The e-Depot in practice

Barbara Sierman
Digital Preservation Officer
Madrid, 16-03-2006
e-Depot in practice

- Short introduction of the e-Depot
- 4 Cases with different aspects
  - Characteristics of the supplier
  - Specialities, problems
  - Actions taken
The e-Depot

- IBM made the system (DIAS)
- Some special functions for the KB
  - Post-Office
  - Preprocess
  - Link with KB catalogue

- DIAS + special functions + organisation = e-Depot
The e-Depot in practice

**OAIS model**

- Preservation Planning
- Data Management
- Archival Storage
- Administration
- Ingest
- Access

**Flow Diagram:**
- **PRODUCER** (SIP) → **AIP** → **DIP** → **CONSUMER**
First: Preparation

- Agreements
- Details later
  - What will be delivered?
  - Formats used?
  - Metadata delivered?
  - Bibliographic information?
  - Way of delivering? Tape, cd, ftp?
    - Set of tests
The e-Depot in practice

The flow of the object

Post-office ➔ Pre-Process ➔ SIP Creation ➔ AIP Storage ➔ DIP Creation ➔ DIAS ➔ DIP Request & Representation

Preservation Manager ➔ KB Catalogue
The e-Depot in practice

Access to e-Depot

USER

CATALOGUE with URL

What do you want to access?

Restricted

Where are you?

Remote

Who are you?

Open Access

On-site

Trusted

Unknown

"Please visit the KB in The Hague"

Implicit question

Explicit question

Identification

NBNs

IP check

LDAP
Who is sending material to the e-Depot?

- International publishers  Case 1
- Projects
  - Academic publications (DARE)  Case 2
  - TIFF Masters  Case 3
- Delivery related to deposit role
  - Electronic books
  - CD-ROMS  Case 4
Case 1: International Publishers: characteristics

- Publisher chooses file format
- Delivery part of their business process
- Generally used formats like pdf
- Basic bibliographic metadata
- Large quantities of objects
- Steady quality
Case 1: Specialities with publishers

- Fileformat not always what they said
- Hidden surprises (DOI)
- Limited preservation metadata
- Restricted versus Open Access
Case 1: Actions taken

- File format check (JHOVE)
- Update bibliographic metadata with DOI (tbd)
- Research into Premis / Preservation Metadata
- Added functionality for different access methods
Case 2: Dare project

- 16 Dutch universities and scientific institutions and scientific output
  - Thesis
  - Articles
  - Reports
  - Publications, no raw data
- Network of repositories
- Work together but each partner will keep his own responsibility and management
- Project ends 2006
Case 2: DARE characteristics

- Permanent storage and access at the KB
  - Safe place

- Harvesting the material
  - From the university repositories
  - Range of file formats
  - Bibliographic description Dublin Core
  - Scientists added metadata
  - Titles already present in e-Depot
  - One team, 16 individual members
Case 2: DARE specialities

- Standard set in Dublin Core too limited
- Different file-formats, sometimes peculiar ones
- Different repository systems used
- No relation between KB and the individual scientists
- Threshold must be as low as possible for scientist to deliver material
- Limited (preservation) metadata
- Heterogeneous group
Case 2: DARE conclusions

- Check on file format important
- Quality of bibliographic information important for catalogue: checks
- Unique identifier needed to avoid duplicates in e-Depot (DOI)
- Developing more checks
- Copyright issues
Case 3: TIFF-pilot

- Reuse and permanent storage of tiff masters (digitized material)
- Involved: 5 cultural heritage institutions with different collections
- Feasibility study
- Business to business model
  - Secured web-interface
  - Institutions store their tiffs at KB
  - Ingest in e-Depot
  - Dissemination of tiffs via web-interface
Case 3: TIFF characteristics

- Every tiff has an analogue original
- Restricted access needed
- Business to business
- Only own publications: identifier for delivery
- Only own publications retrievable
- Access outside the KB
- Large files (infrastructure)
- Limited metadata
- NISO Z3987 more metadata needed
- Lack of knowledge or capacity for adding more metadata
Case 3: TIFF Solutions

- Guidelines for creating tiffs
- Metadata requirements
- Added functionality e-Depot
  - Search functionality authorised
  - Access functionality enlarged
- Implementation service end 2006
Case 4: Deposit of CD-ROMs

- 2003: special way of handling
- Interactive multimedia publications on CD-ROM
- Installing CD-ROM on standard PC on C-drive
- Testing
- If needed, install additional software, hardware
- Make disk-image of C-drive
- Testing
- Complete disk-image stored in e-Depot
- 600 done, 1100 backlog
- “valid” until 2010
CD ROM workflow

CD-ROM → Disk-image → DIAS

If needed: Extra soft-/hardware
Case 4: Problems

- CD-ROM dependent of standard PC
- Special software Powerquest to make image: not open source
- 2010 too far away: hardware obsolete earlier
- Time consuming: 1 hour per CD
- 2005: doubts about strategy
Case 4: Solutions

- Separation of publication and environment:
  - CD-ROM as .iso image
  - Base disk images of different system software (OS, drivers, plugins, additional applications)
  - Under discussion

- Rendering using:
  - Virtualisation (VMware)
  - Emulation (modular emulation)
Summary of problems

- Fileformat & versions
- Unknown problems with file formats
- Metadata
- Level of awareness and knowledge
- Access methods
- Copyright
Summary of solutions

- Permanent research at KB
  - In house knowledge
  - Metadata (Premis, Preservation Manager)
  - Fileformats: behavior, PRONOM, GDFR
- Guidelines for delivery
- File Format control (JHOVE)
- Added functionality DIAS
- (Inter-)national Collaboration
Thank you for your attention!
barbara.sierman@kb.nl
www.kb.nl