open archives forum

2nd Technical Validation Questionnaire

- interim results -

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

IST- 2001-320015

Why this technical questionnaire?

➢ 1st Technical Validation Questionnaire

- provide an overview on status, experiences and future plans belonging OAI implementations of participants of the 1st OAForum Workshop
- target group: workshop participants

➢ High Interest, Feedback

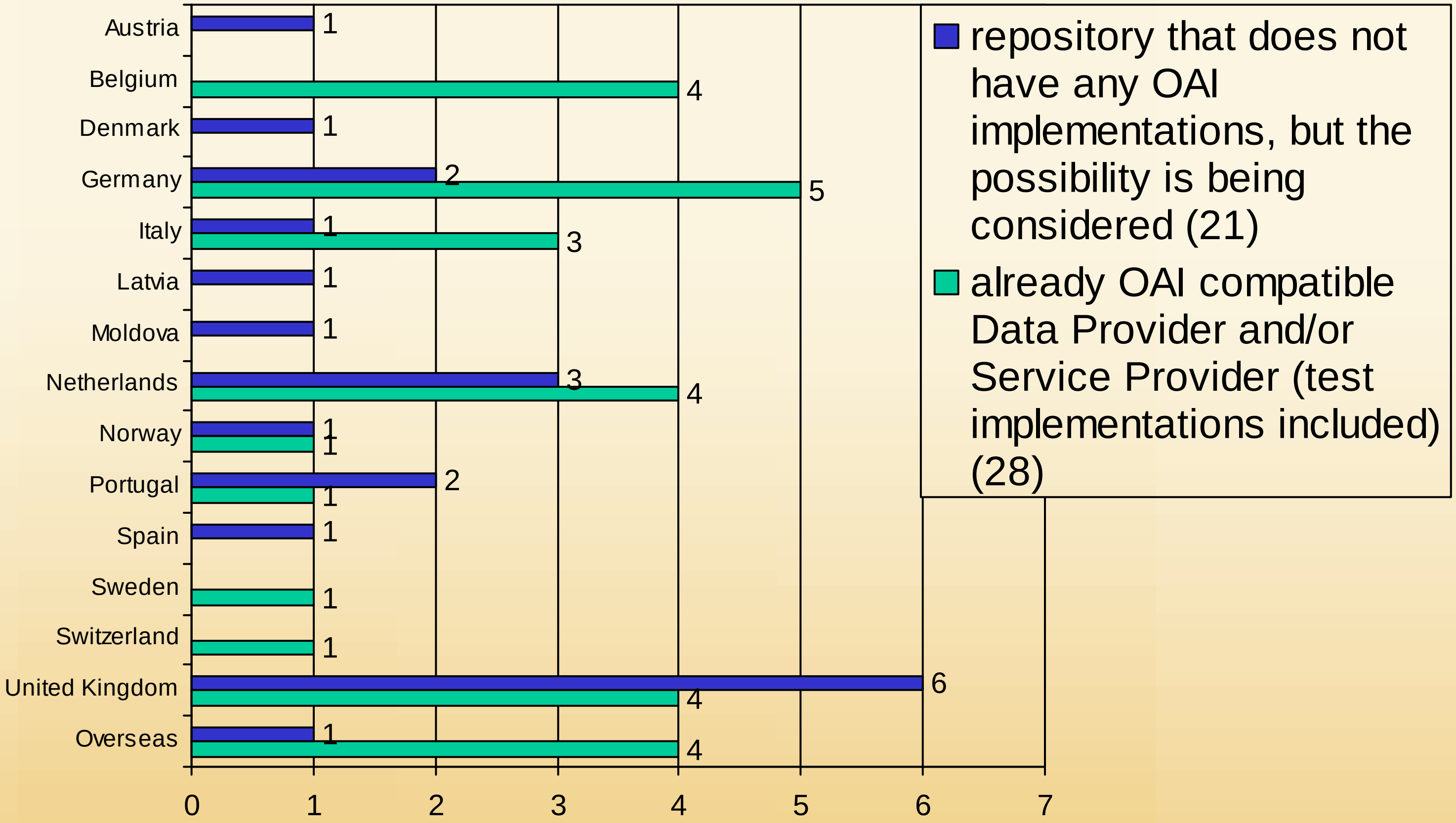
- to collect experiences of a broader spectrum
- to learn more about starting conditions of planned implementations
  - Is there large common ground?
  - Are requirements so individual that it will be necessary for many isolated solutions to be developed?
  - Should tools and protocols correspond more than now to the needs of different communities?

What are the Goals?

- **Extended 2nd Questionnaire**
  - extended questions + target audience + duration
  - new subdivision in two questionnaires
    - technical presuppositions of those, which have not yet integrated OAI-PMH
    - experiences of implementers
- **to get information about**
  - used software
  - implementation costs
  - offered spectrum and interoperability
  - experiences and expectations
    - → in different communities
    - → in different countries
- **to share experiences and information about technical issues related to open archives**

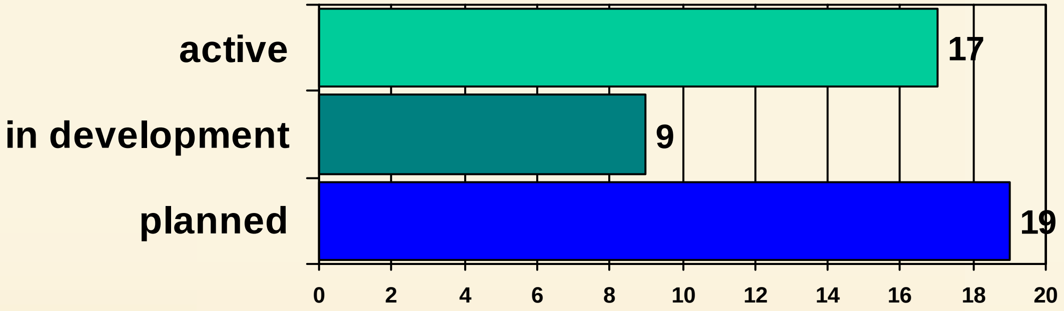
Who participated to date?

- Countries

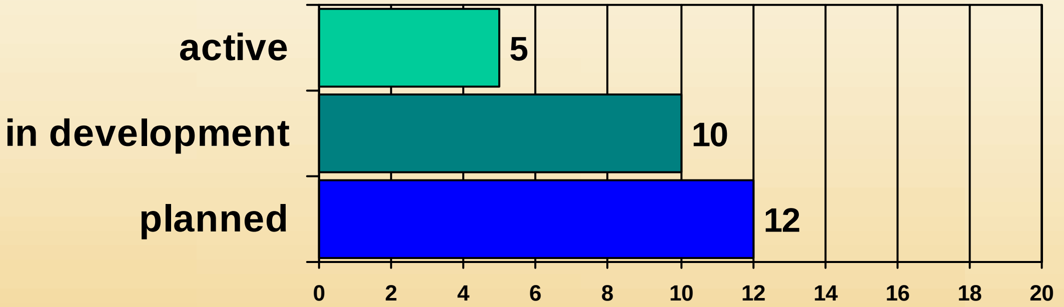


Who participated to date?

- **Data Provider**



- **Service Provider**



30 % of active DP are also SP

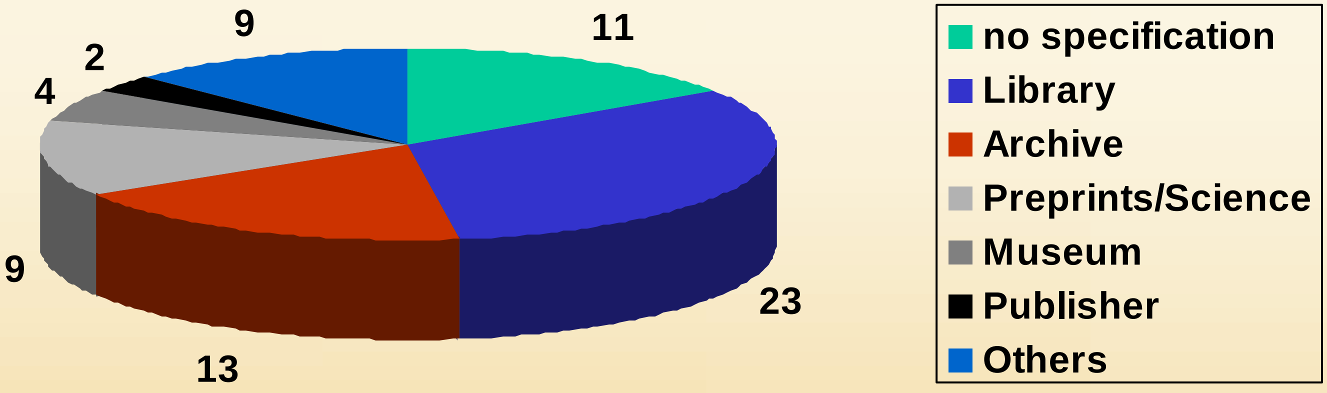
41 % of active DP plan or still develop SP implementations

IST- 2001-320015

Who participated to date?

- **Communities**

  Multiple answers possible



Used Software

- **Technical infrastructure before OAI-Implementation**
  - not many statements to Interface and Collection Systems
  - dominant programming languages:
    - Perl, XML, also Java, PHP
  - dominant databases:
    - MySQL, Oracle

- **Almost no one changed existing software tools to be OAI compatible**

Used Software

➢ Implementations to be OAI compatible

- about 60 % of the used tools were self-developed by both Data- and Service Providers
    - most of them make their developments and the source code available for others
    - dominant programming languages: Java, Perl, PHP, also XML
- tools like PERL implementations, OAI Cat, EPrints, and OAI Harvester were mentioned 3 or 4 times each

→ list of OAI-PMH software: http://www.openarchives.org/tools/

IST- 2001-320015

Implementation Costs

Necessary Know How: Data- & Service Provider

focused on various combinations of the following **five competence fields**:

- system administration (UNIX | Linux)
- web server configuration (Apache)
- knowledge on Databases and SQL (MySQL | Sybase | Oracle)
- programming skills (Perl | Java | PHP | Servlets | CGI | XML)
- experiences with metadata

IST- 2001-320015

Implementation Costs

➢ Time and Manpower

- **implementations of OAI-specifications:**
  - 75% concluded within a quarter by one programmer (span: from 2 to 750 personal days per month)
- **reasons for few bigger expenditures:**
  - context of bigger research projects
  - construction of archives
  - processing of bigger data amounts
- **further maintenance for a stable protocol:**
  - max. 25, mostly 1 personal day per month

IST- 2001-320015

Implementation Costs

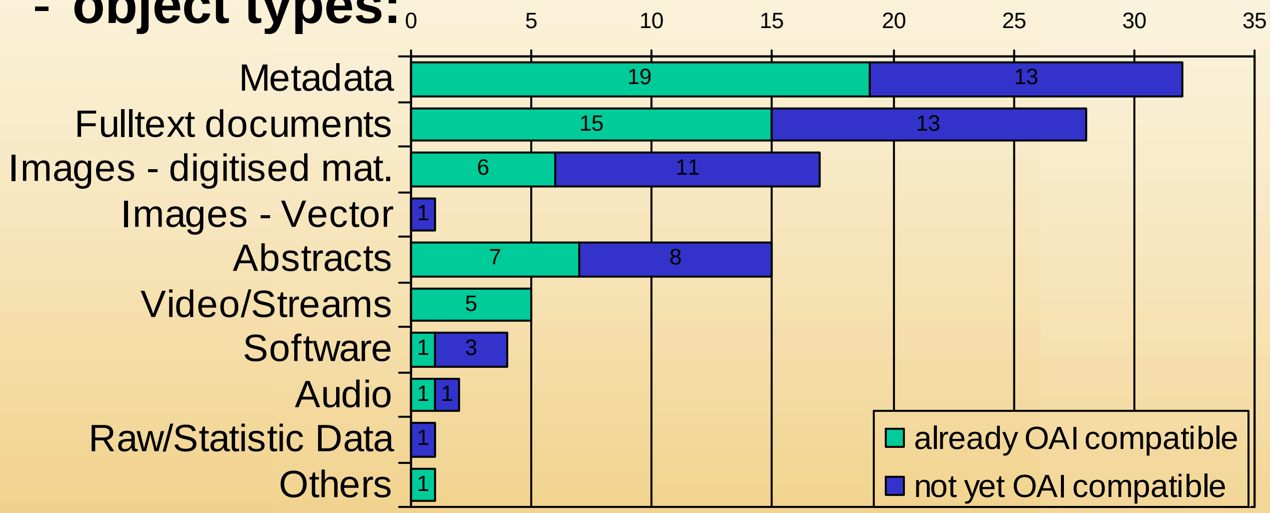
- **Expectations of those who haven't implemented yet**
  - **implementations of OAI-specifications** (same):
    - ° concluded within a quarter by one programmer
  - **further maintenance for a stable protocol** (higher):
    - ° up to 40 personal days per month
  - No specific trend recognizable with expectations if
    - ° data structures suggested by the OAI-PMH are easy to integrate in existing infrastructure
    - ° the adaption of the data to the OAI-PMH will be expensive
    - ° the preparation of the data for an internet usage will be expensive

IST- 2001-320015

Offered spectrum - DP

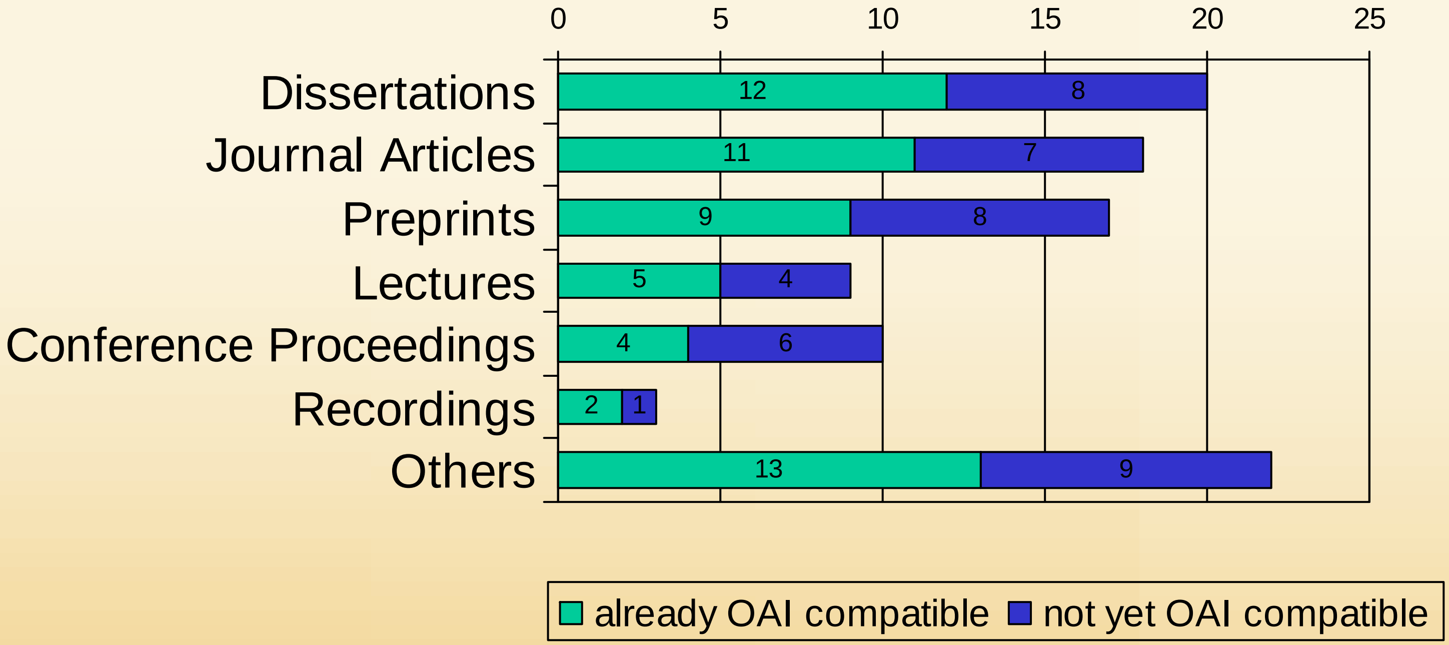
Offers of Data Providers

- **number of documents:**
  - between 5 and several million documents
- **storage space:**
  - between 1 megabytes and 2 Terabyte.
- **object types:**



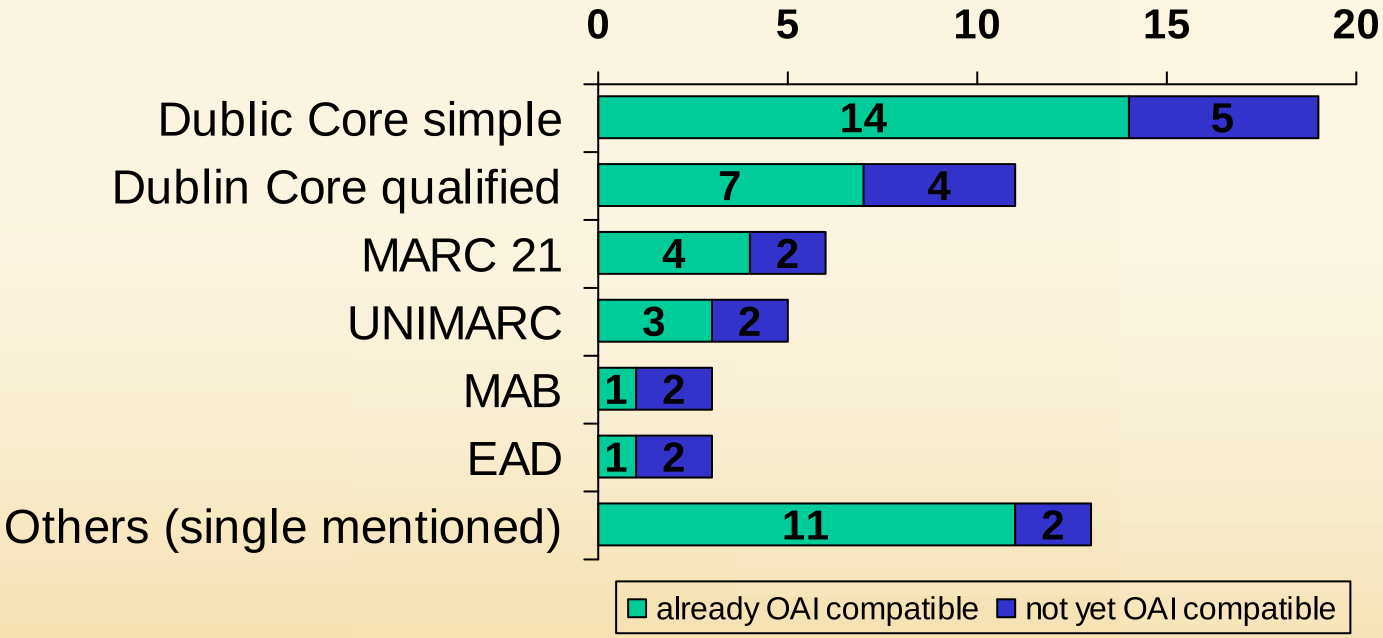
Offered spectrum - DP

- **Content types**



Offered spectrum - DP

- Metadata formats



Single mentioned formats:

Dublin Core Library Profile, DiTeD, CEOS CIP, AMF, RIS, MODS, METS, SPECTRUM, TEI, internal format, self developed

Offered spectrum - DP

- **Dissemination**
  - more than half of the Data Providers are offering all parts or rather extracts of the documents
  - if the openness of the OAI interface is reduced due to several reasons, people use two limitation strategies:
    - access control
      (control of the IP-addresses, licensing, agreements)
    - limitation of the data output

IST- 2001-320015

Offered spectrum - SP

➢ Kind of Services

- OAI-Service / Portal
- local or community specific services
- searching and browsing for information
- search in different sources through one search interface
- cross-linking, annotations, harvesting
- workspace for managing documents and metadata, collaboration within groups of users
- document management

Offered spectrum - SP

- **Stategies to process with harvested data from DP**
  - use no provenance information
  - filter harvester output and load local database
  - strategies to include information about DP in data output:
    - when a metadata record is found, the user can also browse information on the archive the record came from
    - queries against the portal return data sets as harvested, including information about the original data provider
    - provenance information is encoded in the identifier

Experiences - DP

- **Importance / Advantages of OAI**
  - provide additional services to existing services
  - replace existing services through OAI interface
  - better retrieval, make Metadata exchange available
  - share scientific knowledge, harvest other knowledge databases, cross-search in institutional assets
  - major dissemination of researchers' results
  - simple and cheap in implementation
  - easy adaption for project internal usage
  - simple to implement facility of exchanging metadata in comparison to more complex protocols

→ „provide access to all of human knowledge“

→ „nothing other than political expediency“

Experiences - SP

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 data provider
  - development of middleware tools that service providers could use for data normalization

Experiences - SP

Future Plannings

- extend search & browse functions
- export in other formats such as XML
- document delivery services, print on demand
- collaboration environment for users and groups of users, discussion forums, annotations, awareness
- extend existing services, building distributed services
- establish an exchange of different library catalogues and the integration into a virtual union catalogue for the whole country
- create a single catalogue of all library's catalogues: library opac, archives database, image database, Internet gateways

IST- 2001-320015

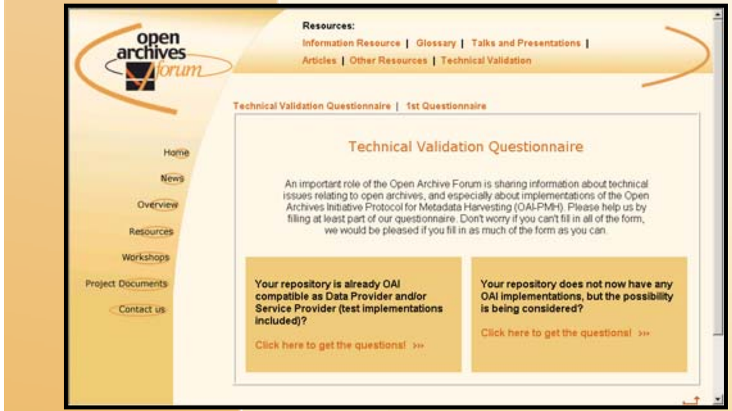
Useful information sources

- **Problems to find useful informations?**
  - Many of those who haven‘t implemented yet made the experience that it is laborious to find good informations about metadata and especially technical support
  - Some asked for a gentle introduction to the protocol („too jargonish“)
- **Recommendations of the questionnaire participants**
  - **Websites**
    - www.openarchives.org
    - www.ndltd.org
    - www.cimi.org
    - www.eprints.org
    - www.rlg.org
    - www.oaforum.org
    - www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/faq/oai
    - http://library.cern.ch/heplw/4/papers/3/
  - **Online journals** eg. Ariadne, D-Lib Magazine
    - www.ariadne.ac.uk
    - www.dlib.org
  - **Conferences** and workshops
  - **Informal discussions** with other gateway managers
  - **Test programs** eg. http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai

IST- 2001-320015

Thank You!

- **Please contribute!**
  - Information about your projects
  - Your implementation and usage experience


open archives forum
Resources:
Information Resource | Glossary | Talks and Presentations |
Articles | Other Resources | Technical Validation
Technical Validation Questionnaire | 1st Questionnaire
Home
News
Overview
Resources
Workshops
Project Documents
Contact us

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.

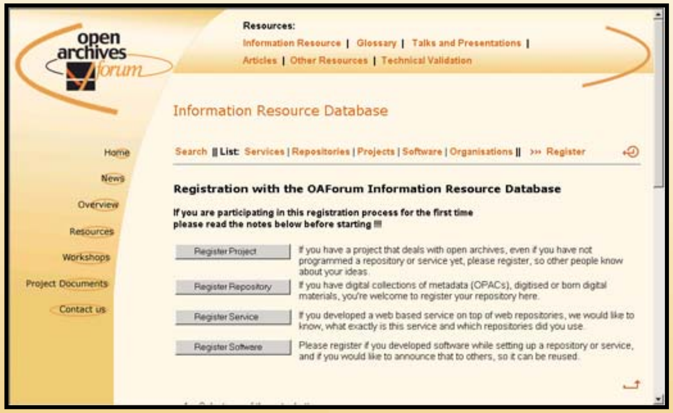
Your repository is already OAI compatible as Data Provider and/or Service Provider (test implementations included)?

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! »»

Technical Validation Questionnaire
http://www.oaforum.org/resources/tecvalq2.php


open archives forum
Resources:
Information Resource | Glossary | Talks and Presentations |
Articles | Other Resources | Technical Validation
Search || List: Services | Repositories | Projects | Software | Organisations || »» Register
Home
News
Overview
Resources
Workshops
Project Documents
Contact us

Information Resource Database

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 !!!

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.
Register Repository	If you have digital collections of metadata (OPACs), digitised or born digital materials, you're welcome to register your repository here.
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.
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.

Information Resource Database
http://www.oaforum.org/oaf_db/