Open Archives Forum
- Technical Validation -

Birgit Matthaei
Humboldt University Berlin, Germany
Computer and Media Service, Electronic Publishing Group
birgit.matthaei@cms.hu-berlin.de
Creating Information Sources

- European portal for open archives information
  - Information Resource Database (registry of repositories, services, software, projects and associated organisations)

- Evaluation of status, experiences and future plans regarding European OAI implementations
  - Online “Technical Validation Questionnaire”

- Systematic inventories
  - Repositories, services and tools

- Reports on own experiences
  - OAI-PMH 2.0 alpha and beta tester,
  - Implementation of OAI Services

→ Highlight some aspects of European activities on OAI in relation to worldwide activities
Overview Repositories - World

Overview on OAI activity (continents)

- America (North) 93
- Europe 82
- Australia 6
- America (Middle & South) 5
- Asia 4
- Africa 1

Upgrading to OAI 2.0 (Nov. 2002)

- OAI 1.1 60%
- OAI 2.0 40%

Upgrading to OAI 2.0 (Aug. 2003)

- OAI 1.1 17%
- OAI 2.0 83%
Overview Repositories - Europe

Overview of European countries engaged in OAI implementation

- UK (22)
- Germany (18)
- France (11)
- Italy (7)
- Sweden (3)
- Austria (2)
- Netherlands (1)
- Belgium
- Belorussia
- Denmark
- Finland
- Ireland
- Norway
- Portugal
- Russia
- Slovenia
- Spain
- Switzerland
Details of Questionnaire

Data Provider
- active: 19
- in development: 10
- planned: 22

Service Provider
- active: 5
- in development: 11
- planned: 16
Questionnaire: Community Types

- Library: 31%
- Archive: 19%
- Preprint/Science: 13%
- Publisher: 2%
- Museum: 8%
- Others: 14%
- No specification: 13%

Multiple answers possible
Questionnaire: Object Types

Multiple answers possible

- Metadata
- Fulltext documents
- Abstracts
- Images - digitised mat.
- Images - Vector graphics
- Video/Streams
- Software
- Audio
- Raw/Statistic Data
- Others

already OAI compatible
not yet OAI compatible
Questionnaire: Used Software

- Many different tools were mentioned:
  - ADLIB, ARNO, CDSware, DIENST, DSpace, Elektra, EPrints, PERL implementations, OAI Cat, OAI Harvester, VT-ETB-db

- Today: 50 % use self developeded tools
  One year ago: 80 % used self-developed tools

- Trend: Need of tools which are user-friendly complete solutions
  - cover typical functionalities
  - to be installed by relatively small expenditure
  - adaptable to special requirements if necessary
  - little expenditure with the further care of the data
Tools: Eprints - DSpace

- **GNU Eprints** - developed at the Electronics and Computer Science Department of the University of Southampton, UK

- **DSpace** - newly developed as a joint project of the MIT Libraries and the HP Company, USA

- Some numbers of the inventory of repositories:

<table>
<thead>
<tr>
<th></th>
<th>Nov 2002</th>
<th>Aug 2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>eprints</td>
<td>26%</td>
<td>39%</td>
</tr>
<tr>
<td>other</td>
<td>74%</td>
<td>61%</td>
</tr>
<tr>
<td>tools</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tools: Eprints - DSpace

- Open source developments for archiving
- Nearly identically in their functionality
  - search functions, document archiving, online interfaces for self archiving, integration of the OAI PMH, …
- Systems base on different technologies
  - Eprints: traditional technologies, runs on pure open source systems: mySQL and Apache, programmed by using the script language “Perl”
  - Dspace: operates with new technologies such as the Postgres database and Tomcat for jsp/java web application, higher performance
**Questionnaire: Metadata Formats**

Multiple answers possible

- **Dublin Core simple**: 17 already OAI compatible, 7 not yet OAI compatible
- **Dublin Core qualified**: 10 already OAI compatible, 5 not yet OAI compatible
- **MARC 21**: 4 already OAI compatible, 2 not yet OAI compatible
- **UNIMARC**: 3 already OAI compatible, 2 not yet OAI compatible
- **EAD**: 1 already OAI compatible, 3 not yet OAI compatible
- **MAB**: 1 already OAI compatible, 2 not yet OAI compatible
- **TEI**: 1 already OAI compatible, 2 not yet OAI compatible
- **METS**: 3 already OAI compatible, 6 not yet OAI compatible
- **Others (single mentioned)**: 12 already OAI compatible, 6 not yet OAI compatible
Problems for Service Provider

- **Problem: Standardisation**
  “heterogeneity of the content of the metadata records requires the service provider to expend a lot of effort in normalizing the data in order to make it more comparable and usable”
  - could be done at lesser cost by the individual DP
  - or by the development of middleware tools that service providers could use for data normalisation

- **Lacking interoperability**
  - different metadata standards, terminology, languages, access strategies, interfaces / transfer protocols, copyright regulations

- **Difficulty to establish joint services based on open archives**
Creating Recommendations

Example of a possible solution:

**German Initiative for Networked Information:**

**Recommendations for usage of OAI-PMH**

- created by DINI-OAI working group ([http://www.dini.de/](http://www.dini.de/))
- target: agreement on syntax and semantics of OAI set definitions for German data and service providers
- enhance retrieval quality and support subject gateways (e.g. Physnet, Dissertation search engine, ...)
- definition of three classification types
  - subjects (according to DNB)
  - formal publication types (e.g. dissertation)
  - formal document types (e.g. text, audio)
- example service provider based on recommended sets: [http://edoc.hu-berlin.de/e_suche/oai.php](http://edoc.hu-berlin.de/e_suche/oai.php)
OAI advantages?

- **Importance of OAI**
  - provide additional services to existing services
  - replace existing services through OAI interface
  - better retrieval, make Metadata exchange available

- **Advantages of OAI**
  - share scientific knowledge, harvest other knowledge databases, cross-search in institutional assets
  - major dissemination of researchers' results
  - simple and cheap in implementation

→ „provide access to all of human knowledge“
→ „nothing other than political expediency“
http://www.oaforum.org/oaf_db/

Information Resource Database

In order to bring interested parties together, to build a community of interest, and to enable exchange of information OAForum established this web-based European information source for open archives.

This application allows you to

... browse current lists of
- Services
- Repositories
- Projects
- Software
- Organisations
- Protocols
- Metadata schemes

... search in the database
Make sure you have only used blank (" ") to separate different search terms.

Search for: [input field] [Go]

... be part of the OAForum open archives information source?

>>> Register here

All European services, repositories, projects and organisations that use an open archives approach, have started metadata initiatives or have developed appropriate software can register themselves using our web
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! »»
Project Documents:

Reports | Internal BSCW-Server (project partners only)

Reports

The Open Archives Forum project will produce several reports during 2002 and 2003. These reports will be made available here as they are delivered.

- **Project presentation**
  This briefly sets out the objectives, describes the work and expected results of the OA-Forum project, and gives some administrative details.

- **Workshop reports**
  - **Workshop report 1** (pdf-file, 197 KB)
    This describes the 1st OA-Forum workshop, held in Pisa, 13-14 May 2002: "Creating a European Forum on Open Archives Activities"
  - **Workshop report 2** (pdf-file, 285 KB)
    This describes the 2nd OA-Forum workshop, held in Lisbon, 6-7 December 2002: "Open Access to Hidden Resources"
  - **Workshop report 3** (pdf-file, 278 KB)
    This describes the 3rd OA-Forum workshop, held in Berlin, 27-29 March 2003: "Networking Multimedia Resources"
  - Workshop report 4
Thank you!

Birgit Matthaei
Humboldt University Berlin, Germany
Computer and Media Service, Electronic Publishing Group
birgit.matthaei@cms.hu-berlin.de