



Long term preservation → and Institutional Repositories



Barbara Sierman
Digital Preservation Officer
22-05-2007



Overview

- KB, the National Library of the Netherlands
- Digital Preservation at the KB
- Dare and Dare-net
- Dare publications in the KB e-Depot



Koninklijke Bibliotheek

- Founded in 1798
- National library of the Netherlands
- 275 FTE staff
- 3,5 million books, 9 million electronic objects
- Funded by the Ministry of Education, Culture and Science
- (Voluntary) Deposit library since 1974



E-Depot



- 1994: E-publications part of deposit
Need for an infrastructure and digital archive
Studies & pilots
- 2000 Tender procedure: IBM chosen
- 2003 Digital Information Archiving System: DIAS,
based on OAIS
- 2002 – now
Archiving agreements with major STM publishers
Elsevier Science, Kluwer Academic, BioMed
Central, Blackwell Publishing, Oxford University
Press, Taylor & Francis Group, Sage, Springer
- 9 million journal articles

Current developments e-Depot



- Safe place & trusted repository for scientific journals
- National deposit of e-publications
- Web-archiving of selection of Dutch websites

Organisation of Digital Preservation at KB

- 3 KB departments involved in e-Depot
 - IT department
 - e-Depot department
 - R&D department on Digital Preservation
 - Technology watch
 - e-Depot related activities:
 - Metadata (Premis), File formats
 - Preservation Strategies: migration, emulation
 - National and international knowledge exchange
 - Participating in European projects: Planets, Driver-II

DARE (Digital Academic Repositories)



- 2003 Digital Academic Repositories (DARE)



Goal: to make the research output digitally accessible and available for different services

- Partners:
 - 14 Universities and Organisations of Higher Education
 - Netherlands Organisation for Scientific Research
 - Royal Netherlands Academy of Arts and Sciences
 - KB, National Library of the Netherlands
 - Coordination: SURF Foundation

DARE (Digital Academic Repositories)



- DARE-net; network of university repositories in the Netherlands (2004)
- Agreement on repository principles:
 - Repository is OAI compatible (Open Archiving Initiative)
 - Descriptive metadata in Dublin Core Simple
 - Structural metadata in MPEG21-DIDL
 - Metadata in XML
- Special attention for preservation
- Role for KB, National Library of the Netherlands
 - e-Depot

Agreement KB and DARE

- No official contract, but practical approach
- In accordance with e-Depot current practice
- Submission on voluntary basis
- What is a publication? What is scientific?

“Those objects that are made available on the public Internet through Institutional Repositories”

- DARE partners decide individually what they'll deliver
 - No data sets
 - No restricted publications
 - Not under embargo etc.

Harvesting of DARE material

- Based on OAI-PMH protocol
- KB starts harvesting, in consultation with university (day & time)
- Initial harvest 2006
- Regular updates 4 times a year

File formats in DARE

- Mostly PDF files (90%)
- Around 20 popular other file formats
- Written agreement allows this set
 - Zip files allowed uncompressed
 - Only those 20 file formats in Zip
 - Dynamic webpages received as HTML pages
- Negotiation about new file formats
- Investigation of normalisation

Checking of DARE objects

- Virus check
- Dublin Core metadata conform agreement?
- Check on completeness (object + metadata)
- JHOVE file identification check
- Extraction of technical metadata with JHOVE

- DARE partners informed about errors

Accessibility of DARE material

- Descriptive metadata stored in KB catalogue
- All DARE publications accessible via KB catalogue
- Choice offered:
 - Either he uses link to original repository
 - Or he uses the link to e-Depot
 - DARE partners prefer link to their repository

DARE search

Search

First choose a database (set) and then fill in your search term(s).

Identify database(s):

KB-Sets

Search in: All Fields [v] [v] And [v] [Search]

Set: Full-text contains 6 databases:

Database Name	Type	Actions
<input type="checkbox"/> KB e-Depot	Searchable	[i] [d]
<input checked="" type="checkbox"/> DARE	Searchable	[i] [d]
<input type="checkbox"/> DATOR Arts & Sciences - Collections I-II-III	Searchable	[i] [d]
<input type="checkbox"/> DATOR [Full Text] [New]		
<input type="checkbox"/> Project BRISA [Full Text]	Search and Link	[i] [d]
<input type="checkbox"/> SwetsWise [Full Text]	Searchable	[i] [d]
<input type="checkbox"/> Periodicals Archive Online (ProQuest) [Full Text]	Searchable	[i] [d]

Batch delivery

- On request of the DARE partner
 - In case of emergency
 - KB not a back up functionality
 - In case of file format obsolescence (future use)
- Selection of material
 - Date of ingest
 - File format + version
- Most recent version returned
 - Older versions will be kept

Long term preservation

- KB will archive the DARE objects:
 - In its original form (bitstream)
- KB commitment to do research into Digital Preservation
- Dare partners follow KB guidelines in this respect
- KB will warn DARE-partners if file format gets obsolete and require preservation action
- KB will support DARE partners to keep their material accessible (knowledge exchange)

Current situation & New developments

- 100.000 DARE scientific publications in e-Depot
- New developments:
 - Persistent Identifier (DOI, CERN Handler?)
 - Link with National Union Catalogue
 - Connection between publication and related data
 - Extra file format checks
 - Normalisation of file formats (Word etc.)

More information

- KB and Digital Preservation:
<http://www.kb.nl/dnp/e-depot/e-depot-en.html>
- DARE and DARE-net
www.darenet.nl/



Long term preservation and IR's

Thank you for your attention!

Barbara.Sierman@kb.nl

Koninklijke Bibliotheek – National Library of the Netherlands