Bits and Bobs: Digital Preservation and Costs

Digital Preservation Coalition:
Forum on preservation of e-learning materials and cost models for digital preservation
London, 15th October 2002

Alison Macdonald
Secure Sciences Limited
alison.macdonald@securesciences.co.uk
1986 schilling & computer

Now has to be taken to bank for migration (to €1.453457 less charges ... cost to you of time, to bank of processing)

All obsolete technologies: 12” Laserdisk, Philips player BBC Acorn PC, software in BPCL

Original book (Domesday) made in 1086, still in use
Summary

1. Some words about costs
2. Identifying digital preservation costs today
   - Cost drivers and variables
   - Whose cheque book?
3. Some conclusions
What is cost?

- A pressure
- It leads to interference from others
- They then impose (further) limitations

- Costs pay for inputs which produce output
- Cost usually has multiple elements
- Which are subject to multiple factors
- Cost has multiple contexts
The context of costs

• Some words around costs:
The context of costs

Some words around costs:
- Price
- Payback
- Proportion
- Financing
- Unit of measure
- Charge-back
- Investment
- Return
- Relative cost
- Replacement cost
Some jargon related to costs

- Budget
- Standard costing
- Target costing
- Value analysis
- Value engineering
- Unit of measure
- Opportunity cost
Part 2

Identifying digital preservation costs today

Do we know what digital preservation involves?
Does the finance department have a cost centre for it? Should it?
Digital preservation: OAIS model
Digital preservation a sub-activity

• Digital preservation implies purpose behind retention
• Digital preservation also implies awareness
• Economies of scale imply one or more real or virtual repositories
  = archive(s)
Digital preservation is complex

Need to preserve:
  – Accessibility
  – Retrievability
  – Interpretability
  – Meaningfulness
  – Integrity
  – Authenticity
  – Security

Overcoming:
  • Technology churn
    – Hardware, software, media
  • Knowledge, semantic loss
  • Infrastructure loss (incl. funding continuity)
  • New problem areas
Metadata = the engine of preservation

OAIS model – from CCDSD 2001

Descriptive info. (metadata)

Preservation Planning

Data Management

Ingest

Archival Storage

Access

Administration

SIP (Submission Information Packages)

Management

AIP (Archival Information Packages)

DIP (Dissemination Information Packages)
Digital preservation: cost elements (1)

- Sub-activity of archive (in this example)
  - Set-up & infrastructure
  - Running costs: overheads, labour, materials

- Preservation actions as required
  - Materials and staff as required
  - Lots of knowledge required – almost certainly more than available in-house
  - Timing of these actions dependent on external factors
Digital preservation: cost elements (2)

• Digital preservation cost components:
  – Media refresh, software elements
  – Labour (internal, external)
  – Knowledge bank (internal, external?)

• Archive cost components:
  – Infrastructure
  – Hardware, media
  – Software
  – Telecommunications
  – Staff (archivists, IT, management)

= Uneven costs:
Annual archive cost **ex digital preservation** and average cumulative cost per 100,000 records.
Variables - what to preserve in a record?

• The original functionality?
  – To what extent?
    • Accuracy of numerical results
    • Links – original targets
    • Environmental dependencies
    • Etc.

• The original look and feel?
  – To what extent?
    • Colour fidelity?
    • Screen resolution
    • Screen layout
    • Fonts
    • etc

• Just the file / bit sequence?
• Audit trail, annotations?
Digital preservation: variables (2)

- Preservation actions as required
  - But which method?
    * Migration
    * Emulation
    * Archaeology
    * Museum
    * Other (eg Rosetta Stones)
  - But each of these have quite different cost characteristics,
  - And different implications about management

- Choice affects the value of the record
Digital preservation: cost drivers (1)

- Cost of materials (hardware, software, media)
- Hardware technology changes
- Availability of expert staff
Digital preservation: cost drivers (2)

• Heterogeneity of data types (today and tomorrow)
• Specialization of data
• Complexity of record structures
• Number of user communities
• Technology churn – software and communications
• Quality of the original data
• Indexing:
  – Ease of metadata creation
  – Depth and breadth of metadata
• Volumes (types, accumulation, annotations, audit trail)
• Frequency & regulation of access (ie, near-line, off-line)
OAIS: metadata the engine – but how many, and where do the costs fall?

- Descriptive info. (metadata)
- Preservation Planning
- Data Management
- Archival Storage
- Access
- Administration
- Ingest
- SIP (Submission Information Packages)
- DIP (Dissemination Information Packages)
- AIP (Archival Information Packages)
Where do the costs fall?

- **Who bears (or shares) which costs when?**
  - Original funders
  - Data originators
  - Curator(s)
  - Other business units (e.g., IT departments)
  - Data owners
  - Current users, future users
  - Future funders

- **What accounting problems are involved?**
  - Different archive models imply different problems
Hard costs, soft benefits (mostly):  
HYPOTHETICAL MAP

- Decreased network back-up costs
- Higher Tb per administrator for network
- Reduced h/w costs for network

- Improved DR capabilities
- Reduced “discovery” costs in litigation
- Ease of dissemination
- Patent protection

- Regulatory compliance
- Space saving
- Digital search capabilities
- Re-use of digital record
- Enhancements of “record” + annotations

Conservative
- Strong observed case

Nominal
- Reasonable case

Aggressive
- Protection
The other side of the coin: securing the budget

- Justifying the costs in terms of the benefits
- These are likely to be indirect
- Are the benefits measurable?
- Some archives may have income streams
- Balancing costs and benefits
- The risk of waste (several kinds)
- Staying within budget
Cost words again

• Reviewing those words around costs:
  – Price
  – Payback
  – Proportion
  – Financing
  – Unit of measure
  – Charge-back
  – Investment
  – Return
  – Relative cost
  – Replacement cost
Concluding questions and answers

- Is cost a risk in digital preservation?
- Is it possible to budget for digital preservation?
- Which cost elements can we predict with any degree of accuracy?
- Should the finance department have a cost centre for digital preservation?
- What are the implications of the cost choices identifiable today?

An understanding of costs in digital preservation is an important factor in maintaining maximum funding for digital archives.