
the national archives
www.nationalarchives.gov.uk

the national archives

Cost Modelling

The TNA experience

Adrian Brown

Head of Digital Preservation

The National Archives

the national archives

Overview

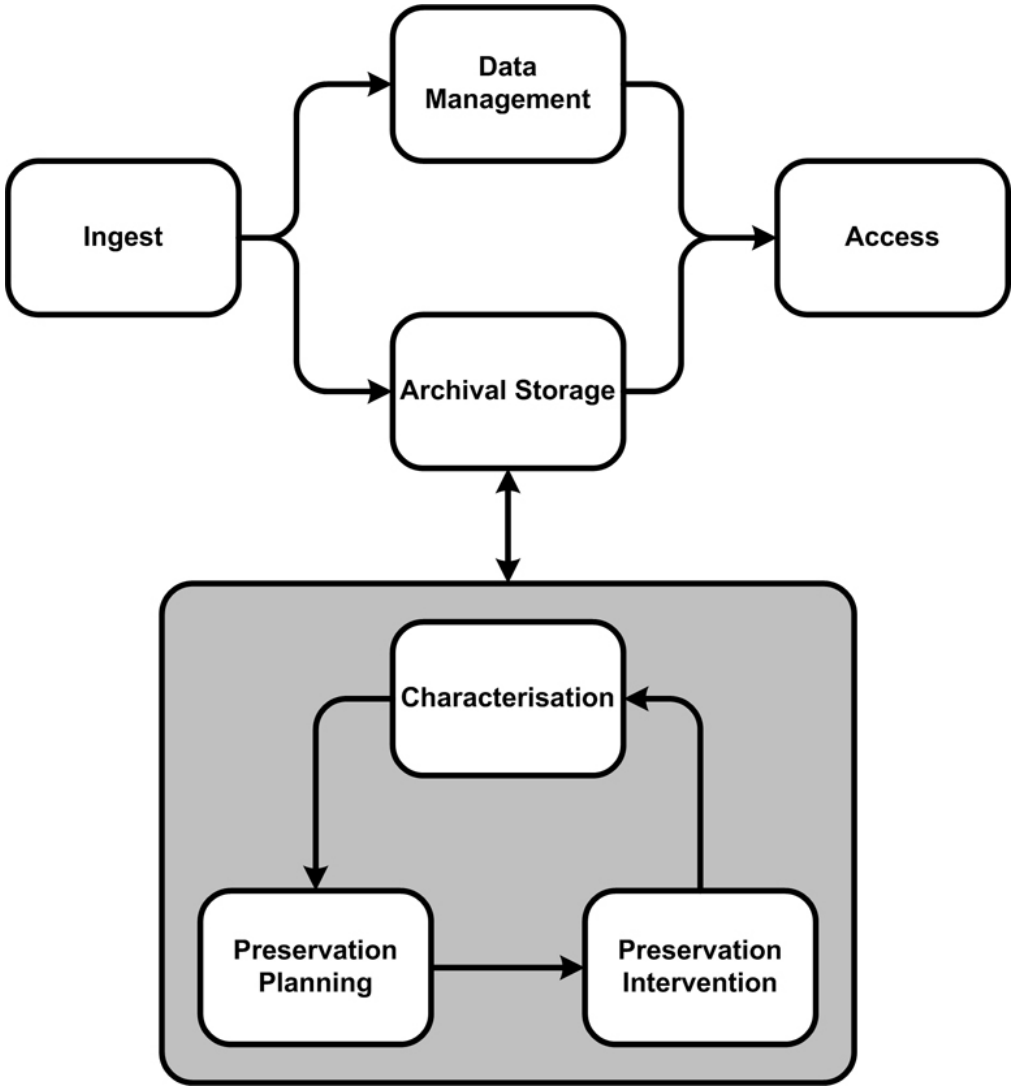
- Current cost modelling
  - Ingest
  - Data and storage management
  - Access
- Future cost profiling
  - Preservation
- Conclusions

the national archives

Current approaches

- Cost modelling drivers
  - Budget planning
  - Accountability and performance targets
  - Current operational functions

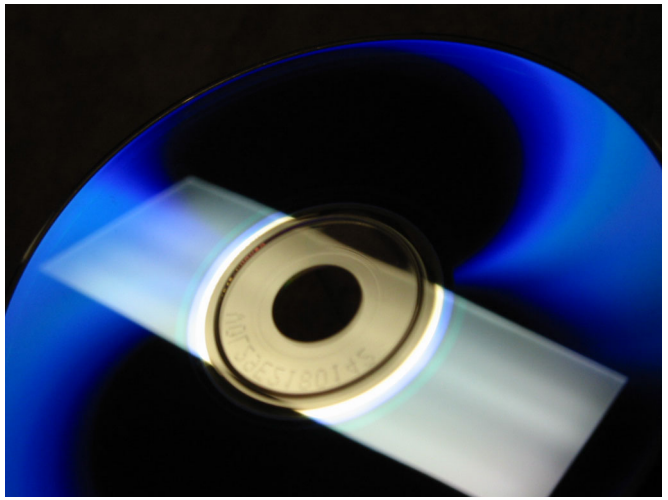
the national archives



the national archives

Ingest

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



the national archives

Ingest

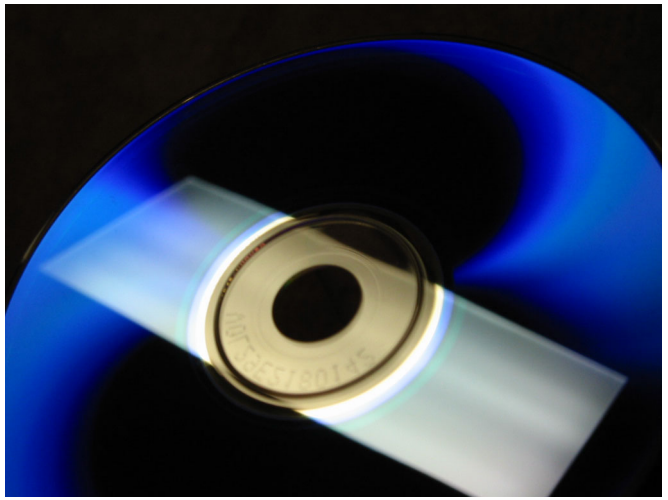
$$\frac{(60\%*DPD)+(5\%*RMD)+(60\%*NDAD)}{FN}$$

- 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**

the national archives

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



the national archives

Data and Storage Management

- Cost elements
  - Systems administration
  - Hardware costs
  - Software costs
  - Media management

the national archives

Data and Storage Management

$$\frac{(DA*0.06)+(\frac{DA}{5})+(30\%*DPD)+(30\%*NDAD)}{VOL}$$

- 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**

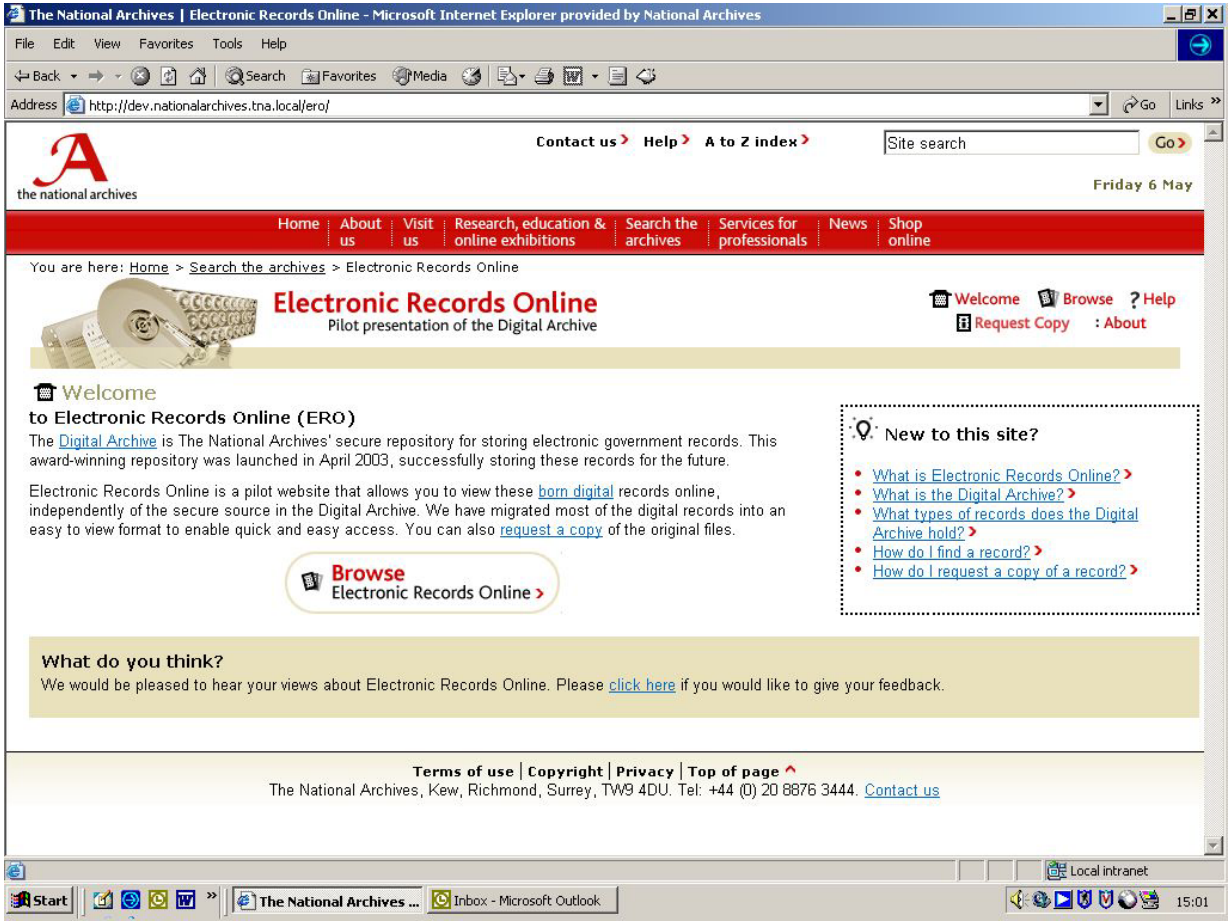
the national archives

Data and Storage Management

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

the national archives

Access



the national archives

Access

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

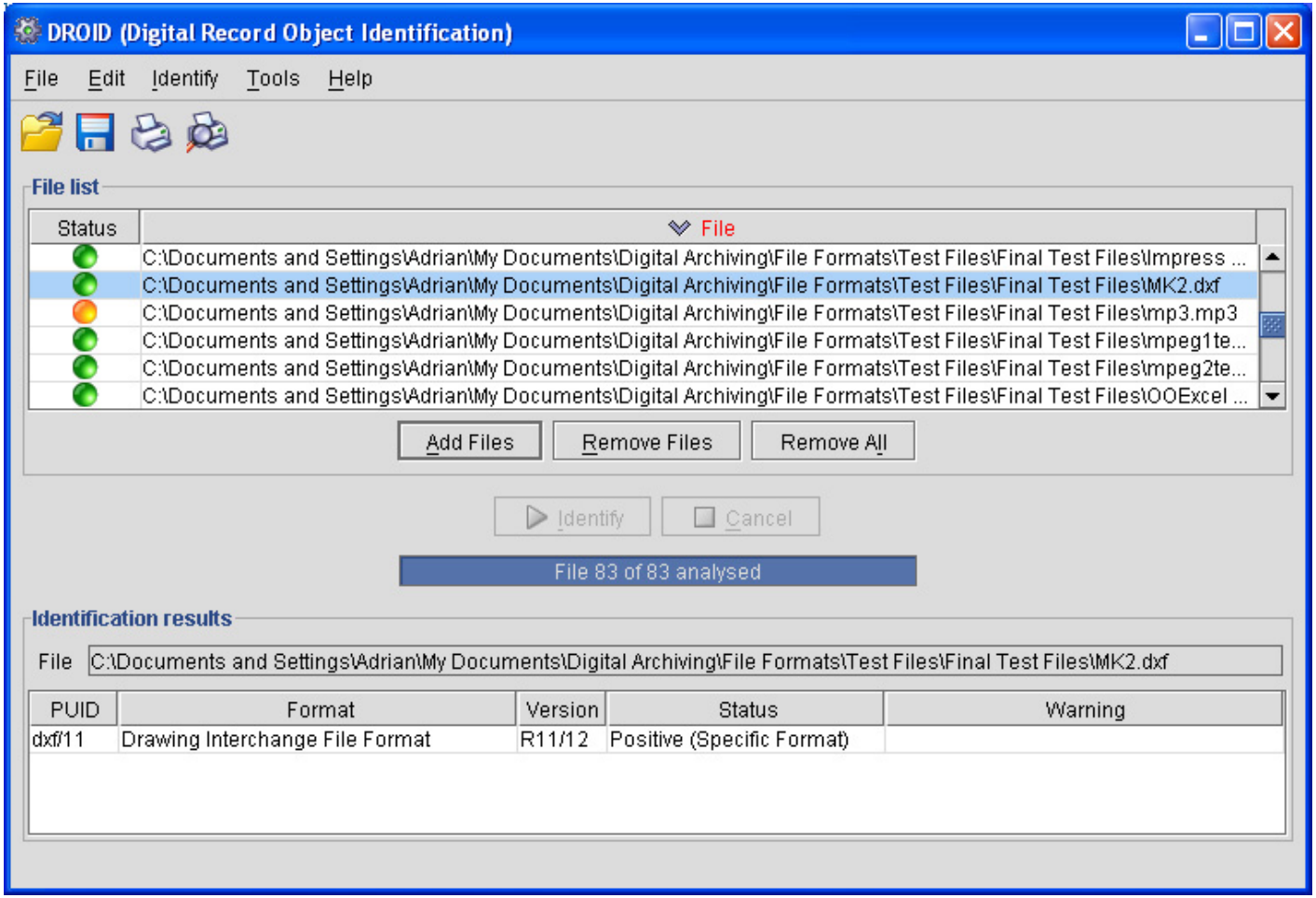
the national archives

Predicting future costs

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

the national archives

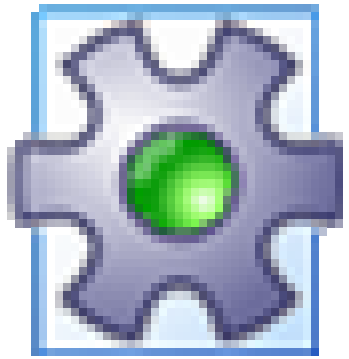
Characterisation



the national archives

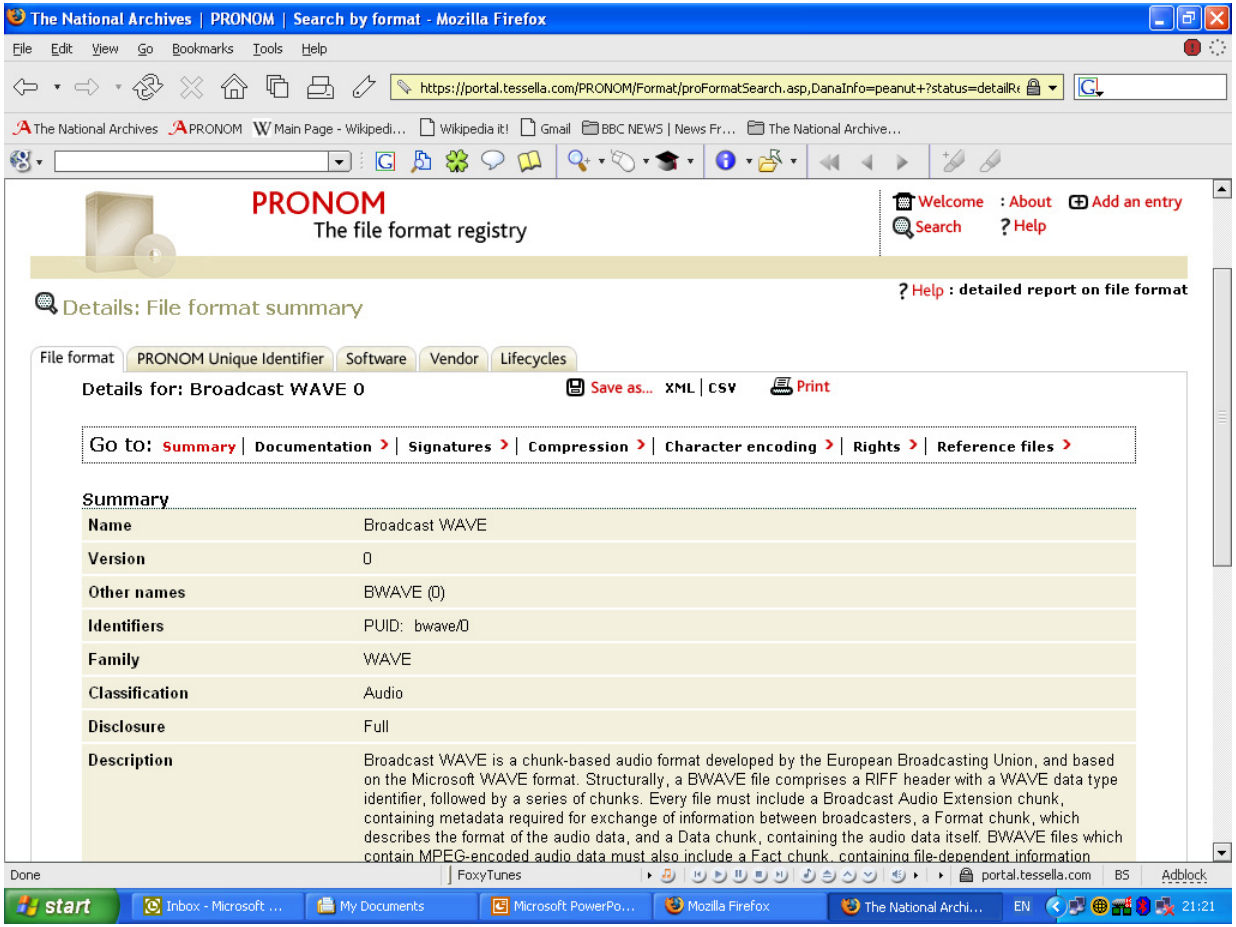
Characterisation

- 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



the national archives

Preservation Planning



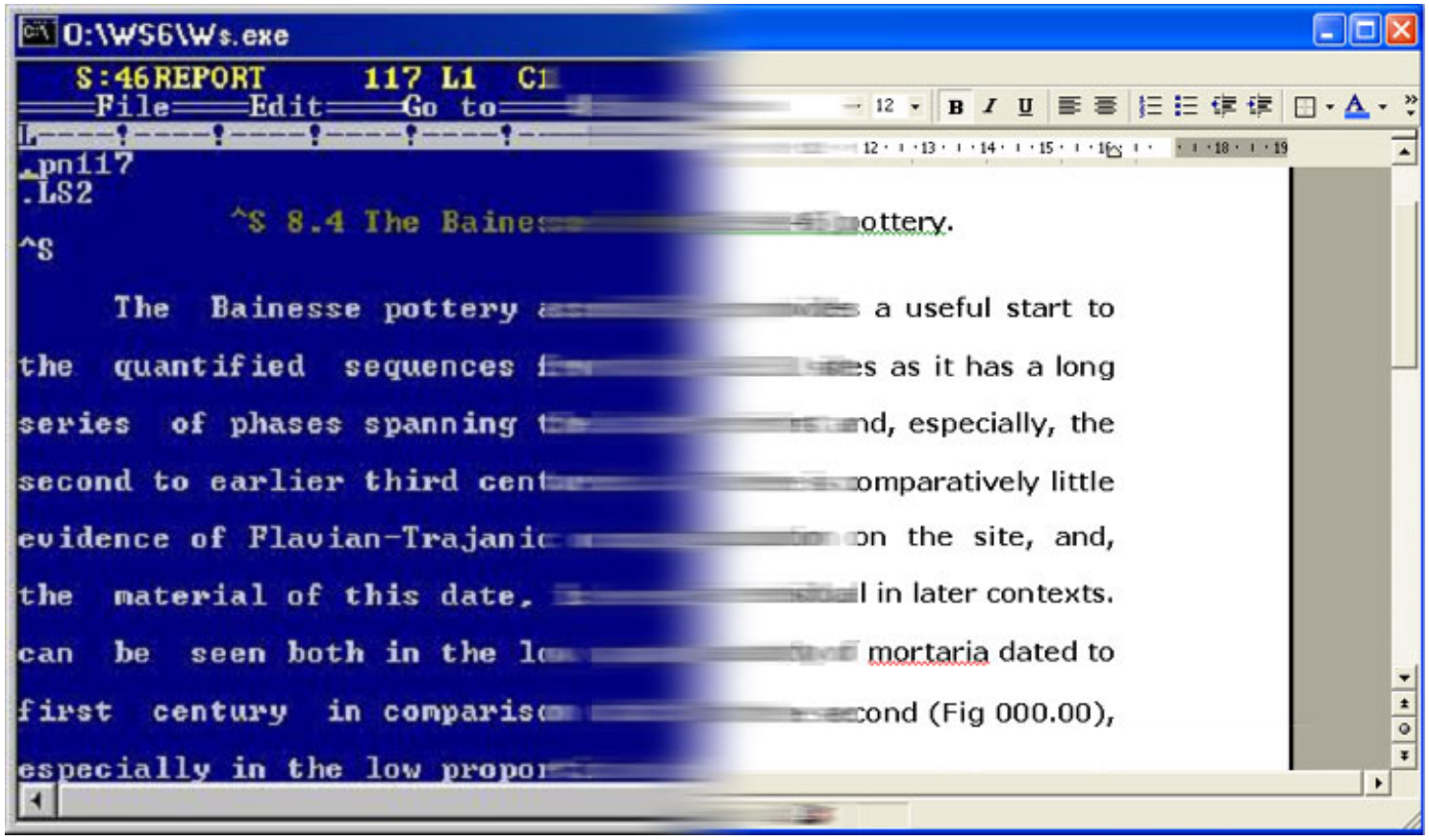
the national archives

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

the national archives

Preservation Intervention



the national archives

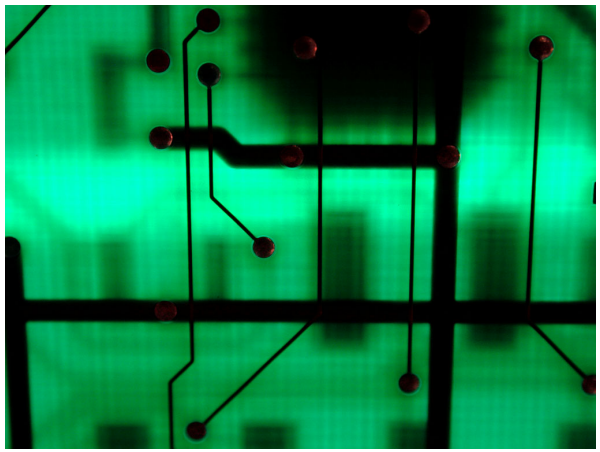
Preservation Intervention

- 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

the national archives

Migration

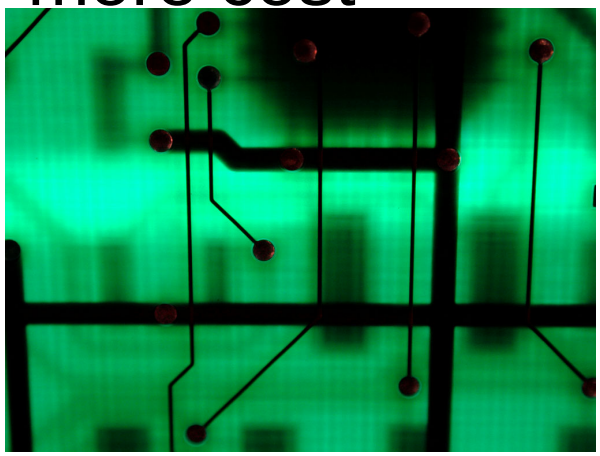
- Major cost component is for development of migration pathway
- Development costs will recur over time
- Implementation costs are minimal due to automation
- Volume has little impact on costs



the national archives

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



the national archives

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