



# open archives *forum*

## **Offene Archive, die Open Archives Initiative und Europäische Aktivitäten**

Uwe Müller

Humboldt-Universität zu Berlin

Arbeitsgruppe *Elektronisches Publizieren*

Universitätsbibliothek / Rechenzentrum

[u.mueller@rz.hu-berlin.de](mailto:u.mueller@rz.hu-berlin.de)



1. Die Problematik der offenen Archive
2. Die Open Archives Initiative
3. Europäische Aktivitäten: OAFforum
4. Fragebogen und Workshop im März

- entstehen aufgrund der aktuellen Publikationssituation in den Wissenschaften
- Ansätze des *Self Archiving*
- Probleme:
  - unterschiedliche Metadatenstandards
  - unterschiedliche Übertragungsprotokolle / Schnittstellen
  - unterschiedliche Zugriffsstrategien
- Mangelnde Interoperabilität!
- Neue Dienste sind sehr schwer zu etablieren.

- unterschiedliche Interpretationsmöglichkeiten für einzelne Metadatenfelder
  - z.B.: Autor-Feld in Bibliothekskatalog  
(bei MAB-Format: Feld 100)
  - vs. Künstler-Feld in Museums-Datenbank
  - vs. Fotograf in Bilddatenbank
- Übertragungsprotokolle
  - Bibliotheken: Z39.50 – Schnittstelle
  - NCSTRL: DIENST-Protokoll
  - Harvest-System: auf HTTP-Basis mit SOIF-Metadatenformat

## Problem: Suche in Archiven

IST- 2001-320015

- datenbankbasiert
  - Echtzeit
  - ODBC, SQL-Schnittstellen
  - abfragbar mit Programmiersprachen PHP, Perl, Java
- z.B. Harvest
  - Indexfile im Filesystem
  - Generierung spezieller Suchindizes
  - funktioniert nicht in Echtzeit
  - Suchsprache SOIF

1. Die Problematik der offenen Archive
2. Die Open Archives Initiative
3. Europäische Aktivitäten: OAFforum
4. Fragebogen und Workshop im März

# Die Open Archives Initiative

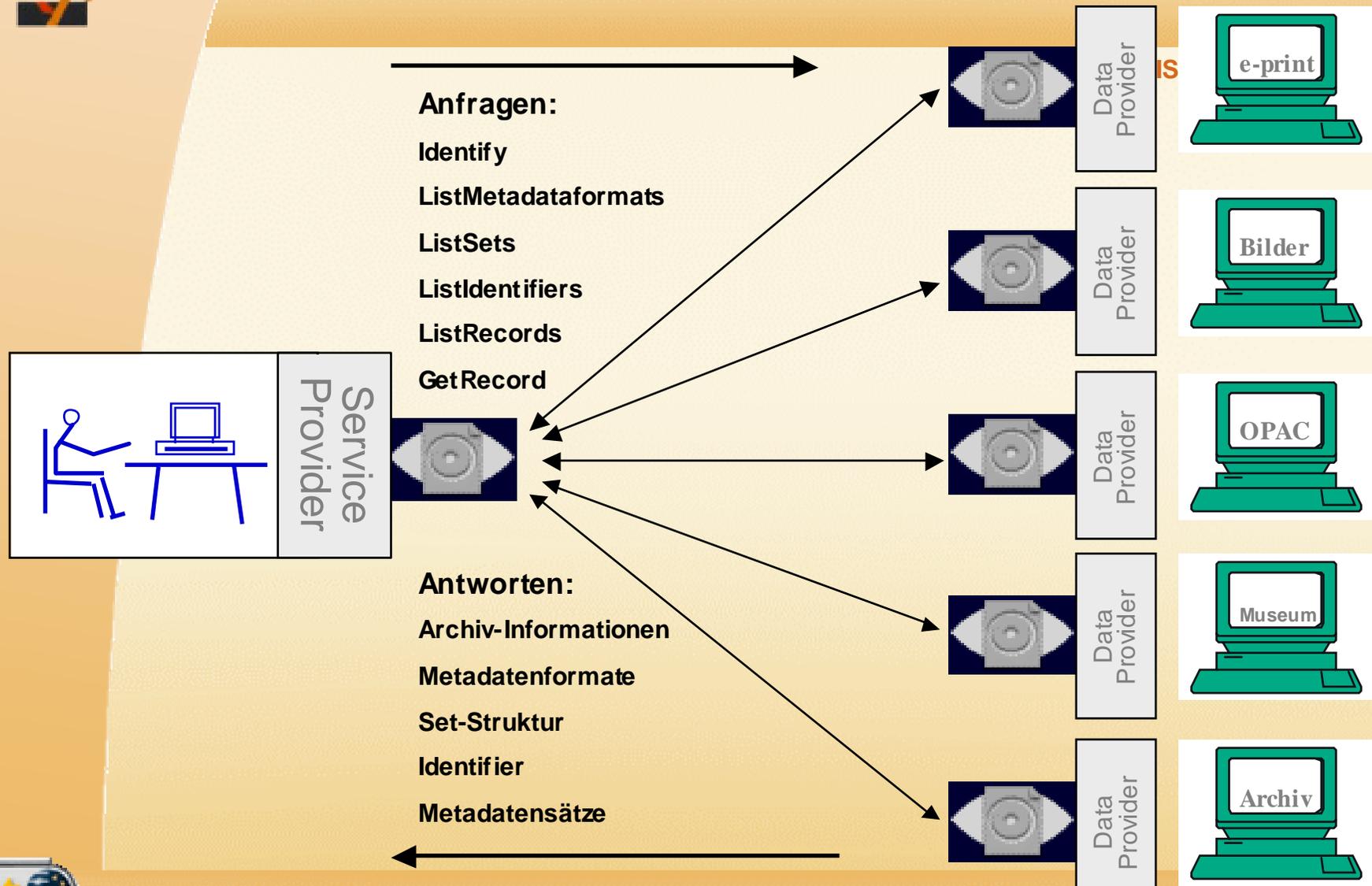
IST- 2001-320015

- <http://www.openarchives.org/>
- **Prinzipien / Strategie**
  - weltweite Zusammenführung wissenschaftlicher Archive
  - freier Zugriff auf diese Archive (mindestens: Metadaten)
  - einheitliche Schnittstellen für Archive und Serviceanbieter
  - einfache Implementierbarkeit
- **Grundregeln von OAI**
  - Definition einer Policy für die Metadaten / Dokumente
  - Nutzung des *Dublin Core* – Metadatensatzes
  - OAI-spezifische XML-Syntax
  - netzbasierendes Serversystem
  - spezielles *Open Archive Protokoll* auf HTTP-Basis

- **Steering Committee**  
12 Mitglieder, USA, UK, D  
Allgemeine Politik der OAI
- **Executive Committee**  
Carl Lagoze (Cornell University)  
Herbert van de Sompel (Los Alamos)
- **Technical Committee**  
Erarbeitung der OAI-Spezifikation (Austauschprotokoll)  
Kontakt mit OAI-Community

- *Data Provider (Archive)*
  - freie Zugänglichkeit der Metadaten
  - nicht notwendigerweise: Freier Zugriff auf Volltexte
  - einfache Implementierbarkeit: niedrige Schwelle
  - insbesondere für kleinere Institutionen
- *Service Provider*
  - nutzen OAI-Schnittstellen der *Data Provider*, um Services zu etablieren:
    - z.B.: Suche, Navigation, Alerting-Dienste u.a.
    - Informationsdienstleistungen

# OAI: Technisches Modell



## OAI: Protokoll (PMH)

IST- 2001-320015

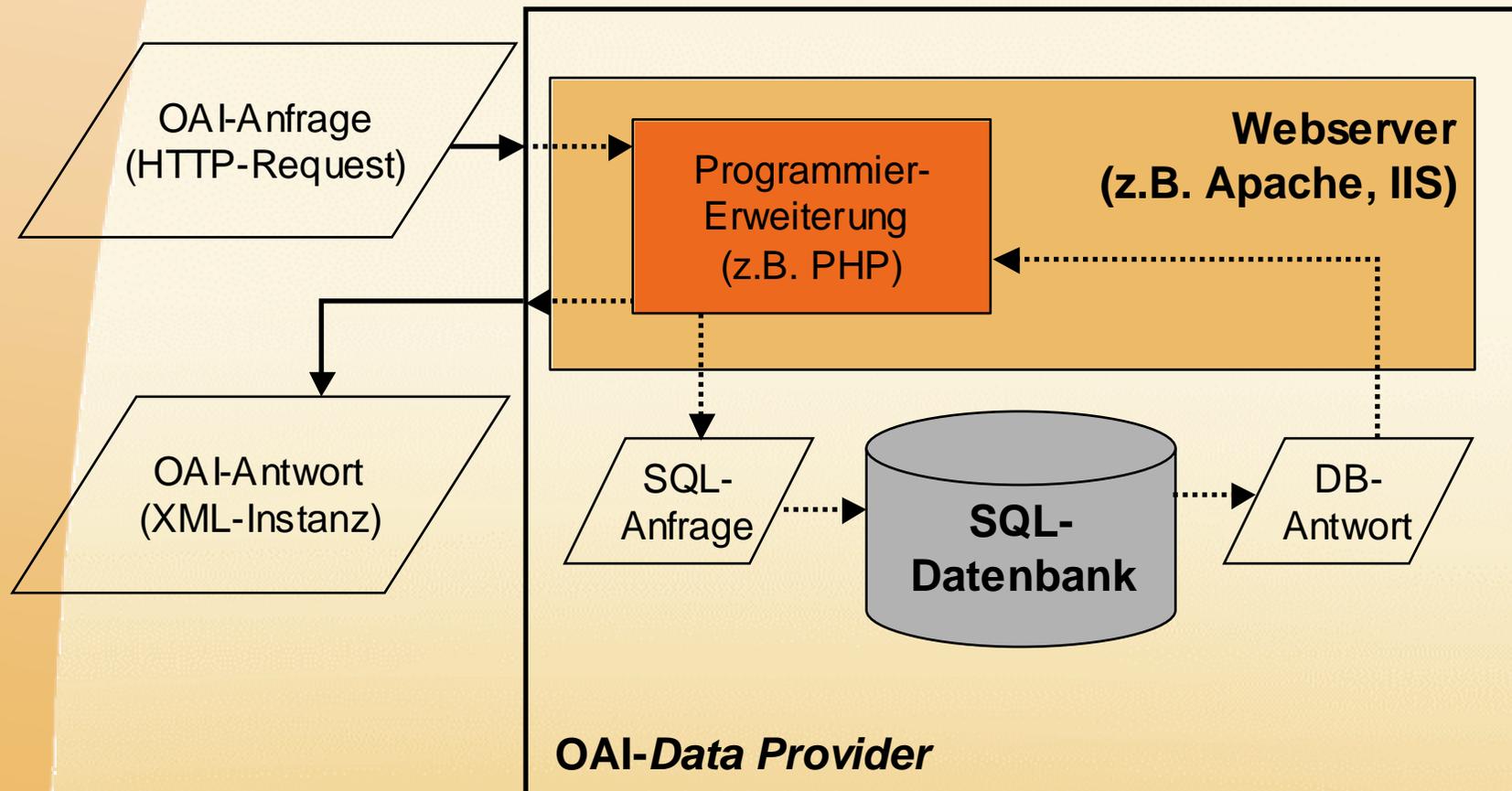
- basiert auf HTTP
  - Anfrageargumente als GET- oder POST-Parameter
  - Antworten werden in XML kodiert
- sechs Anfragetypen
- Set-Hierarchie (optional)
- beschränkte Eingrenzungsmöglichkeiten für Anfragen (keine Suche!)
  - Änderungsdatum des Metadatensatzes
  - Metadatenformat
  - Sets
- Fehlermeldungen
- Flusskontrolle

- in XML kodiert
- Header
  1. Eindeutiger Identifier: Schlüssel für Archivanfragen  
(z.B.: oai:HUBerlin.de:1000192)
  2. Datestamp: Letzten Änderung eines Metadatensatzes  
(z.B.: 2002-01-26)
  3. Sets, in denen der Metadatensatz enthalten ist
- Metadaten
  - Metadatenprefix
  - Dublin Core (mindestens)
  - beliebige andere Metadatensätze

## Beispiel: [http://edoc.hu-berlin.de/OAI-2.0? verb=ListIdentifiers&from=2002-01-03&until=2002-01-08& metadataPrefix=oai\\_dc&set=doctypes:dissertations](http://edoc.hu-berlin.de/OAI-2.0?verb=ListIdentifiers&from=2002-01-03&until=2002-01-08&metadataPrefix=oai_dc&set=doctypes:dissertations)

IST- 2001-320015

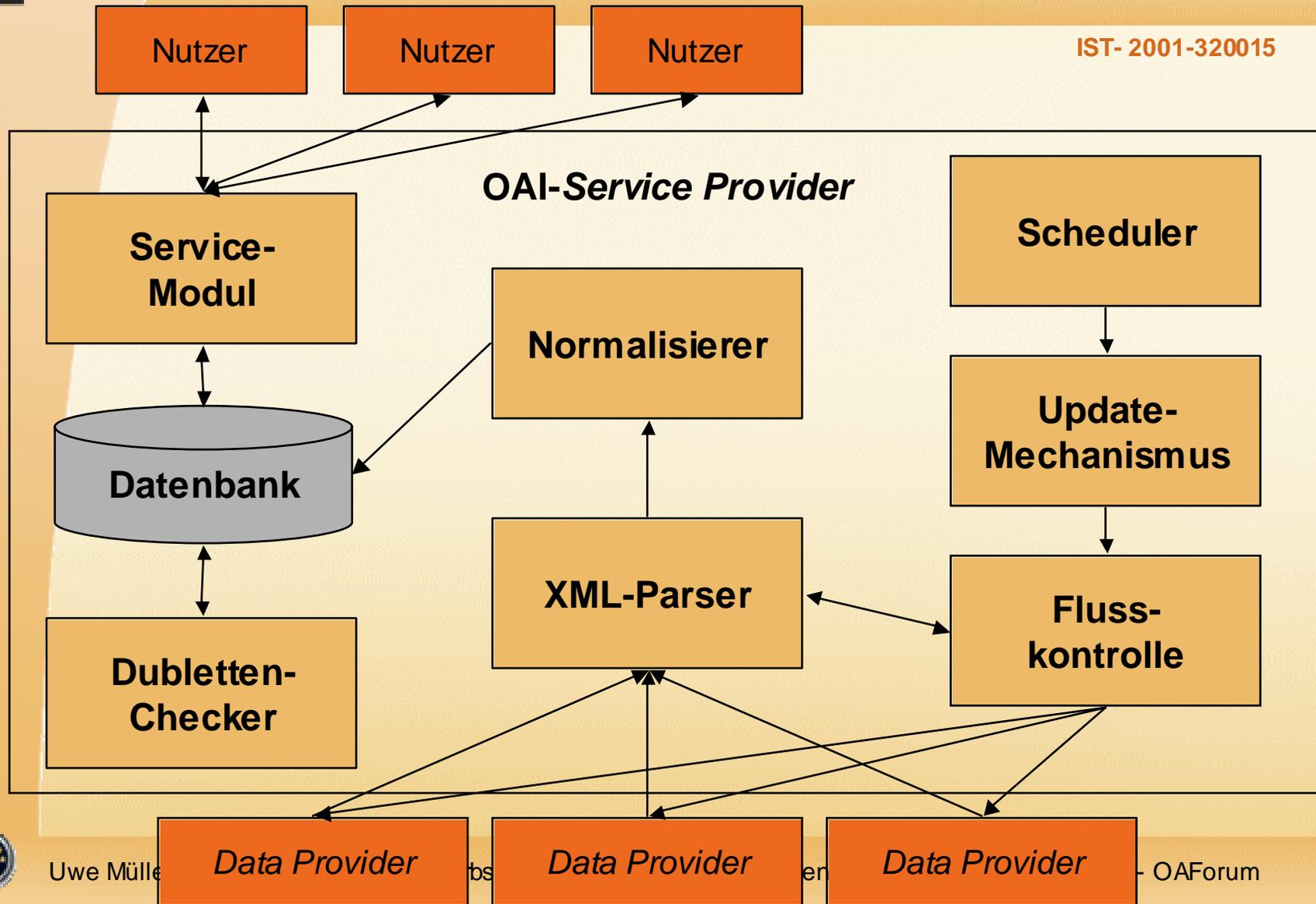
```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-10-22T17:49:49+01:00</responseDate>
  <request verb="ListIdentifiers" from="2002-01-03" until="2002-01-08" metadataPrefix="oai_dc"
    set="doctypes:dissertations">http://edoc.hu-berlin.de/OAI-2.0</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:HUBerlin.de:3000819</identifier>
      <timestamp>2002-01-08</timestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb33</setSpec>
    </header>
    <header>
      <identifier>oai:HUBerlin.de:3000831</identifier>
      <timestamp>2002-01-07</timestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb27</setSpec>
    </header>
  </ListIdentifiers>
</OAI-PMH>
```



- Definitionen
  - eindeutiger Archiv-Bezeichner (z.B. HUBerlin.de, physdoc)  
(siehe <http://oaisrv.nsd.comell.edu/Register/BrowseSites.pl>)
  - eindeutige Record-Bezeichner (z.B. oai:HUBerlin.de:254)
  - Metadatensätze (mindestens: *Dublin Core*)
  - Set-Hierarchie (optional)
- Webserver-Erweiterung (CGI, PHP, JavaServlets)
  - Parser für HTTP-Anfragen
  - Datenbank-Komponente
  - XML-Generator
- Offizielle Registrierung
  - Konformitätstest (Korrektheit, Robustheit, ...)

# OAI: Architektur Service Provider

IST- 2001-320015



## OAI: Implementation *Service Provider*

IST- 2001-320015

- Auswahl des zur Verfügung zu stellenden Service
- Auswahl der integrierten *Data Provider*
- Absprachen mit den *Data Providern*
  - Auswahl des Metadatensatzes (mindestens: DC)
  - Auswahl der Set-Hierarchie (Empfehlungen: DINI)
- technische Aspekte
  - Auswahl der Datenbank, Programmierschnittstelle
  - weitere technische Details
- Implementation (siehe Architektur)
- Offizielle Registrierung

## ➤ Data Provider / Repositories

Beispiele:

siehe <http://www.openarchives.org/Register/BrowseSites.pl>

## ➤ Service Provider

Beispiele:

Cross Archive Searching Service: <http://arc.cs.odu.edu/>

MyOAI: <http://www.myoai.org/>

DINI: <http://edoc.hu-berlin.de/oaisearch/>

Physnet: <http://physnet.uni-oldenburg.de/oai/query.php>

BSZ – in Planung

ProPrint

1. Die Problematik der offenen Archive
2. Die Open Archives Initiative
3. Europäische Aktivitäten: OAFforum
4. Fragebogen und Workshop im März



## Europäische Aktivitäten: OAForum

IST- 2001-320015

- Das Projekt *Open Archives Forum*
  - European Union Information Society Technologies (IST) Programme
  - Begleitende Maßnahme
  - Projektbeginn 1. Oktober 2001 (Laufzeit: 2 Jahre)
  - Partner:
    - UKOLN, University of Bath
    - IEI-CNR, Pisa
    - Humboldt-Universität zu Berlin, Rechenzentrum
  
- <http://www.oaforum.org/>





## Das Projekt *Open Archives Forum*

IST- 2001-320015

- ... ist kein weiteres OAI Implementationsprojekt
- ... ist ein Community-Projekt
  - Neue Interessengruppen auf die Problematik aufmerksam machen
  - Unterstützen des Aufbaus einer globalen digitalen verteilten Bibliothek
  - Förderung und Unterstützung des Informations- und Technologieaustausch zwischen den Anwendern
  - Insbesondere Förderung nationaler Initiativen
- OAFforum ist ein Verbreitungsprojekt



## Betreiber und Nutzer offener Archive

### ➤ **Communities**

- Institutionen des kulturellen Erbes
- Museen
- Europäische Digitalisierungsprojekte
- Forschungseinrichtungen
- Sektor der öffentlichen Bibliotheken
- Spezielle Nutzergruppen
- Verlage
- Kommerzielle Sektor
- Bildungssektor

### ➤ **Service Provider**

- E-print Archive
- Fachportale mit aggregierender Funktion
- Mehrwertdienste

### ➤ **Data Provider**

- Bestehende Metadaten-repositories
- Neue Metadaten-sammlungen

- Erreichen von Nutzergruppen
- Erfassen, was in Europa auf dem Gebiet passiert
- Lenken der Aufmerksamkeit auf diese Problematik und Anregen zur Diskussion über die Hauptprobleme
  - Schaffen einer gemeinsamen Terminologie für digitale Sammlungen
  - Metadaten bzw. Volltextrecherchemodelle
  - Erfassen der nutzerspezifischen Anforderungen
  - Anregung zum Aufbau erweiterter Dienstleistungen
- Unterstützen Europäischer Projekte
  - Entwickeln möglicher Beispiellösungen
  - Etablieren neuer Geschäftsmodelle für die Behandlung digitaler Ressourcen

- Workshops, Verbreitung
  - Informationen, Technologie verbreiten
  - Interesse bei neuen Anwendergruppen wecken
- Organisatorische Evaluation
  - Geschäftsmodelle untersuchen
  - Thematik der Urheberrechte erfassen
- Technische Evaluation
  - Anwendung des technischen Rahmens der OAI
  - Entwicklung eines Informationsportals: enthält Informationen über Projekte, Software, Metadaten Schemata, Implementationen
  - Probleme der Interoperabilität

- Mai 2002, Pisa, Italien
  - Erfahrungen der europäischen e-prints community
  - Etablierung des Forums
- Dezember 2002, Lissabon, Portugal
  - Archive und Bibliotheken
- März 2003, Berlin, Deutschland
  - Museen und Einrichtungen des kulturellen Erbes
- September 2003, Bath, Großbritannien
  - In Kooperation mit LIBER, SPARC Initiative
  - Größere Tagung

- Geschäftsmodelle
  - Kooperation von *Data Providern* zum Aufbau von Dienstleistungsnetzwerken
  - Austausch von Metadaten zwischen Archiven und Diensten
  - Zurverfügungstellung von Mehrwertdiensten (Metadatenanreicherung, Automatische Klassifikation, OpenURL)
- Probleme der Urheberrechte
  - IPR und copyright (Einfluss auf Produzenten und Vertreiber)
  - Eigentumsrechte an Metadaten (Gemeinsame Nutzung von Metadaten, Metadaten austausch, Vereinbarungen mit Verlagen etc.)
  - Langzeitverfügbarkeit digitaler Ressourcen (Digitale Ressourcen, Wissenschaftskommunikation, Metadaten)
- Diskussionsgruppe
  - Dennis Nocholson (University Glasgow)

- Kompatibilität
  - Integration mit bestehenden Technologien
  - Normalisierung der Metadaten
  - Ist *unqualified Dublin Core* ausreichend?
- Probleme des Datenbank-Management
  - Update
  - Skalierbarkeit
  - Dubletten
- Software, Tools
- notwendiger Aufwand
  - Manpower
  - Fähigkeiten
  - benötigte Zeit, ein Repository / Service aufzubauen

1. Die Problematik der offenen Archive
2. Die Open Archives Initiative
3. Europäische Aktivitäten: OAForum
4. Fragebogen und Workshop im März



# Fragebogen zur technischen Evaluation

IST- 2001-320015

- Bitte füllen Sie den Fragebogen aus!
- <http://www.oaforum.org/resources/tecvalq2.php>



Open Archives Forum - Resources - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.oaforum.org/resources/tecvalq2.php

Open Archives Forum - Resources



**Resources:**  
[Information Resource](#) | [Glossary](#) | [Talks and Presentations](#) | [Articles](#) |  
[Other Resources](#) | [Technical Validation](#)

[Technical Validation Questionnaire](#) | [1st Questionnaire](#)

## Technical Validation Questionnaire

An important role of the Open Archive Forum is sharing information about technical issues relating to open archives, and especially about implementations of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Please help us by filling at least part of our questionnaire. Don't worry if you can't fill in all of the form, we would be pleased if you fill in as much of the form as you can.

**Your repository is already OAI compatible as Data Provider and/or Service Provider (test implementations included)?**  
[Click here to get the questions! >>>](#)

**Your repository does not now have any OAI implementations, but the possibility is being considered?**  
[Click here to get the questions! >>>](#)

[Home](#)  
[News](#)  
[Overview](#)  
[Resources](#)  
[Workshops](#)  
[Project Documents](#)  
[Contact us](#)

[Imprint](#)

The Open Archives Forum (OAF) is an IST-Accompanying Measures project (IST-2001-320015).  
The partners of OAF are: [University of Bath-UKOLN](#) (United Kingdom), [Istituto di Scienza e Tecnologie della](#)



<http://www.oaforum.org/resources/tecvalq2.php>



## Informationsportal

IST- 2001-320015

- Bitte helfen Sie uns mit Informationen zu Ihren Projekten, tragen Sie sich in unserer Datenbank ein!
- [http://www.oaforum.org/oaf\\_db/oaf\\_index.php](http://www.oaforum.org/oaf_db/oaf_index.php)



Open Archives Forum - Resources - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.oaforum.org/oaf\_db/register/ Search

Open Archives Forum - Resources



**Resources:**  
[Information Resource](#) | [Glossary](#) | [Talks and Presentations](#) | [Articles](#) |  
[Other Resources](#) | [Technical Validation](#)

## Information Resource Database

[Search](#) || [List: Services](#) | [Repositories](#) | [Projects](#) | [Software](#) | [Organisations](#) || [»» Register](#)

### Registration with the OAForum Information Resource Database

**If you are participating in this registration process for the first time please read the notes below before starting !!!**

<input type="button" value="Register Project"/>	If you have a project that deals with open archives, even if you have not programmed a repository or service yet, please register, so other people know about your ideas.
<input type="button" value="Register Repository"/>	If you have digital collections of metadata (OPACs), digitised or born digital materials, you're welcome to register your repository here.
<input type="button" value="Register Service"/>	If you developed a web based service on top of web repositories, we would like to know, what exactly is this service and which repositories did you use.
<input type="button" value="Register Software"/>	Please register if you developed software while setting up a repository or service, and if you would like to announce that to others, so it can be reused.

[http://www.oaforum.org/oaf\\_db/register/](http://www.oaforum.org/oaf_db/register/)

1. Select one of the entry buttons.
2. A sequence of forms will lead you through the respective registration procedure.

\*\*\* On each form page (except the first one after clicking the button) you have three filling options. In each case it is only possible to use one of those:

- a. Selection from existing data --- By this selection you can join already available data with your new data and

Document: Done (0.375 secs)

Open Archives Forum - Resources - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.oaforum.org/oaforum\_db/list\_db/ Search

Open Archives Forum - Resources



**Resources:**  
[Information Resource](#) | [Glossary](#) | [Talks and Presentations](#) | [Articles](#) |  
[Other Resources](#) | [Technical Validation](#)

## Information Resource Database

[Search](#) || [List: Services](#) | [Repositories](#) | [Projects](#) | [Software](#) | [Organisations](#) || [»» Register](#)

### List of Services

**Service:** [DINI Search Service](#)

**Description:** DINI- Search Service for OAI Data Provider of German University Libraries, which mostly contain digital dissertations, videos of lectures, master theses, public readings, university publications. Test implementation for OAI.  
<http://www.dini.de/oaisuche/>

**Service:** [tool2know](#)

**Description:** In the world wide web, resources like e-magazines, data collections, e-books and PDF documents are difficult to trace and TOOL2KNOW offers a way to track them through a storyboard based on a dartboard approach. In this version of TOOL2KNOW, you can find two hyperguides: the guide to the Italian Museums websites and the one to the Italian scientific research websites. A third hyperguide about Italian newspapers websites is under construction. For these guides are about Italian websites, you will find the English language for the interface framework of the application, and the Italian language for the names of the internal values. Next hyperguides can keep the English framework and provide local languages for the websites values. TOOL2KNOW has been realized for Hyperguide, one of the Trials projects selected under the 4th IST Call for Pro-posals by the European Commission. Visit our project site to get some info about the project.  
<http://www.tool2know.com>

Home  
News  
Overview  
Resources  
Workshops  
Project Documents  
Contact us

Document: Done (0.359 secs)

## 3. OAForum-Workshop

IST- 2001-320015

- 27. Bis 29. März in Berlin
- inhaltliche Schwerpunkte
  - Fragen und Probleme von Museen, die offene Archive betreiben oder dies anstreben
  - Metadaten-Initiativen von Projekten und Institutionen
- endgültiges Programm: Januar 2003
- Anmeldeschluss: Februar 2003
- [http://www.oaforum.org/workshops/berl\\_invitation.php](http://www.oaforum.org/workshops/berl_invitation.php)

# Herzlichen Dank!

IST- 2001-320015

➤ <http://www.oaforum.org/>

- Leona Carpenter (Bath)
- Donatella Castelli (Pisa)
- Susanne Dobratz (Berlin)

Uwe Müller  
Humboldt-Universität zu Berlin  
Arbeitsgruppe *Elektronisches Publizieren*  
Universitätsbibliothek / Rechenzentrum  
u.mueller@rz.hu-berlin.de

# Weitere Informationen

IST- 2001-320015

- Open Archives Initiative  
[www.openarchives.org/service/listproviders.html](http://www.openarchives.org/service/listproviders.html)  
[www.openarchives.org/service/listproviders.html](http://www.openarchives.org/service/listproviders.html)
- Open Archives Forum  
<http://www.oaforum.org/>
- Signal-Hill OAI Site  
<http://www.signal-hill.org/nav/archives2.html>
- Eprints.org  
<http://software.eprints.org/#sites>