the national archives

www.nationalarchives.gov.uk
Cost Modelling
The TNA experience

Adrian Brown
Head of Digital Preservation
The National Archives
Overview

• Current cost modelling
  • Ingest
  • Data and storage management
  • Access
• Future cost profiling
  • Preservation
• Conclusions
Current approaches

• Cost modelling drivers
  • Budget planning
  • Accountability and performance targets
• Current operational functions
Ingest → Data Management → Archival Storage
Archival Storage → Access

Characterisation

Preservation Planning → Preservation Intervention
Ingest

• Cost elements
  • Selection and evaluation
  • Transfer
  • Pre-accession processing
  • Cataloguing
  • Loading
Ingest

\[
\frac{(60\% \times DPD) + (5\% \times RMD) + (60\% \times NDAD)}{FN}
\]

- DPD = annual operational cost of DPD
- RMD = annual operational cost of RMD
- NDAD = annual cost of NDAD contract
- FN = annual no. of files ingested
- **Current ingest cost = £18.76 per file**
Ingest

• Issues

  • Costs are primarily linked to the complexity of the transfer, not the volume
  • Current cataloguing costs are high due to manual processing
  • Improved standards for creation, management and transfer should reduce costs
Data and Storage Management

• Cost elements
  • Systems administration
  • Hardware costs
  • Software costs
  • Media management
Data and Storage Management

\[
\frac{(DA \times 0.06) + \left(\frac{DA}{5}\right) + (30\% \times DPD) + (30\% \times NDAD)}{VOL}
\]

- DA = capital cost of systems
- DPD = annual operational cost of DPD
- NDAD = annual cost of NDAD contract
- VOL = total volume (MB) of records stored
- Current cost = £3.34 per MB
- Current minimum cost = £0.06 per MB
Data and Storage Management

• Issues
  • Costs relate primarily to capacity rather than use
  • Future transfer volumes are difficult to predict
Access

Electronic Records Online
Pilot presentation of the Digital Archive

Welcome to Electronic Records Online (ERO)
The Digital Archive is The National Archives' secure repository for storing electronic government records. This new beginning repository was launched in April 2003, successfully storing these records for the future.

Electronic Records Online is a pilot website that allows you to view these new digital records online, independently of the secure source in the Digital Archive. We have migrated most of the digital records into an easy to view format to enable quick and easy access. You can also request a copy of the original file.

What do you think?
We would be pleased to hear your views about Electronic Records Online. Please click here if you would like to give your feedback.
Access

• Current delivery costs
  • Online delivery = 13p per item
  • Removable media = charged on cost-recovery basis
• Integrated eCommerce system should reduce costs further
Predicting future costs

• Cost modelling drivers
  • Preservation strategy
  • Research & development costs
  • Future operational functions
Characterisation
Characterisation

• Automated tools to characterise
  • Representation properties
  • Inherent properties
• Automated tools to validate preservation interventions through comparison of characterisations
• Principal costs are R&D
• Operational costs are negligible
Preservation Planning

### File Format Summary

**Details for: Broadcast WAVE**

<table>
<thead>
<tr>
<th>Details for: Broadcast WAVE</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Go to:</strong></td>
<td>Summary</td>
<td>Documentation</td>
</tr>
<tr>
<td><strong>Name</strong></td>
<td>Broadcast WAVE</td>
<td></td>
</tr>
<tr>
<td><strong>Version</strong></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Other names</strong></td>
<td>EWA/E (0)</td>
<td></td>
</tr>
<tr>
<td><strong>Identifiers</strong></td>
<td>PUD:WAVE:0</td>
<td></td>
</tr>
<tr>
<td><strong>Family</strong></td>
<td>WAVE</td>
<td></td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>Audio</td>
<td></td>
</tr>
<tr>
<td><strong>Disclosure</strong></td>
<td>Full</td>
<td></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Broadcast WAVE is a chunk-based audio format developed by the European Broadcasting Union, and based on the Microsoft WAVE format. Structurally, a Broadcast WAVE file comprises a RIFF header with a WAVE data type identifier, followed by a series of chunks. Every file must include a Broadcast Audio Extension chunk, containing metadata required for exchange of information between broadcasters, a Format chunk, which describes the format of the audio data, and a Data chunk, containing the audio data itself. Broadcast WAVE files which contain MPEG-encoded audio data must also include a Fact chunk, containing file-dependent information</td>
<td></td>
</tr>
</tbody>
</table>
Preservation Planning

- PRONOM content development
- Performing risk assessments
- Identifying and testing migration pathways
- Currently largely manual – 2/3 FTEs
- Elements will be automated in future
- Most significant cost element
- Greatest opportunity to reduce costs through collaboration
Preservation Intervention

The Baines pottery provides a useful start to the quantified sequences as it has a long series of phases spanning the second to earlier third century AD, and, especially, the comparatively little evidence of Flavian-Trajanic on the site, and, the material of this date, still in later contexts. Can be seen both in the low mortaria dated to first century in comparison, especially in the low proportions...
Preservation Intervention

• Automated tools to perform interventions identified in preservation planning (e.g. migration/emulation)

• Major costs lie in development (preservation planning) and validation (characterisation)

• Implementation costs will be low
Migration

- Major cost component is for development of migration pathway
- Development costs will recur over time
- Implementation costs are minimal due to automation
- Volume has little impact on costs
Emulation

- Major cost component is for development of emulator
- Development costs will recur over time
- Operational costs should be negligible
- Costs are independent of volume – more cost effective for higher volumes
Conclusions

• Highest costs are for
  • Ingest
  • Preservation planning

• Storage costs are highly sensitive to capacity planning

• Cost differential between migration and emulation strategies is unclear

• Greatest efficiencies can be achieved through collaborative R&D on characterisation and preservation planning
the national archives

www.nationalarchives.gov.uk