org>**
- **Mailinglisten**
 - oai-general
 - oai-implementers
- **<http://www.dini.de> - Punkt DINI-OAI**
(Koordination der OAI-Aktivitäten in Deutschland)



Danksagung

Meinen Mitstreitern in der DINI Arbeitsgruppe
Susanne Dobratz HU Berlin,
Frank Scholz UB Stuttgart und
Bernd Diekmann BIS Oldenburg



sowie meinen Mitarbeiterinnen Frau Jürgen und Frau Puppe

und Ihnen Danke ich für Ihre Aufmerksamkeit !