



Die Open Archiv Initiative

Workshop

Dr. Bruno Klotz-Berendes
Universitätsbibliothek Dortmund

6. Inetbibtagung, Göttingen, 18. - 20.09.2002





Gliederung des Workshops

- Ziele und Entstehung von OAI
- Grundlagen des Protokolls
- Datenprovider
- Serviceprovider



Ziele von OAI

The Open Archives Initiative has been set up to create a forum to discuss and solve matters of interoperability between preprint solutions, as a way to promote their global acceptance.

Paul Ginsparg, Rick Luce & Herbert Van de Sompel



Ziele von OAI

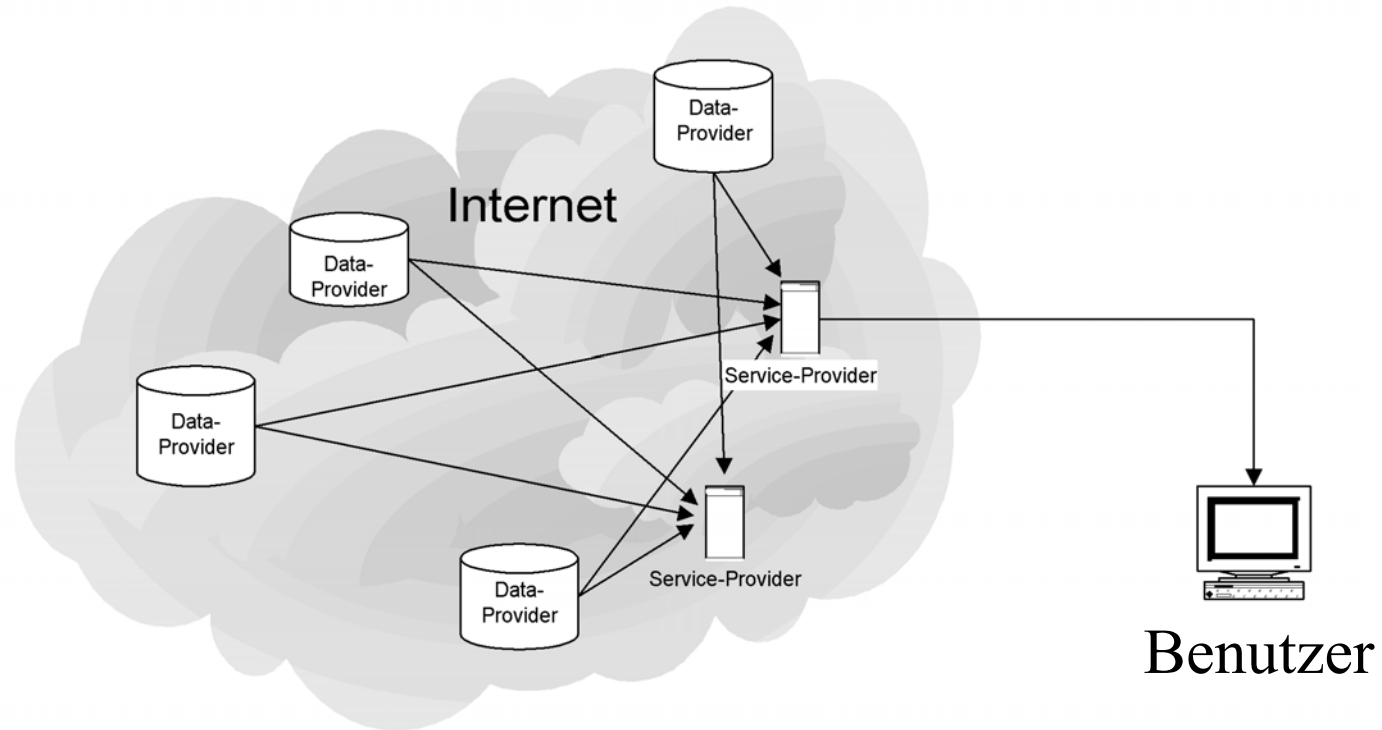
„The Open Archives Initiative Protocol for Metadata Harvesting provides an application-independent interoperability framework based on *metadata harvesting*.“

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



Budapester Open Access Initiative

Für den freien Zugang zu wissenschaftlichen Informationen sucht die Budapester Open Access Initiative Unterstützung.

<http://www.soros.org/openaccess/index.shtml>



Entstehung von OAI

- **Oktober 1999 Santa Fee Meeting**
- **Mai, Juni 2000 Alpha-Version**
- **Juli 2000 Beta-Version für Erstanwender**
- **Januar 2001 Version 1.0**
- **23. Februar Vorstellung in Berlin**
- **Veröffentlichung der Version 1.1 Juni/Juli 2001**
- **März 2002 Protokollversion 2.0 im Alphatest**
- **April 2002 Betaversion**
- **Juli 2002 Protokollversion 2.0**



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



Informationen zu OAI

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



Grundlagen des OAI Protokolls

Im Rahmen von OAI unterscheidet man zwischen

Datenprovidern, d. h. den Dokumentenservern
und den

Serviceprovidern, welche die Suchfunktionalität aufbauen.



OAI-Grundregeln

- Definition einer Policy
- Nutzung des Dublin Core Metadatensatz
- OAI-spezifische XML-Syntax
- netzbasiertes Serversystem,
- spezielles Open Archive Protokoll auf HTTP-Basis



OAI Protokoll

basiert auf http-Protokoll

- als cgi script implementieren

Auswahl von sets, die abgefragt werden können (Archivspezifisch)

- :dissertationen
- :cardiology

Ausgewählte Anfragen

- ListSets, ListRecords, ListIdentifier....

Resumption Token



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)



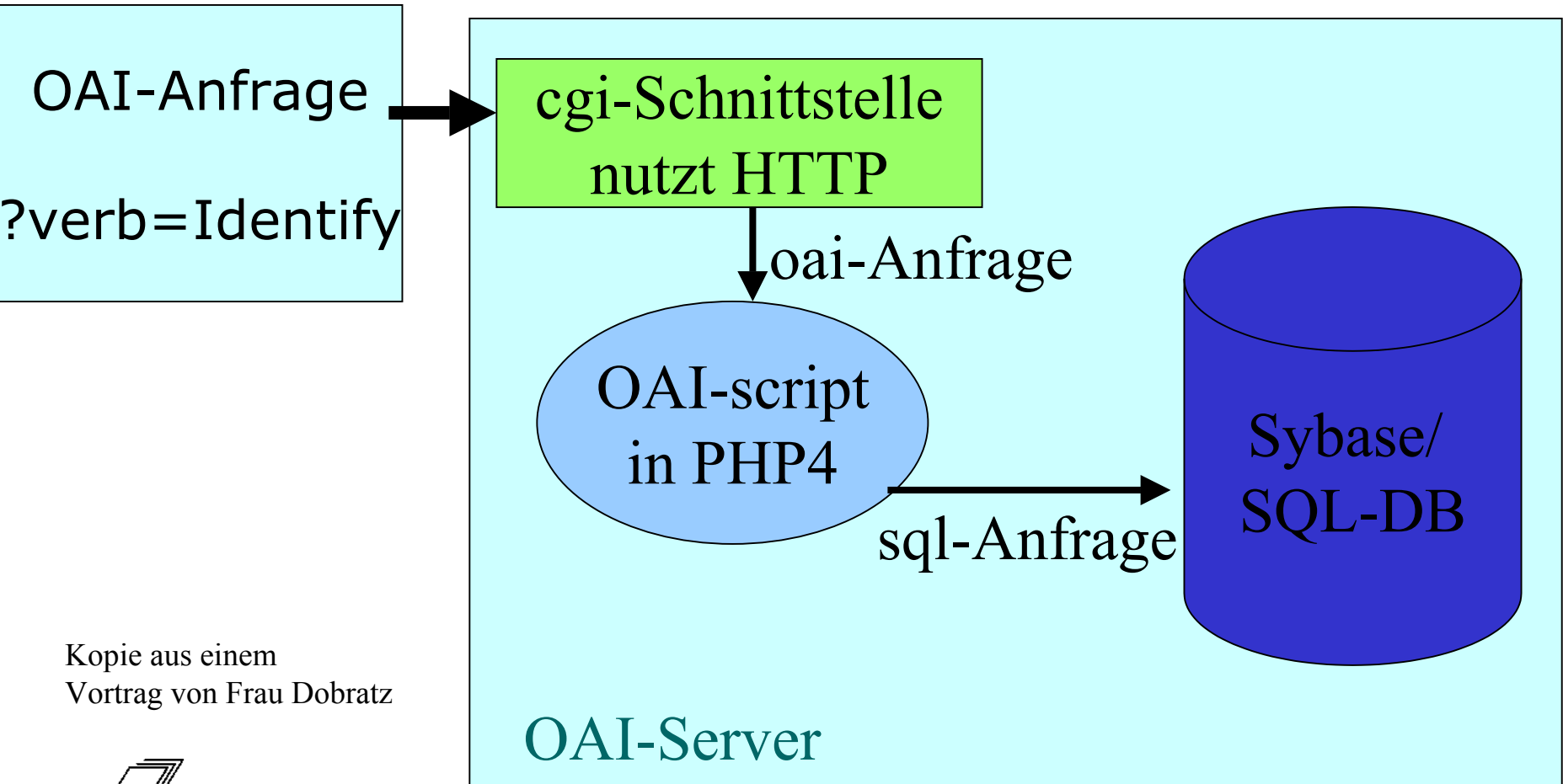
Requests

Insgesamt sechs Requests müssen beantwortet werden:

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



Die Implementation der HU Berlin



Kopie aus einem Vortrag von Frau Dobratz

OAI-Server



OAI Records

Header

- Eindeutiger Identifier: Schlüssel für Archivanfragen
`oai:HUBerlin:dissertationen:kemps-christof-2000-06-18`
- Datestamp: Datum der Erstellung, des Löschens, der letzten Modifikation eines Items `2001-06-18`
- Verwendung von Sets

Metadaten

- Metadatenprefix
- Dublin Core ohne Qualifier!



Datenprovider

In diesem Abschnitt des Workshops sollen die folgenden Punkte besprochen werden:

- Was zeichnet einen Datenprovider aus?
- Wie wird aus einem Dokumentenserver ein Datenprovider?
 - Der Repository Explorer?
 - Empfehlungen der DINI-Arbeitsgruppe zur Setbenennung
 - Fertige Lösungen

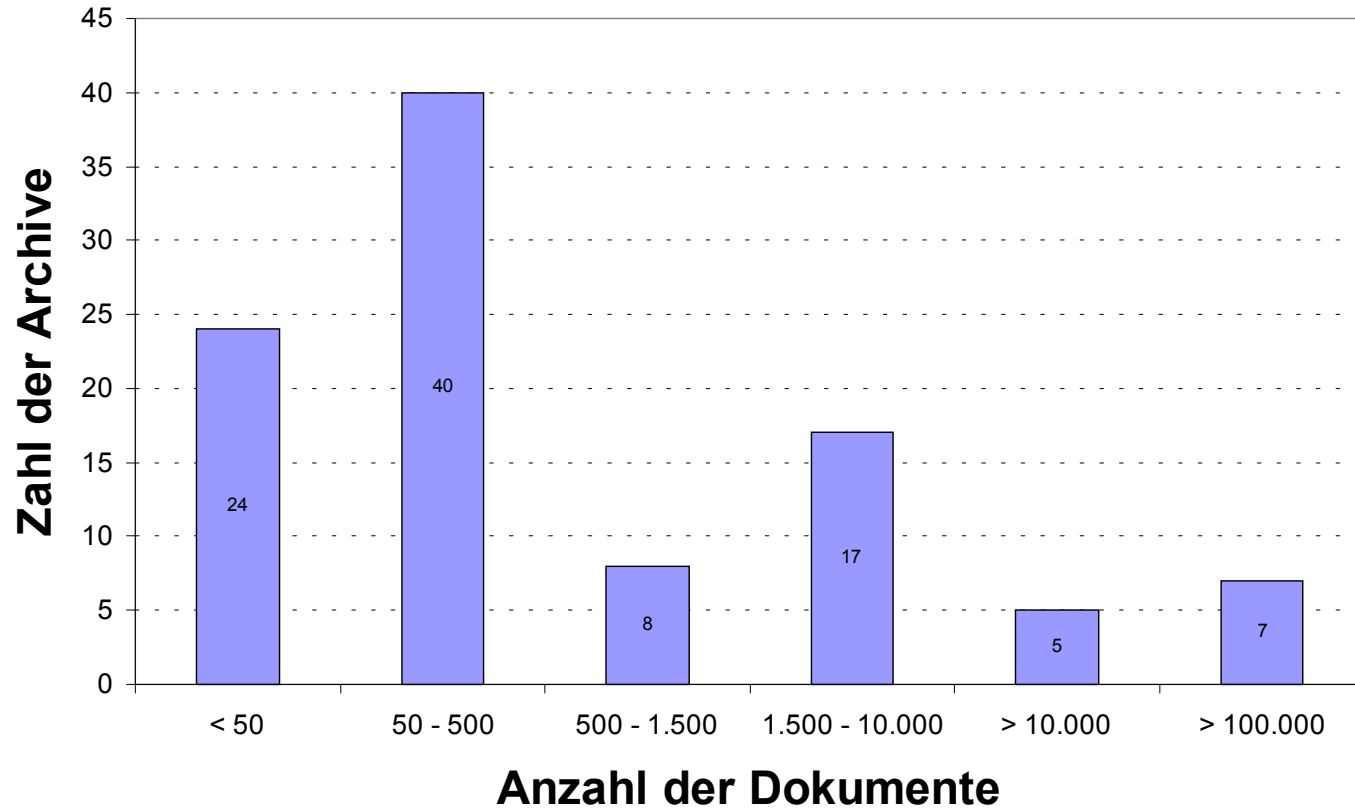


Datenprovider - ein Überblick

- 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





Datenprovider

Wodurch unterscheidet sich ein OAI Datenprovider von einem normalen Dokumentenserver?

- **definierte Schnittstelle**
- **Mindestsatz an Metadaten**
- **Metadaten dürfen benutzt werden**
- **Persistent identifier**

Anbindung von Datenbanken möglich, die nicht von Suchmaschinen erfasst werden oder erfasst werden dürfen.



Vorteile für den Dokumentenserver

- Nutzung des Archivs unter verschiedenen Gesichtspunkten
 - Inhaltliche Strukturierung (Virtuelle Fachbibliotheken)
 - Regionale Aspekte
 - Formale Gliederung (Suchmaschine für Dissertationen)
- Metadatenutzung
- Nachweis der eigenen Publikationen in einem internationalen Kontext
- Regelung der Serverlast möglich



1. Schritt

Analyse der gegebenen Situation

- Hat die Institution schon einen Dokumentenserver?
 - Wie sieht diese Lösung aus?
 - Kommerzielles Produkt
 - Eigenentwicklung

Welche Ressourcen stehen zur Verfügung? Kooperieren!

Die OAI Community hat fertige Komplettlösungen entwickelt. Die Universitätsbibliothek Duisburg hat damit gute Erfahrungen gemacht.

Wir haben ein Perltool an unsere Lösung angepasst.



Fertige Lösungen

Open Source Lösungen gibt es inzwischen von:

- Virginia Tech University (an der UB Duisburg)
- Eprint.

Vorteile:

- Schnelle Umstellung auf OAI
- kostenlos

Nachteil:

- Anpassung an die lokalen Begebenheiten



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>



Implementierung des OAI-Protokolls an der UB Dortmund

Eldorado - Struktur der Daten in HIS

- **Objekte und Objektinformationen separat gespeichert**
- **Objektinformationen als „Hash“ von Attributen und Werten**



Beispiel UB Dortmund

Objektinformationen bestehen aus:

- **nicht zu editierenden Systemattributen**
 - TimeCreated, TimeModified
 - Persistent Identifier (GoiD) u.a.
- **editierbaren Attributen**
 - Rights (wie unter Unix)
 - Owner
 - Title
 - Name u. a.
- **beliebigen Attributen**
 - man kann beliebige Attribute hinzufügen

Auf die Objektinformationen kann über die Objekt-ID oder Suchanfragen zugegriffen werden.



Objektinformationen: Auszug aus einem „Attributhash“

ObjectID=0x00034a63
Type=Document
DocumentType=collection
TimeCreated=2001/10/25 12:22:16
TimeModified=2002/01/10 12:08:25
Rights=W:g eldorado konvertierung
Name=FB4/ls12/forschung/2001/Bashford
CollectionType=MultiCluster
DC.Contributor.Advisor=Peter Marwedel
DC.Contributor.Referee=Peter Padawitz
DC.Creator.PersonalName=Bashford, Steven
DC.Creator.PersonalName.Address_Post=Buchackerstr. 3a, 63867 Breunberg
DC.Date=2001-02-21
DC.Format=application/pdf
DC.Language=GER
DC.Language.Translated=ENG
DC.Title=Constraintbasierte Codegenerierung für eingebettete Prozessoren
DC.Type=Dissertation
DC.Identifier=http://eldorado.uni-dortmund.de:8080/0x81d98002_0x00034a63
DC.Description=Eingebettete Systeme gewinnen zunehmenden Einfluss ...
DC.Subject=Code-Generierung; Compiler; Phasenkupplung; Code-Selektion; ...
OAI=yes
OAI-Set=28
Subdocs=7
GOid=0x81d98002 0x00034a63



Beispiel der UB Dortmund

Entscheidung für PERL Data Provider Template VTET

aufgrund der Abfragemöglichkeiten in HIS 5.5

- „Abfragesprache nicht sehr mächtig“
- schlechte Performanz bei komplexen Abfragen
- nur der komplette Attributhash läßt sich schnell und einfach abfragen

Entscheidung für das Perl-Skript als single file version

- einfach und überschaubar
- geringer Anpassungsaufwand
- geringer Pflegeaufwand
- gleicht die geringe Mächtigkeit der HIS-Queries aus (reg ex, schnelles parsen des kompletten Hashs)



Beispiel der UB Dortmund

Vorbereitende Arbeiten

Erweiterung der bisher vorhandenen Metadaten um folgende Attribute:

- OAI yes/no
- OAI-set Setdefinition (zur Zeit DNB-Sachgruppe)
- metadata code der festlegt, welchen „Zustand“ die Metadaten haben

Daten für den Identify-Request festlegen

- name
- admin email
- content
- metadataPolicy
- dataPolicy
- submissionPolicy



Beispiel UB Dortmund

Anzupassende „Teilbereiche“

- Variablendeklaration
- Auswahl eines oder mehrerer Metadatenformate
- Festlegung der Set-Definition
- Datenbankabfragen



Struktur des Skripts von VTET

Funktionen

muß man anpassen

Variablendeklaration	ja
sub pclean	nein
sub lclean	nein
sub straighten	nein
sub ErrorExit	nein
sub MetadataFormatisValid	nein
sub DateisValid	nein
sub xmlheader	nein
sub xmlfooter	nein
sub Identify	ja
sub ListMetadataFormats	nein
sub ListSets	ja
sub FormatXML	nein
sub FormatRecord	nein
sub GetRecord	ja
sub ListRecords	ja
sub ListIdentifiers	ja
sub main	nein



2. Schritt

Wie ist mein vorhandenes Archiv aufgebaut?

- Handelt es sich um eine Datenbanklösung?
- Wie wird die Datenbank vom Webserver angesprochen?
- Wie sind die Metadaten mit den Volltexten verbunden?
- Gibt es einen elektronischen Geschäftsgang, in dem das Archiv eingebunden ist.
- Wie erfolgt die Meldung an die DDB?
- Wie groß sind die Eingriffsmöglichkeiten in das System?
- Welchem Ziel dient das Archiv?



Datenbanklösung mit Webserver

Wie wird die Datenbank an den Webserver angebunden?

- Bei PHP würde sich die Anpassung der Berliner Lösung an die lokalen Begebenheiten anbieten. Das Skript findet sich auf den Seiten von DINI.

Wie sind die Metadaten in der Datenbank abgelegt? Kann ich beliebige Ausgaben erzeugen. Komme ich an alle benötigten Daten?

- Anpassung der Perllösung der Vergina Tech an die lokalen Gegebenheiten - unsere Lösung



Eigenlösung

Entwicklung eines eigenen Tools

- Miles Lösung - Im Rahmen der MyCore Initiative wird eine javabasierte Lösung entwickelt.



Wie geht es weiter?

Auf den Webseiten von OAI gibt es ein Testmodul für die Anpassung des eigenen Archivs, welches sehr zu empfehlen ist.

Zu Beginn müssen einige Parameter festgelegt werden.



Wie werde ich Datenprovider?

- Wer ist der Ansprechpartner?
- Wie soll das Archiv heißen?
- Festlegen der URL für die OAI-Schnittstelle
- Strukturierung des Archivs durch Sets



Open Archives Initiative - Repository Explorer

explorer version - 1.45 : protocol version - 1.0/1.1/2.0 : June 2002

This site presents an interface to interactively test archives for compliance with the OAI Protocol for Metadata Harvesting [[Click here for details](#)]

JavaScript is required

Note: To avoid HTTP errors, please wait for each page to finish loading before clicking on any link.

Please enter the URL to the OAI interface (everything before the ?) or choose a predefined archive from the table below. You can click on a verb from the list below to test that function (entering parameters as necessary)

URL :

- American Memory [LoC]
- ...Kiv
- ...burn University - Transforming America
- ...lestial - Cache/Aggregator

[New Archive Website](#) || [Test and Add an archive to this list](#)]

Verbs	Parameters
Identify List Metadata Formats List Sets Get Identifiers	from (eg., YYYY-MM-DD) : <input type="text"/> until (eg., YYYY-MM-DD) : <input type="text"/> metadataPrefix : <input type="text"/> ... : <input type="text"/>

Click on a verb from the list below to test that function (entering parameters as necessary)

URL :

- American Memory [LoC]
- Arkiv
- Burn University - Transforming America
- Celestial - Cache/Aggregator

[View Archive Website](#) || [Test and Add an archive to this list](#)]

Verbs	Parameters
Identify List Metadata Formats List Sets List Identifiers List Records List Record	from (eg., YYYY-MM-DD) : <input type="text"/> until (eg., YYYY-MM-DD) : <input type="text"/> metadataPrefix : <input type="text"/> identifier : <input type="text"/> set : <input type="text"/> resumptionToken : <input type="text"/>

Language	Display	Schema Validation
<input type="text" value="English"/>	<input checked="" type="radio"/> Parsed <input type="radio"/> Raw XML <input type="radio"/> Both	<input type="radio"/> None <input checked="" type="radio"/> Local mirror of schemata (Xerces) <input type="radio"/> Online schemata (Xerces) <input type="radio"/> Local mirror of schemata (XSV) <input type="radio"/> Online schemata (XSV)



Open Archives Initiative - Repository Explorer

explorer version - 1.45 : protocol version - 1.0/1.1/2.0 : June 2002

http://memory.loc.gov/cgi-bin/oai2_0?verb=Identify

Archive details : <http://memory.loc.gov/>

Parsed Output

Archive Self-Description

Repository Name	Library of Congress Open Archive Initiative Repository 1
Base URL	http://memory.loc.gov/cgi-bin/oai2_0
Protocol Version	2.0
Admin Email	dwoo@loc.gov
Admin Email	caar@loc.gov
Earliest Datestamp	2002-06-01T00:00:00Z
Deleted Record Handling	no
Granularity	YYYY-MM-DDThh:mm:ssZ
	description: oai-identifier:



Open Archives Initiative - Repository Explorer

explorer version - 1.45 : protocol version - 1.0/1.1/2.0 : June 2002

http://memory.loc.gov/cgi-bin/oai2_0?verb=ListMetadataFormats

Archive details : <http://memory.loc.gov/>

List of Metadata Formats

Click on the link to view schema

fix=[oai_dc]
meSpace=[http://www.openarchives.org/OAI/2.0/oai_dc]
schema=[http://www.openarchives.org/OAI/2.0/oai_dc.xsd]

fix=[oai_marc]
meSpace=[http://www.openarchives.org/OAI/1.1/oai_marc]
schema=[http://www.openarchives.org/OAI/1.1/oai_marc.xsd]

fix=[marc21]
meSpace=[<http://www.loc.gov/MARC21/slim>]
schema=[<http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>]

[not a standard OAI metadata name]

request : http://memory.loc.gov/cgi-bin/oai2_0, verb=ListMetadataFormats
response Date : 2002-09-12T10:07:53Z

```
schema targetNamespace="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified" attributeFormDefault="unqualified">
  <annotation>
  <documentation>
    XML Schema 2002-03-18 by Pete Johnston.
    Adjusted for usage in the OAI-PMH.
    Schema imports the Dublin Core elements from the DCMI schema for unqualified Dublin Core.
  </documentation>
  </annotation>
  <import namespace="http://purl.org/dc/elements/1.1/"
    schemaLocation="http://dublincore.org/schemas/xmls/simpledc20020312.xsd"/>
  <element name="dc" type="oai_dc:oai_dcType"/>
  <complexType name="oai_dcType">
  <choice minOccurs="0" maxOccurs="unbounded">
    <element ref="dc:title" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:creator" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:subject" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:description" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:publisher" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:contributor" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:date" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:type" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:format" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:identifier" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:source" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:language" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:relation" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:coverage" minOccurs="0" maxOccurs="unbounded"/>
    <element ref="dc:rights" minOccurs="0" maxOccurs="unbounded"/>
  </choice>
  </complexType>
</schema>
```

```

chema xmlns="http://www.w3.org/2001/XMLSchema"
targetNamespace="http://www.openarchives.org/OAI/1.1/oai_marc"
xmlns:oai_marc="http://www.openarchives.org/OAI/1.1/oai_marc"
elementFormDefault="qualified"
attributeFormDefault="unqualified">

```

```

<annotation>
  <documentation>
    Schema for MARC metadata format.
    MARC semantics available at http://www.loc.gov/marc/
    .....
    This Schema has been successfully applied for MARC21 records. It is likely to also work
    for older versions of USMARC and CANMARC. Application of this Schema for other MARC
    formats has not been tested and may require some adjustments.
    .....
    the Open Archives Initiative. 2000.
    Herbert Van de Sompel
    MARC XML transportation format on which this schema is inspired
      available at http://www.dlib.vt.edu/projects/Oai/marcxml/marcxml.html
    This Schema validated at http://www.w3.org/2001/03/webdata/xsv on 05-09-2001.
  </documentation>
</annotation>

```

```

element name="oai_marc">
  <complexType>
    <sequence>
      <element ref="oai_marc:fixfield" minOccurs="1" maxOccurs="unbounded"/>
      <element ref="oai_marc:varfield" minOccurs="0" maxOccurs="unbounded"/>
    </sequence>
    <attribute name="status" type="string" use="optional"/>
    <attribute name="type" type="string" use="required"/>
    <attribute name="level" type="string" use="required"/>
    <attribute name="ctlType" type="string" use="optional"/>
    <attribute name="charEnc" type="string" use="optional"/>
    <attribute name="encLvl" type="string" use="optional"/>
    <attribute name="catForm" type="string" use="optional"/>
    <attribute name="lrRqrd" type="string" use="optional"/>
  </complexType>
</element>

```

```

element name="fixfield">

```



Open Archives Initiative - Repository Explorer

explorer version - 1.45 : protocol version - 1.0/1.1/2.0 : June 2002

http://memory.loc.gov/cgi-bin/oai2_0?verb=ListSets

Archive details : <http://memory.loc.gov/>

Parsed Output

List of Sets

Click on the link to list the contents

- [Ansel Adams's Photographs of Japanese-American Internment at Manzanar](#)
- [Gotttscho-Schleisner Collection \(Photographs\)](#)
- [LC Dance Instruction Manuals](#)
- [LC Early Motion Pictures](#)
- [LC Maps](#)
- [LC Panoramic Photographs](#)
- [LC Photographs \(selected collections\)](#)
- [LC Printed Ephemera Selections](#)
- [LC Sheet Music](#)
- [Theodor Horydczak Collection \(Photographs\)](#)

Request : http://memory.loc.gov/cgi-bin/oai2_0, verb=ListSets

Response Date : 2002-09-12T10:16:33Z

Raw XML Output

```

<?xml version="1.0" encoding="UTF-8"?>
  <OAI-PMH xmlns="http://www.openarchives.org/OAI/1.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/1.0
      http://www.openarchives.org/OAI/1.0/OAI-PMH.xsd">
    <responseDate>2002-09-12T10:16:33Z</responseDate>
    <request verb="ListSets">http://memory.loc.gov/cgi-bin/oai2_0?verb=ListSets</request>
    <ListSets>
      <set>
        <setSpec>manz</setSpec>
        <setName>Ansel Adams's Photographs of Japanese-American Internment at Manzanar</setName>
      </set>
      <set>
        <setSpec>gotttscho</setSpec>
        <setName>Gotttscho-Schleisner Collection (Photographs)</setName>
      </set>
      <set>
        <setSpec>musdibib</setSpec>
        <setName>LC Dance Instruction Manuals</setName>
      </set>
      <set>
        <setSpec>papr</setSpec>

```



Open Archives Initiative - Repository Explorer

explorer version - 1.45 : protocol version - 1.0/1.1/2.0 : June 2002

<http://cogprints.soton.ac.uk/perl/oai?verb=ListIdentifiers>

Archive details : <http://cogprints.soton.ac.uk/>

List of Record Identifiers

Click a link to view more information

- :cogprints:1 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:2 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:3 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:4 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:5 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:6 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:7 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:8 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:9 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:10 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:11 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:12 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:13 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:14 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:15 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:16 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:17 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:18 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:19 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:20 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)
- :cogprints:21 [\[display record in Dublin Core\]](#) [\[display metadata formats\]](#)

http://cogprints.soton.ac.uk/perl/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:cogprints:1

Archive details : <http://cogprints.soton.ac.uk/>

Parsed Output

List of Fields

header:

identifier : oai:cogprints:1
datestamp : 1998-03-06

metadata:

dc:

title: Neural Basis of Utility Estimation
creator: Shizgal, Peter
subject: Behavioral Neuroscience
description: The allocation of behavior among competing activities and goal objects depends on the payoffs they pro
date: 1997-01-01
type: Journal (Paginated)
identifier: <http://cogprints.ecs.soton.ac.uk/archive/00000001/>
format: pdf http://cogprints.ecs.soton.ac.uk/archive/00000001/00/cur_op_e-print_version.PDF
format: html http://cogprints.ecs.soton.ac.uk/archive/00000001/01/CUR_OP_E-print_version.html

Request : http://cogprints.ecs.soton.ac.uk/perl/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:cogprints:1

Response Date : 2002-09-12T12:32:29+01:00

Raw XML Output

```
<?xml version="1.0" encoding="UTF-8" ?>

<GetRecord xmlns="http://www.openarchives.org/OAI/1.1/OAI_GetRecord" xsi:schemaLocation="http://www.openarchives.org/
<responseDate>2002-09-12T12:32:29+01:00</responseDate>
<requestURL>http://cogprints.ecs.soton.ac.uk/perl/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:c
<record>
  <header>
    <identifier>oai:cogprints:1</identifier>
    <datestamp>1998-03-06</datestamp></header>
  <metadata>
    <dc xmlns="http://purl.org/dc/elements/1.1/" xsi:schemaLocation="http://purl.org/dc/elements/1.1/ http://www.op
    <title>Neural Basis of Utility Estimation</title>
    <creator>Shizgal, Peter</creator>
    <subject>Behavioral Neuroscience</subject>
    <description>The allocation of behavior among competing activities and goal objects depends on the payoffs th
    <date>1997-01-01</date>
    <type>Journal (Paginated)</type>
    <identifier>http://cogprints.ecs.soton.ac.uk/archive/00000001/</identifier>
    <format>pdf http://cogprints.ecs.soton.ac.uk/archive/00000001/00/cur_op_e-print_version.PDF</format>
    <format>html http://cogprints.ecs.soton.ac.uk/archive/00000001/01/CUR_OP_E-print_version.html</format></dc></
</GetRecord>
```




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

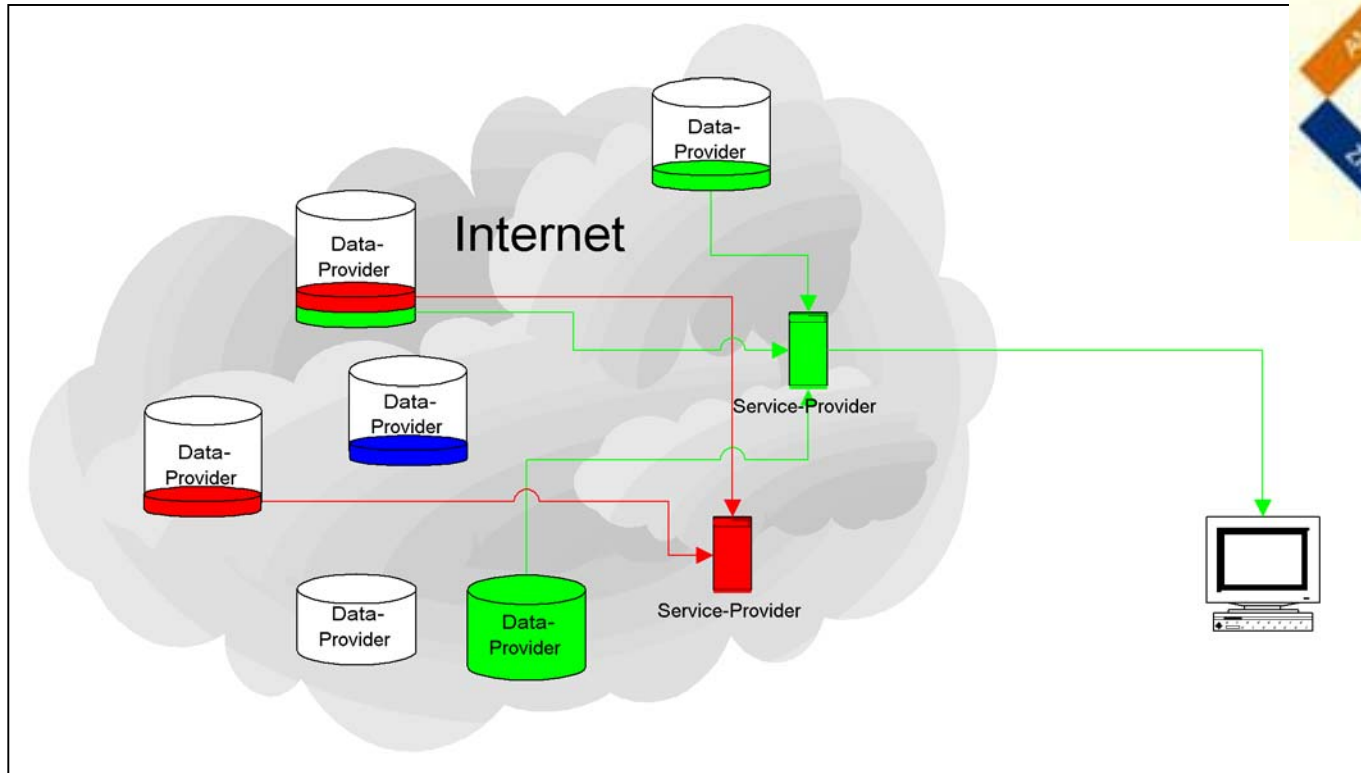
Empfehlungen der Arbeitsgruppe bei DINI

- 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





Welche Argumente sind beim OAI-Protokoll vorgesehen?

Ein Set besteht aus:

***setName*:** eine sprechende Bezeichnung für den Inhalt

***setTag*:** ein alphanumerischer String ohne Leerzeichen

***setSpec*:** durch Doppelpunkt separierte *setTags*

Das Argument *listSets*

ergibt eine Liste, mit der inhaltlichen Struktur des Archivs in Form der *setSpec*. Es wird bei der Liste allerdings der *setName* ausgegeben, so dass die Hierarchie nicht sofort erkennbar ist.



Ein Beispiel für eine Hierarchie innerhalb eines Archiv

```
<set>  
  <setSpec>bio</setSpec>  
  <setName>Biology</setName>  
</set>  
<set>  
  <setSpec>bio:bio-ani-behav</setSpec>  
  <setName>Animal Behavior</setName>  
</set>  
<set>  
  <setSpec>bio:bio-ani-cog</setSpec>  
  <setName>Animal Cognition</setName>  
</set>
```




Welche Ergebnisse liefert *setSpec*?

Eine Anfrage nach *setSpec*: bio

ergibt alle Dokumente mit der Setdefinition

***setSpec*: bio**

Und

**alle Dokumente, die sich in den tieferen Ebenen der
Sethierarchie befinden, wie in unserem Beispiel:**

***setSpec*: bio:bio-ani-behav (Animal Behavior)**

und

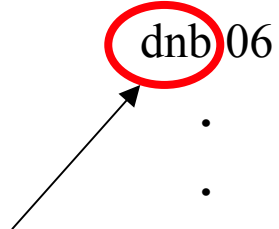
***setSpec*: bio:bio-ani-cog (Animal Cognition)**



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





Beispielelement Plain-XML Darstellung

```
<set>
```

```
<setSpec>dnb</setSpec>
```

```
<setName>DNB classified objects</setName>
```

```
</set>
```

```
<set>
```

```
<setSpec>dnb:01 </setSpec>
```

```
<setName>Knowledge and Culture in General </setName>
```

```
</set>
```

```
<set>
```

```
<setSpec>dnb:30 </setSpec>
```

```
<setName>Chemistry </setName>
```

```
</set>
```



Set Formaler 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
pub-type:masterthesis	Diploma Theses
pub-type:report	Report
pub-type:paper	Paper
pub-type:conf-proceeding	Conference Proceedings
pub-type:lecture	Lectures
pub-type:music	Music
pub-type:program	Programs
Pub-type:play	Play
Pub-type:news	News
Pub-type:standards	Standards



Beispielelement Plain-XML Darstellung

```
<set>
```

```
<setSpec>pub-type </setSpec>
```

```
<setName>Documents having a formal publication-type</setName>
```

```
</set>
```

```
<set>
```

```
<setSpec>pub-type: monograph/span></setSpec>
```

```
<setName>Books, Monographs </setName>
```

```
</set>
```



Set Formale Dokumenttyp

SetSpec	SetName
doc-type	formal document-type
doc-type:text	Text
doc-type:notes	Notes
doc-type:image	Image
doc-type:audio	Audio
doc-type:video	video
doc-type:multimedia	multimedia
doc-type:data	data
doc-type-binary	Binary data, (executable) program



Beispielelement Plain-XML Darstellung

```
<set>
```

```
<setSpec>doc-type </setSpec>
```

```
<setName>formal document-type</setName>
```

```
</set>
```

```
<set>
```

```
<setSpec>doc-type: video/span></setSpec>
```

```
<setName>Video </setName>
```

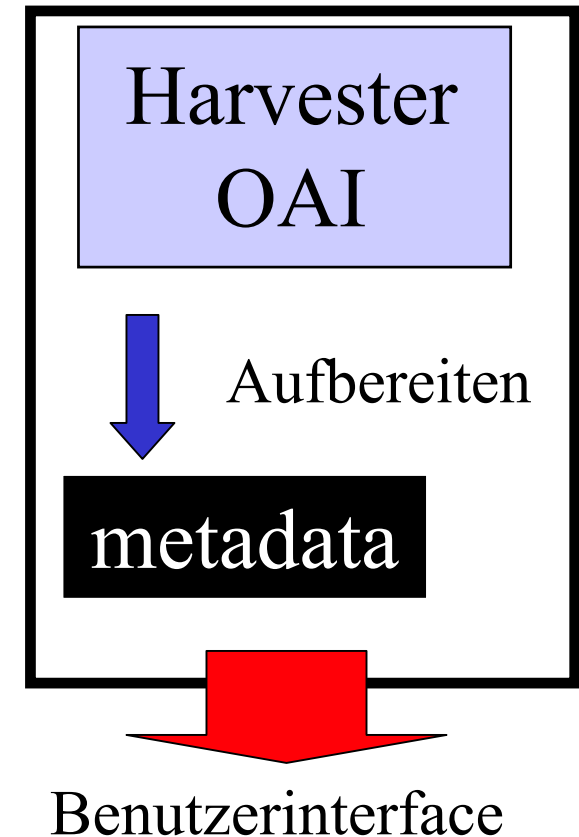
```
</set>
```



Serviceprovider

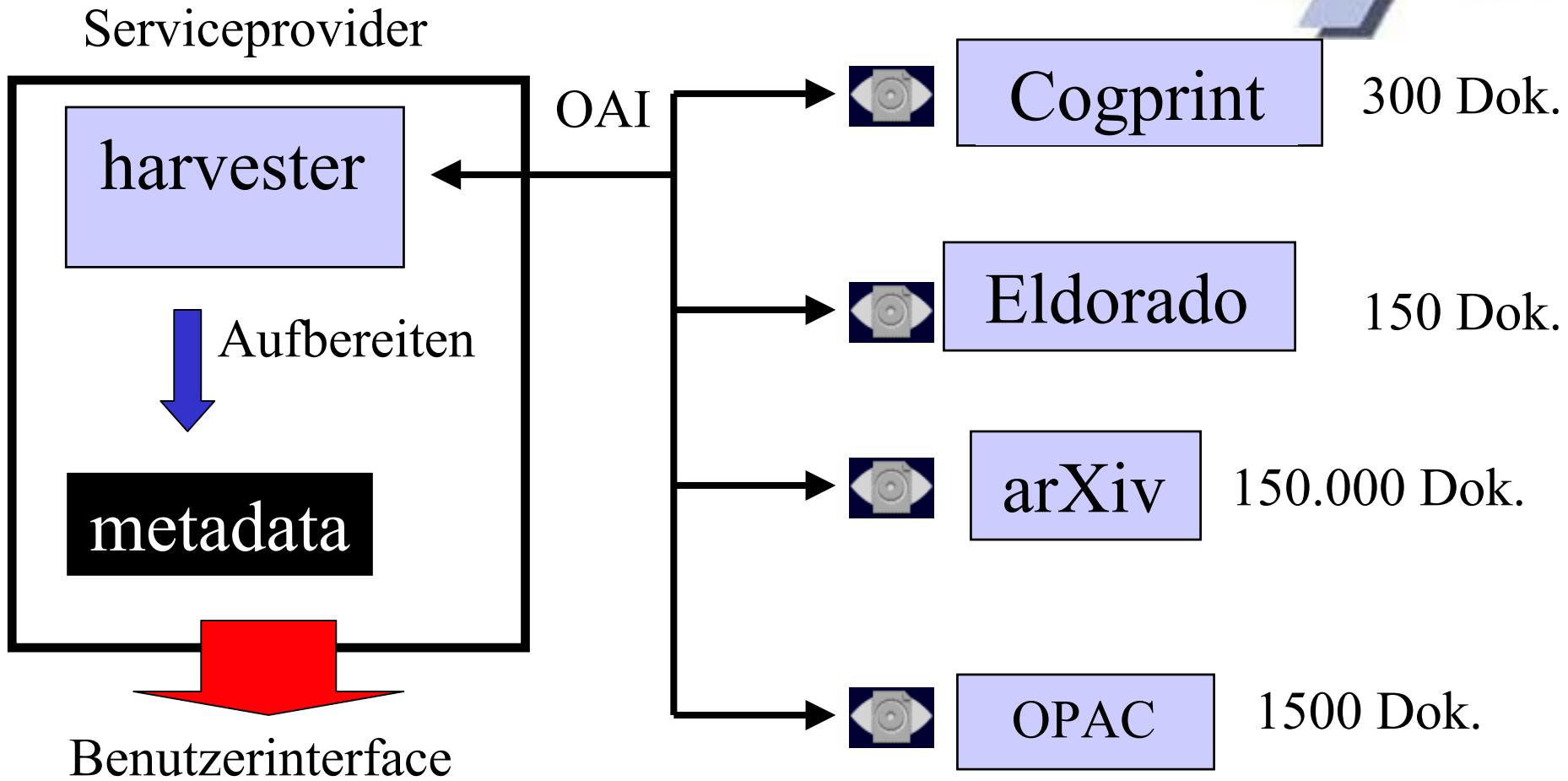
- Sammeln der Daten - OAI
- Analysieren der eingesammelten Daten
- Aufbereitung der gefundenen Dokumente
 - Zitatanalyse
 - automatisches Klassifizieren
 - Browsingfunktion
 -

Serviceprovider





Was macht ein 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>**



Einige Anmerkungen zu 2.0

Die Version 2.0 ist nicht abwärtskompatibel zu der Version 1.1.

Alle bisher registrierten Archive müssen sich erneut registrieren.

Die Version 1.1 hatte immer einen experimentellen Charakter.

Die Version 2.0 soll eine dauerhafte Basis für die Implementierung des Protokolls bieten.



Änderungen in der Version 2.0

- Standardisierung von Daten und Zeiten (ISO 8601) galt bisher auch schon. YYYY-MM-DD
- ISO 8601 für die from / until Angaben und Anfragen
- Erweiterte Flusskontrolle - erleichtert die Steuerung von Zugriffen und schafft einen Lastausgleich zu Spitzenzeiten.
- Fehlerbehandlung soll unabhängig vom http-Protokoll erfolgen.



Änderungen bei Version 2.0

- Zusammenarbeit mit der DCMI (Dublin Core Metadata Initiative)
- Provenance-Container, bei aggregierenden DataProvider kann das Ursprungsarchiv angegeben werden. (Lösung für Verbundsysteme?)



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 !



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)