



open archives *forum*

Open Archives Activities and Experiences in Europe: an Overview by the Open Archives Forum

Susanne Dobratz

Humboldt-University, Berlin, Germany

Electronic Publishing Group

University Library / Computing Centre

dobratz@rz.hu-berlin.de



Susanne Dobratz, 17. Oct. 2002, Geneva, 2nd Workshop on OAI: Gaining Independence with e-prints and OAI

Partners

IST- 2001-320015

- European Union Information Society Technologies (IST) programm
- Accompanying measure
- Project start October, 1st, 2001 (2 years)
- Partners:
 - UKOLN, University of Bath
 - IEI-CNR, Pisa
 - Computing Centre of Humboldt-University, Berlin
- <http://www.oaforum.org>

... Ask and answer questions

IST- 2001-320015

- ... yet another OAI implementation project?
- It is a clustering activity
 - Communities, new communities
 - IST projects
 - National initiatives
- It is a dissimulation activity
 - Share experiences on OA in general
 - Investigation of usage: different paradigms
 - Wide availability: globally
 - Share developments: e.g. enable developing countries to get access to scientific and cultural heritage information

Participants...implementors of Open Archives

IST- 2001-320015

➤ Communities, New Communities

- Cultural heritage institutions
- Museums
- European Digitization Projects
- Research organisations
- Public library sector
- Community organisations
- Publishers
- Commercial sector
- Educational sector
- *As service providers*
- E-print archives
- Aggregators
- Added value services
- *As data providers*
- Established metadata repositories
- New data providers

Benefits of a “Forum”

IST- 2001-320015

- Is a tool to reach communities
- What is happening in Europe
- raise awareness on and discuss main issues
 - Common terminology on digital repositories
 - Metadata / fulltext harvesting models
 - Users and communities needs
 - Advanced services
- Make (European) projects ready for action
 - Develop possible solutions
 - Establish business models

Project objectives

IST- 2001-320015

- Provide focus for dissemination
- Encourage collaborative development of software
- Support European liaison with OAI
- Exchange information
- *build community of interest*



Open Archives Forum

IST- 2001-320015

*Forum for exchange of experience
about all aspects of an
Open Archive approach,
including ...*



Susanne Dobratz, 17. Oct. 2002, Geneva, 2nd Workshop on OAI: Gaining Independence with e-prints and OAI

How OAForum works

IST- 2001-320015

- Workshops, Dissimination
 - Spread information
 - Create interest within new communities
- Organisational Validation
 - Explore Business Models
 - Issues of Ownership
- Technical Validation
 - Deployment of OAI technical framework
 - Create an information space on projects, software, implementations, services
 - Interoperability issues

OAF: Dissemination, Workshops

IST- 2001-320015

- May 2002 Pisa, Italy
 - Experiences in Europe from e-prints community
 - Create a forum
- December 2002 in Lisbon, Portugal
 - Archives and libraries
- March 2003 in Berlin, Germany
 - Museums and cultural heritage institutions
- September 2003 in Bath, UK
 - Big Dissimination event
 - All communities
 - In cooperation with other initiatives

OAF: Organisational Validation

IST- 2001-320015

- Business models
 - Data providers co-operating to build network of service provider
 - Metadata 'exchange' between data and service providers
 - Provision of added value services (metadata enhancement, auto classification, addition of OpenURLs)
- Issues of ownership
 - IPR and copyright (impact on authors and publishers)
 - Ownership of metadata (sharing metadata, metadata exchange and commercial agreement)
 - Long term digital sustainability (digital resources, scholarly communication, metadata)
- Discussion Group
 - by Dennis Nocholson (Strathclyde Univ. Glasgow)

OAF: Technical Validation

IST- 2001-320015

- Interoperability Issues Register
 - Integration with other existing technologies
 - Is unqualified Dublin Core sufficiently rich?
- Database management issues
 - Currency and update
 - Scaleability
 - De-duplication
- Software, Tools
- Efforts
 - Manpower
 - Skills
 - Time to set up an run a repository, service

Overview of European Activities

IST- 2001-320015

- Data Provider
 - Institutional Repositories
 - Subject „Gateways“
 - Library OPACs
 - Commercial Products
- Service Provider
 - Search Services
 - Extended Services
- OAI:
 - Tendency: not everyone wants to be listed
 - Does not fit for all communities

Overview of European activity on OAI

Leona Carpenter
Donatella Castelli
Susanne Dobratz

Overview of European Activities

IST- 2001-320015

- Data Provider
 - Subject „Gateways“
 - Institutional Repositories
 - Library OPACs
 - Commercial Products
- Service Provider
 - Search Services
 - Extended Services
- OAI:
 - Tendency: not everyone wants to be listed
 - Does not fit for all communities

Services: Extended Services

IST- 2001-320015

➤ CYCLADES

- EU project (CNR Pisa, Italy)
- Open Collaborative Virtual Archive Environment
 - Access Service to store information in a local database
 - Collection Service to structure workspace
 - Search and Browse
 - Filtering on individual user and community profiles
 - Recommendation service
 - Collaborative Workspace

Services: Extended Services

IST- 2001-320015

➤ TORII

- SISSA Italy
- provides a portal, where tools and documents are collected under a unified access point
- Personal workspace for scientists: rank documents according to user profile, impact factor
- Contains:
 - Okapi search engine
 - iCite extracts Citations from the documents to the archive
 - WAP

DP: Subject Gateways

IST- 2001-320015

➤ Medicine

Behavioral and Brain Science Prints Interactive Archive,
By Cambridge University Press, UK

<http://www.bbsonline.org/>

Uses eprints software, OAI

➤ Economics

RePec, Research Papers in Economics, UK

<http://repec.org/>

Subject Gateways

IST- 2001-320015

➤ Mathematics

Math-Net, University of Osnabrück, Germany

<http://www.math-net.de/services>

➤ Physics

PhysNet/ PhysDoc, University of Oldenburg, Germany

OAI-compatible

http://www.eps.org/PhysNet/uk_doc.html

Subject Gateways

IST- 2001-320015

- Cognitive Science
 - CogPrints, University of Southampton, UK
 - Uses eprints2 software, OAI
 - <http://cogprints.soton.ac.uk/>

- Psychology
 - University of Southampton, UK
 - <http://psycprints.ecs.soton.ac.uk/> (eprints 2)

➤ Educational Sciences

–Education-line

- conference papers, working papers and electronic literature which supports educational research, policy and practice.
- University of Leeds, UK
- <http://www.leeds.ac.uk/educol/index.html>

–Journal of Interactive Media in Education

- Knowledge Media Institute, The Open University, U.K. ,
- Center for LifeLong Learning & Design, Dept. Computer Science, University of Colorado at Boulder, U.S.A.
- <http://www-jime.open.ac.uk/>

➤ Others

–Organic eprints (eprints2)

- <http://www.orgprints.org/>

–Philosophy

- <http://sammelpunkt.philo.at:8080> (eprints2)
- Austria

Institutional Repositories

IST- 2001-320015

- Austria
- Belgium
- Denmark
- Finland
- France
- Germany
- Italy
- Ireland
- Portugal
- Norway
- The Netherlands
- Russia
- Sweden
- Switzerland
- United Kingdom

➤ Data Provider

–University of Viewwna eprints

•<http://epub.wu-wien.ac.at:80/dyn/OAI/oaicgi.pl>

–Philosophy

•<http://sammelpunkt.philo.at:8080> (eprints2)

Belgium

IST- 2001-320015

- **Data Provider**
 - Université Catholique de Louvain (UCL)
 - Université Libre de Bruxelles

- **Service Provider**
 - Université Catholique de Louvain (UCL)

Denmark

IST- 2001-320015

- Aalborg University - *Electronic Library*
<http://www.aub.auc.dk/phd/mainpage.html>

- University of Art and Design Helsinki (Dissertations)
http://www.uiah.fi/julkaisut/listat-en/lista_vaitos.html
- Helsinki University of Technology elect. acad. diss.
<http://www.hut.fi/Yksikot/Kirjasto/Diss/>

➤ Data Providers

–ATILF Resorces, CCSD theses-EN-ligne, Institut National de Physique Nucleaire et de Physique des Particules, The Lacito Archive, Univ. Lyon 2

- <http://jeanNicod.ccsd.cnrs.fr/perl/oai>
- <http://eprints.univ-lyon2.fr:8050/perl/oai>
- <http://documentation.in2p3.fr/oai/oai.php>
- <http://theses-EN-ligne.in2p3.fr/perl/oai>
- <http://lacito.archivage.vjf.cnrs.fr/servlet/metadata>
- <http://archiveSIC.ccsd.cnrs.fr/perl/oai>

➤ Services

–Cybertheses.org

- DINI: German Initiative for Networked Information
 - Recommendation to universities, libraries, etc. to follow the standards of the OAI released in December 2000
- Hosts workshops on different topics
 - Targeted to bring people together working on the same questions in the field of networked computing in general
- Working group on OAI
- <http://www.dini.de>



➤ Data Provider:

➤ University libraries document servers:

- HUBerlin (Humboldt Universität Berlin)
- SUUB Bremen
- MONARCH (TU Chemnitz)
- eldorado (Universität Dortmund)
- hsss (SLUB Dresden)
- DUETT (Universität Duisburg)
- University of Stuttgart, OPUS System
- Ludwigs-Maximilians-Univ. Munich

➤ Media server of universities

- TIMMS (Universität Tübingen)
- <http://oai.uni-tuebingen.de/server/metaserver>

➤ **Data Provider:**

➤ **University libraries catalogues:**

- Univ. Library Oldenburg

- <http://bis-oai.bis.uni-oldenburg.de>

➤ **Library service institutions for the region**

- BSZ-BW (Bibliotheksservice-Zentrum Baden-Württemberg) in Konstanz

➤ **Subject specific services**

- Physdoc (Universität Oldenburg)
- MathNet

- **Service Provider (not officially registered with OAI)**
 - HUBerlin Search Service incl. DINI Search
 - BSZ BW (Bibliotheksservice-Zentrum Baden-Württemberg) in Konstanz Search the union catalog of Baden-Württemberg
 - Physnet Search Service (Universität Oldenburg)
 - Open Archive Portal (ZDV Tübingen)
- **Projects:**
 - CYCLADES Access Service, University of Dortmund,
 - Prof. Norbert Fuhr, Gudrun Fischer
 - Elektra SISIS Library System
 - Institut für Informatik, Technische Universität München
 - Rudolf Bayer

The Netherlands

IST- 2001-320015

➤ Data Provider

- Tilburg University (Thomas Place)
- University of Utrecht (Dispute)
 - By end of 2002
 - scientific publications from the Utrecht University
 - <http://dispute.library.uu.nl>
- University of Nijmegen
 - **Elektronische publicaties KUN (WebDOC)**
 - <http://www.kun.nl/ubn/Algemeen-extern/Kundocs1.htm>
- University of Maastricht Electronic Documents
 - <http://www-edocs.unimaas.nl>

➤ Data providers

- Sissa Multimedia Database
- Eprints.org
(Bologna, Florence, Padua, Trento, Milan Universities)

➤ Service providers

- Cyclades (CNR-Pisa)
- TORII
 - Tools for Innovative Publishing in Science (SISSA-Trieste)

- DataProvider
 - <http://eprints.may.ie/perl/oai>

Portugal, Norway

IST- 2001-320015

- No official data and service providers

Internal experimentation

- The National Library of Portugal (PORBASE)
- BIBSYS
- The National Library of Norway

The Netherlands

IST- 2001-320015

- Service Provider
 - Not yet
- Projects
 - ARNO project
 - Tilburg University, University of Twente, Amsterdam University
 - FIGARO (??OAI)

Russia

IST- 2001-320015

- Data Provider
 - <http://edu.ioffe.ru/oai/>

➤ Data Providers

- Uppsala University Digital Archive
- Two universities are experimenting eprint.org software
 - (EPSILON EPrints2 Dissertation Test Archive)
 - Blekinge Institute of Technology: Electronic Research Archive (<http://www.hk-r.se/fou/>)
 - Karolinska Institute (<http://diss.kib.ki.se/info/proj/>)
 - Luleå University of Technology: Publications from LTU (<http://epubl.luth.se/index-en.html>)
 - Lund University: LU:research (<http://eprints.lub.lu.se/>)

- A new project DIVA
 - <http://publications.uu.se/portal>
 - Stockholm University
 - Södertörn University
 - Umea University
 - Uppsala University
 - Örebro University
- Service providers
 - Internal university local services

- CERN document server
<http://cds.cern.ch/>
See Martin Vesely

- Developments with international impact
 - University of Southampton:
 - Tools development
 - **eprints.org** self archiving application
 - **CiteBase** harvester
 - CogPrints archive
 - Evangelism for the open archive approach
 - Scientific American, print & online debate

➤ Data Provider

- CogPrints
- Dep. Comp. Science TR (U Southampton)
- Univ. of Nottingham
- ePrints@Glasgow
- *Education-line* (University of Leeds, UK)

➤ Service Provider

- CogPrints, etc.
- RDN (UKOLN) <http://www.rdn.ac.uk/oai/nph-oai.cgi>

- Repositories and service providers
 - many institutional archives planned and starting;
 - newly-funded projects include some service providers
- Beyond e-prints
 - for interoperability across image repositories:
 - portal for NoF-digitisation projects;
 - other heritage initiatives
 - for “virtual union catalogue” creation (UKOLN)
 - harvesting via OAI in architecture of Resource Discovery Network (RDN)
- **JISC Project RoMEO.**

The **JISC** has funded a one year project to investigate the rights issues surrounding the 'self-archiving' of research in the UK academic community under the OAI-PMH.

Questionnaire: Used Software

IST- 2001-320015

➤ Data Provider

- Selfdeveloped: 8 → available for others: 7 • open source: 5
- Not Selfdeveloped: 5 → open source: 4

➤ Service Provider

- Selfdeveloped: 8 → available for others: 4 • open source: 3
- Not Selfdeveloped: 1 → open source: 1

➤ Programming Languages

- PHP: 4 • Perl: 8 • Java: 11 • C: 2
- Visual Basic: 1 • Tcl: 1

■ eprints

- <http://www.eprints.org/>
- Creator: Christopher Gutteridge, University of Southampton
- Data Provider Software
- Service Provider Software
- used by Université Catholique de Louvain (UCL)
- run centralised, although archives are distributed.
- Requirements
 - UNIX, Linux
 - Apache WWW server
 - Perl programming language
 - MySQL Database

➤ VT ETD-db

- <http://scholar.lib.vt.edu/ETD-db/>
- Creator: Anthony Atkins, Virginia Polytechnic Institute and State University
- Data Provider Software
- used by Université Catholique de Louvain (UCL)
- provide a standard interface for web users to enter and manage metadata related to a collection of electronic theses and dissertations.
- Requirements
 - UNIX server platform
 - Apache WWW server
 - Perl programming language, CGI.pm
 - MySQL Database

- Questions concerning
 - the Implementation
 - Costs

Questionnaire: Know How

IST- 2001-320015

- **Data - & Service Provider**
- System Administration (UNIX | Linux)
- Web Server Configuration (Apache)
- Knowledge on Databases and SQL
(MySQL | Sybase | Oracle)
- Programming
(Perl | Php | Java | Servlets | CGI | XML)
- Experiences with Metadata

Questionnaire: Time & Manpower

IST- 2001-320015

➤ Implementation expenditure:

DP: ≤ week: 3 • ≤ month: 4 • ≤ quarter: 2 • several years: 2

SP: ≤ week: 1 • ≤ month: 2 • ≤ quarter: 2 • ≤ half-year: 1

➤ Involved Programmers:

DP: 1 Progr.: 9 • 2 Progr.: 2 • 20 Progr.: 1

SP: 1 Progr.: 6 • 2 Progr.: 2 • 3 Progr.: 1

➤ Expenditure to keep the OAI implementation running:

1 person day per month (once: 5 person day/month)

DP – Object Quantity & - Type

IST- 2001-320015

➤ Number of Documents:

490 - ca.800 - 1.600 - 3.000 - 12.000 - 140.000 - 400.000 -
600.000 - 7.000.000 - several million

➤ Disc Space:

15 MB - 150 MB - 1,4 GB - 1,5 GB - 2,5 GB - 10 GB - 300 GB
- 2 TB

➤ Type of objects:

| | |
|-------------------------|--------------------------|
| Fulltext documents (10) | Metadata (10) |
| Image files (4) | Video files/ streams (3) |
| Sound (1) | |

DP – Content Type

IST- 2001-320015

- Preprints (5) Dissertations (5)
- Journal articles (6) Lectures (3)
- Conference Proceedings (1)
- Short Articles (1)
- Video streams of University Events (1)
- Library Catalogue (metadata records for books, periodicals, video,...) (1)
- All publications produced by the academic staff of the university (1)
- Recordings (1)
- Earth Observation Satellite Images (1)

- Questions regarding
 - Content Type,
 - Structure and Integration
 - of Archive / of Service

DP – Metadata Formats

IST- 2001-320015

- **Metadataformats:**
- Dublin Core (6)
- Qualified DC stripped down to OAI-DC
- Dublin Core Library Profile
- MARC21 (2), German MAB
- RIS, UNIMARC, DiTeD (internal format for thesis and dissertations), CEOS CIP
- Internal format that can be converted to any standard metadata format
- **Disseminate all parts:** 7 Yes • 5 No

DP – Open Access to OAI Interface

IST- 2001-320015

- **OAI interface open:** 7 Yes • 5 No
- **Restrictions:**
 - IP adress controlled
 - Swets licences required
 - Most of the data is only accessible through a search form
 - Awaits funding and management approval for wider access when the service is developed

SP – Kind of Services

IST- 2001-320015

- OAI-Service / Portal
- Searching and browsing for information
- Search in different sources through one form
- Workspace for managing documents and metadata
- Cross linking, annotations

SP – Harvesting Level

IST- 2001-320015

- harvesting level is ok: 2
- better metadata
- the quality of metadata is just a big mess
- more standardized content in the DC metadata, for example, standard ways to specify names, dates, languages
- more structure than unqualified DC
- did some adaptation of the protocol for local service / use different unique keys

- Questions about
- Experiences
- and Future planning

DP - Importance

IST- 2001-320015

- **Provide additional services to existing services**
- Supplement to a paper catalogue
- Delivery of electronic holdings info to an OpenUrl resolver system
- Possibilities of dissemination of information about scientific results of our researcher

DP - Importance

IST- 2001-320015

- **Replace existing services through OAI Interface**
- Replace the single, non-searchable lists of eprints available on the websites
- Develop the practice of alternative means of dissemination of the scholarly communication
- Research project. We used to run a search-engine which tried to combine different HTML-outputs from different sources. This has been replaced by an compatible OAI search-engine

DP - Importance

IST- 2001-320015

- **Better Retrieval, make Metadata exchange available**
- Harvesting own archive via OAI (replacement of search engine)
- Interface for metadata exchange within several projects (new service)
- Exchange of our library catalogue with other university in Brussels
- Integration into virtual union catalogue of Belgium
- Want to be able to develop services based on OAI compliant archives maintained by others

DP – OAI-Advantages

IST- 2001-320015

- Chance to share scientific knowledge and to harvest other knowledge databases
- Opportunity to import metadata in Libraries software
- Major dissemination of researchers' results
- Simple implementation & easy adaptation for project internal usage
- OAI provides a simple to implement facility of exchanging metadata (Z3950 is probably too complicated to implement)

DP – OAI-Advantages

IST- 2001-320015

- Research
- In many cases the archives are too small for a Z39.50 service. By harvesting relatively small archives it is possible to maintain a Z39.50 service.
- Easy (and quick) implementation, minimal maintenance
- Provides a standard approach for metadata harvesting which will simplify extension of a pilot system. Software for prototype freely available
- OAI is easy to implement...

DP - Experiences

IST- 2001-320015

- No experiences / still testing
- It just runs
- Good!

- Multiple services have become available, with only ONE SIMPLE implementation on our catalogue
- Very good impact on usage of information in our archive

SP – Which Services

IST- 2001-320015

- get metadata records
- harvesting, annotation, cross linking
- search-engine
- We use the information for a metadatatatalogue (<http://publications.uu.se/metadata>). Currently there are metadata from 5 local databases - about 5 000 records

SP - Problems

IST- 2001-320015

- none
- Problem about the different semantic by defining set
- Different formats of metadata...
- As a harvester of RePEc archives for building a Z39.50 service, our main problem is the quality of metadata.

SP – Plans for the Future

IST- 2001-320015

- search & browse, collaboration environment for users and groups of users, discussion forums, annotations, awareness
- We are planning on implementing a virtual union catalogue for Belgium using OAI (several millions of metadata records). We will be trying out software like Open Digital Libraries (<http://oai.dlib.vt.edu/odl/>) and ALCME (<http://alcme.oclc.org/index.html>) in the near future.

SP – Plans for the Future

IST- 2001-320015

- We intend to develop a resource discovery service for contents related with our mission (the Portuguese science and technology, culture, history and society in general). An alerting service, coordinated with the national union catalogue, is also under consideration.
- Z39.50 services integrated in an information portal (iPort of OCLC|Pica). This will allow searching, browsing, document delivery services, current awareness etc

SP – Plans for the Future

IST- 2001-320015

- Integrate the search engine with compatible new OAI protocol
- The search engine is already running, but more sources will be included. Implementation of version 2.0.

➤ Weak points in the OAI approach to interoperability

The protocol is easy to implement for data providers, but the heterogeneity of the content of the metadata records requires the service provider to invest a lot of effort in normalizing the data

Much of this standardization could be done at lesser cost by the individual data providers (less heterogeneity of the data)

Possible solution: development of middleware tools that service providers could use for data normalization

Another suggestion: define additional metadata formats (or use existing ones) and convincing data providers to export them, too.



Open Archives Forum on the Web

IST- 2001-320015

- Glossary
- News
- Questionnaire 1 + 2
- Workshops
- Database



Susanne Dobratz, 17. Oct. 2002, Geneva, 2nd Workshop on OAI: Gaining Independence with e-prints and OAI

- Questionnaire
- about the
- Technical Validation

Who participated?

IST- 2001-320015

- 21 answers
- 6 Germany • 6 Italy • 2 Belgium •
2 Netherlands • 1 France • 1 Sweden • 1 UK •
1 Portugal • 1 Norway
- With experiences (Implementation 2001):
7 Data Provider - 4 Service Provider
- In planning -, development -, test phase:
6 Data Provider - 7 Service Provider

Questions to the Audience

IST- 2001-320015

- How difficult is it to integrate an OAI Data and/or Service Provider in existing infrastructure, that are based on different technologies? (Z39.50, NCSTRL, REPEC, ????, Metalib, etc.)
- What are your demands on the further work of the Open Archives Initiative committees (technical, steering committee)?

Questions to the Audience

IST- 2001-320015

- What are potential further services to be offered/developed in future?
- How are sets defined? Which classification schemas are used?
- Which meta data formats are used in addition to DC simple?
- How were copyright issues technically solved so far?
- How can a machine-readable rights statement be encoded?

Thank You!

IST- 2001-320015

- Please contribute to our database!
- http://www.oaforum.org/oaf_db/oaf_index.php
 - Information about your projects
 - Your implementation and usage experience
- <http://www.oaforum.org>
 - Leona Carpenter
 - Donatella Castelli
 - Susanne Dobratz

2nd Workshop

<http://www.oaforum.org>

OPEN ACCESS TO HIDDEN RESOURCES

Lisbon, Portugal
5-6-7 December 2002

Venue: Biblioteca Nacional



For a long time libraries and archives have collected documents and preserved them for use. Now, much of this material is being digitised, or starts life as digital. So the foundation is laid for wider availability. The aim of this workshop is to explore whether and under what conditions the open archive approach is viable for these organisations.

Contact: borri@iei.pi.cnr.it

A handwritten signature in black ink, likely of the person who designed the poster.

Used Information

IST- 2001-320015

- Open Archives Initiative
www.openarchives.org/service/listproviders.html
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>