The KB experience

Emulation and Migration: A comparison in terms of costs

Erik Oltmans
Head Acquisitions & Processing Division
Koninklijke Bibliotheek

Nanda Kol
Delft University of Technology

DCC/DPC Seminar on Costs Models, BL, July 26, 2005
Overview

• The KB e-Depot
  • Organisation and Policy
  • Facts and Figures

• Preservation Strategies: Emulation and Migration

• Life Cycle Management

• Cost Issues

• Comparison, conclusions
Organisation and Policy
Mission of the e-Depot

Mission of the Koninklijke Bibliotheek:

- Ensuring permanent availability of information and knowledge

This yields:

- Long-term preservation and accessibility of electronic publications
- Safeguarding authenticity & integrity (once in the archive, never deleted or changed)
- Development of ever changing preservation and accessibility toolbox
- To become a dedicated institution, providing permanent access
Facts & Figures
What is the e-Depot?

- Electronic version traditional depository
- Developed in collaboration with IBM
- Technical heart: DIAS (OAIS-compliant)
- Integrated with other library modules
- Ingest of online journal articles, e-books, and CD-roms (installables)
- Operational since March 17, 2003
Some numbers

• Current holdings 2005:
  • 4 Terabyte
  • 2,600 on line journals
  • 3,800,000 electronic publications

• Processing capacity depends on the input: on average 5,000 to even 65,000 publications per day
Customers of the KB

- Individual end users
  - Have access to the e-Depot on site (within the library’s premises)
  - Pass holders or walk-in users
  - Also remote access if allowed by publishers

- Publishers
  - Archiving agreements determine conditions, terms, and duties

- The use of the documents is allowed under restrictions
- Retrieval, access, printing, downloading for private use only
- Systematic reproduction is not allowed
Digital Preservation

- Digital objects become inaccessible
  - File format obsolescence
  - Software obsolescence
  - Hardware obsolescence

Digital Preservation Strategies:

- Do nothing
- Migration
- Emulation
- Combination of both
Migration

• Keep the environment, change the object
• Repetitively convert digital objects into new formats
• New formats will remain accessible for some time
• After 3-5 years: convert again
• Must apply to entire collection: cannot miss one single object
• Keep previous or original object?
Emulation

- Build ‘old’ software environment on new hardware
- Build ‘old’ rendering tools on new operating systems
- Every 3-5 years, emulation tools must be checked, and adapted
- Does not apply to collection: original objects remain untouched
- Only original objects need to be stored
**Associated Costs:**

- **Do nothing:** no costs (yet)
- **Migration:** many repeating costs
- **Emulation:** high initial costs
- **Combination of (3) and (2):** combination of high initial costs and many repeating costs
Life cycle Management e-objects:

• **Migration**: many repeating costs
  \[ K(t,a) = s + ing + h(t,a) + m(t,a) \]

• **Emulation**: high initial costs
  \[ K(t,a) = s + ing + h(t,a) + E + e(t) \]

Where \( K(t,a) \) is the total cost of holding \( a \) objects for a period of \( t \) years

Where \( s = \) selection, \( ing = \) ingest, \( h = \) storage, \( m = \) migration,

\( E = \) setting up initial emulation tool, \( e(t) = \) emulation over time

Formula adapted from Shenton (2003)
Assessing assumptions in Excel:

- Demonstration
Developments in formula

- Drawbacks of formula as presented:
  - Size of archive remains the same (not realistic)
  - Only linear relation $\rightarrow$ Curve would be expected
  - Economies of scale
  - Inclusion of additional storage costs when applying migration
- The KB experience:
  - Real-world storage figures rather than educated guesses
KB experience

- Storage system supplied by IBM
- Initial Investment: 1.1 million Euro
- Including 10 year technical support, 4 terabyte storage space
- Personnel: 3 fte digital collection management
- Not corresponding to size of archive, but to volume of ingest
- Emulation maintenance: 40,000 each year
Developments in formula

- **Migration**
  \[ K(t, a) = S + p(t) + h(t) + m(t/5, a) \]

- **Emulation**
  \[ K(t, a) = S + p(t) + h(t) + E + e(t/5) \]

Where \( K(t,a) \) is the total cost of holding \( a \) objects for a period of \( t \) years.

Where  
- \( S = \) Initial storage costs, \( p = \) personnel, \( h = \) storage, 
- \( m = \) migration, \( E = \) setting up initial emulation tool, 
- \( e(t) = \) emulation over time
The KB e-Depot archive grows from 500,000 objects to 17 million objects in 35 years.

Archive grows from 500,000 objects to 17 million objects in 35 years

Migration

Emulation
Comparison and conclusions:

- Migration applies to the entire collection, emulation does not.

- Therefore:
  - Migration is cost-effective for relatively small collections.
  - For migrating small collections, the period of time does not affect costs.
  - However, emulation is more cost-effective for large collections.
  - This is not affected by the period of time either.

- Small archive: migration
- Big archive: emulation
Questions...?

erik.oltmans@kb.nl

http://www.kb.nl/e-depot