Towards the
Self-filling Repository
Santa Fé, New Mexico, 21-22 October 1999

Open Archiving Initiative

14 June 2002, OAI-PMH v. 2.0
SURF DARE Programme

- **Approved:** 14 June 2002
- **Name:** DARE,
  Digital Academic Repositories
- **Period:** 1 Jan. 2003 – 31 Dec. 2006
- **Budget:** M€ 5.9
- **Standards:** OAI-PMH 2.0; Dublin Core
- **Partners:** All universities, KNAW, NWO and KB
Services

Society
Institutional windows, expertise, professional journals, personal Web sites, national windows, ...

Research
Subject repositories, refereed portals, databases, collaboratories, (Open Access) journals, ...

Education
Virtual Learning Environments, Course Ware, Readers, ...

Harvesters

Institutional Repository

Data

Grid

waaijers@surf.nl  OAI4, Geneva  20 September 2006
From Library to ‘Libratory’

Services

- PR, academics, libraries, DARE
- Institutional windows, expertise, professional journals, personal
- Web sites, national windows, ...

Research

- Subject repositories, refereed portals, databases, collaboratories, (Open Access) journals, ...

Libraries, Google Scholar!

- Teachers, ICT-ers, libraries
- Virtual Learning Environments, Course Ware, Readers, ...

Data

Institutional Repository

Tilburg

NWO

....

....

Southampton

Grid

waaijers@surf.nl

OAI4, Geneva

20 September 2006
Embedding

Services

Society
PR, academics, libraries, DARE

Research
Subject repositories, refereed portals, databases, collaboratories, (Open Access) journals, ...

Harvesters

Libraries, Google Scholar!

Education
Teachers, ICT-ers, libraries

Institutional windows, expertise, professional journals, personal web sites, national windows, ...

Institutional Repository

2005
2006

e-Depot
DANS

Data

Grid

Amsterdam
Delft

Nat.
The.
Auth.
Thes.

Metis

waaijers@surf.nl

OAI4, Geneva
20 September 2006
Embedding

Services

PR, academics, libraries, DARE

Libraries, Google Scholar!

Research
Subject repositories, refereed portals, databases, collaboratories, (Open Access) journals, ...

Education
Teachers, ICT-ers, libraries

Institutional windows, expertise, professional journals, personal Web sites, national windows, ...

Harvesters

Institutional Repository

Society
Amsterdam
Delft
Cornell
MIT

Data

Grid

e-Depot

Nat. Auth. Thes.

2005
2006

waaijers@surf.nl

OAI4, Geneva

20 September 2006
**OAIster:** 536 OAI repositories; 5.9 M objects

**DAREnet:** per end 2003 15 repositories (all universities + KNAW +NWO), today 49.000 objects.
Some figures

- 15 Institutions
- 207 authors (187 male, 20 female)
- 40479 records = 195/author (from 3 to 1224)
- 23853 full text = 58.7% (from 19% to 96% per institute)
- 25% copyright obstructed, 15% only metadata available at the moment, 2% lost
Some lessons

- Authors’ enthusiasm (not interested in drafts versions, only in the real thing)
- New technology (sets/filters, resource harvesting, new metadata, standard jump off page)
- Limited co-operation of publishers
- Copyright problems not too bad
- The Netherlands are in the forefront

waaijers@surf.nl

OAI4, Geneva

20 September 2006
Keur is overlapping DAREnet

DAREnet = Σ open sets in 15 IR’s
Keur = Σ 207 individual oeuvres
Under way

- HunDAREd thousand
- Promise of Science (incl DIDL)
- Copyright (with JISC)
- Subject based (refereed) portals (filters needed!)
- Datasets (=> e-science)
- Newsfeeds, Annotations, Student theses, Expertise
- Personal and Institutional web sites, ...
- LOREnet (incl. compound objects)
LOREnet

ELO’s
- BlackBoard
- N@tschool
- TeleTop

Repository's
- ARNO
- Intralibrary
- Hive

............... IMSCP

LOREnet portal
Harvester

............... OAI-PMH

............... SOAP-XML

............... waaijers@surf.nl

OAI4, Geneva
20 September 2006
Spring 2005: Transfer to SURF Search Engine
- based on FAST
- OAI-PMH compliant harvester added
- DARE web application added
- Full text indexing planned
- Drill down on type, author or subject planned
Search tool for service providers

Google Scholar, Yahoo, Scirus dominate the consumer market. Needed is a professional search tool that enables service providers to cater.
Suppose

- resource harvesting,
- jump of page/XML container/DIDL,
- sets and filters

are all solved.

What’s next?

waaijers@surf.nl
Intelligent content spider

Input: digital resources residing on distributed heterogeneous platforms (servers, websites, pc’s)

Output: initially a great recall of resources and locations, gradually increasing precision through self learning

(=> efficiency)
Metadata generator

Input: a digital resource with a unique identifier
Output: generated metadata, DIDL compliant, in various formats (DC, Marc 21, METS, LOM) potentially combinable with classifier (=> efficiency)
Classifier

Input: a digital resource with a unique identifier

Output: classification metadata relevant for resource with selectable classification schema.

(=> subject repositories)
Modules

Packager

Input: separate records referring to parts (chapters, modules, subsets) of an object

Output: packaged records, e.g. as IMS-CP or DIDL package

(=> elo’s)
Citation analyzer / authority measure
Input: the content of one or more (disciplinary) repositories
Output: the citation indexes or usage indexes of the publications in these repositories
(=> prestige)
Modules

Associative searching

Input: the content of one or more (disciplinary) repositories

Output: a (graphical) presentation of associative relations between the objects

(=> public exposure)
Promises to authors

- Individual visibility (personal and institutional web site, Google Scholar etc.)
- Registration (one off and easy)
- Speed (a reliable draft version)
- Preservation

The only thing we cannot guarantee is publication in a prestigious journal
Needed

A simple alternative for the forbidding current copyright statement of the classical subscription publishers.

This is the only thing authors want re. copyrights (=> SURF-JISC survey)
Promises to managers

- Institutional visibility
- Knowledge housekeeping
- Cheaper circulation of knowledge
Needed

The manager’s attention

waaijers@surf.nl        OAI4, Geneva        20 September 2006
All you need is DARE!

www.surf.nl/dare