



the national archives

www.nationalarchives.gov.uk

Cost Modelling

The TNA experience

Adrian Brown

Head of Digital Preservation

The National Archives

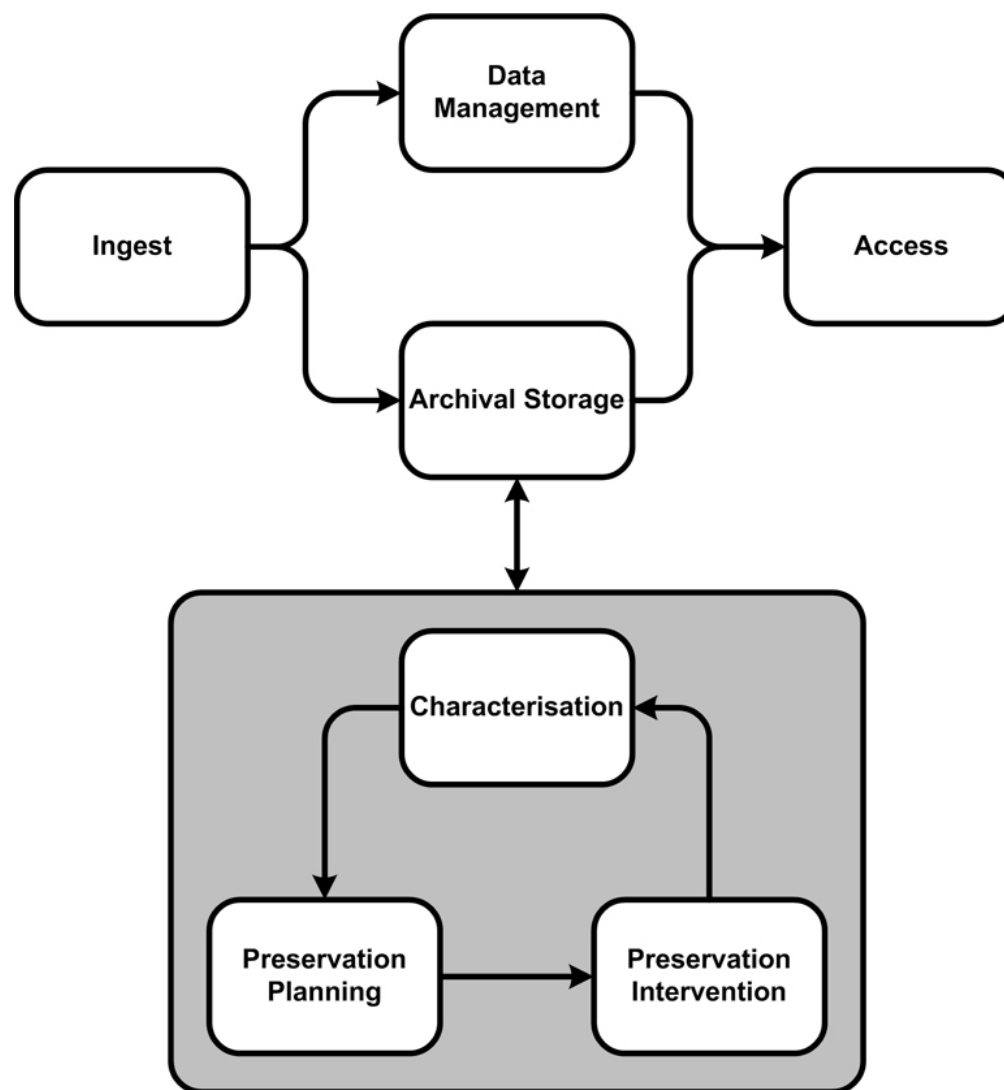
Overview

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

Current approaches

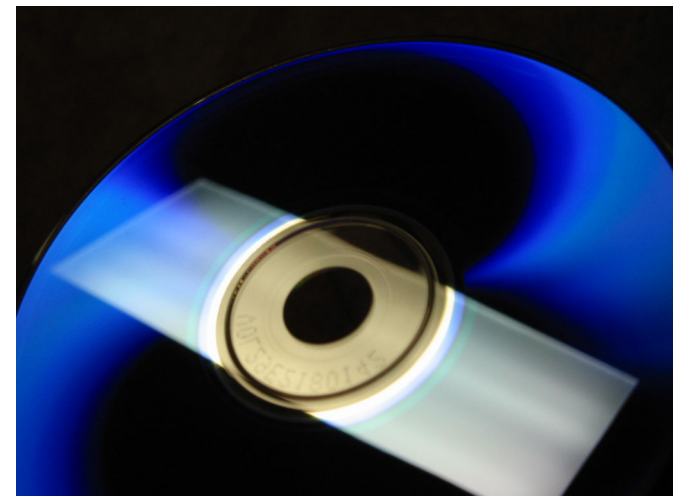
- Cost modelling drivers
 - Budget planning
 - Accountability and performance targets
 - Current operational functions





Ingest

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



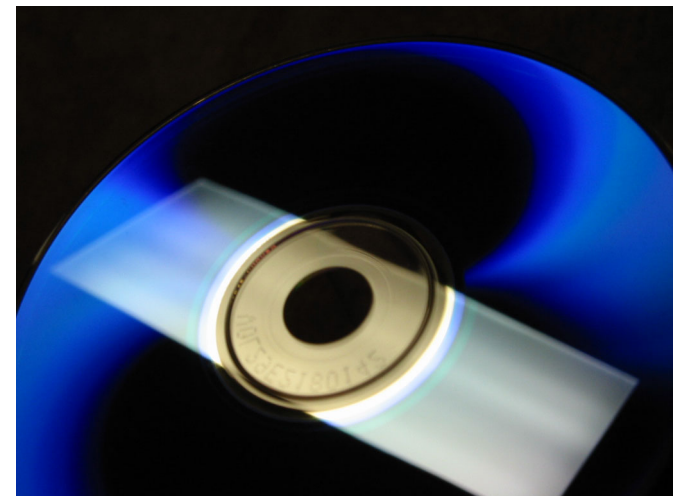
Ingest

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

- 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**

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

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

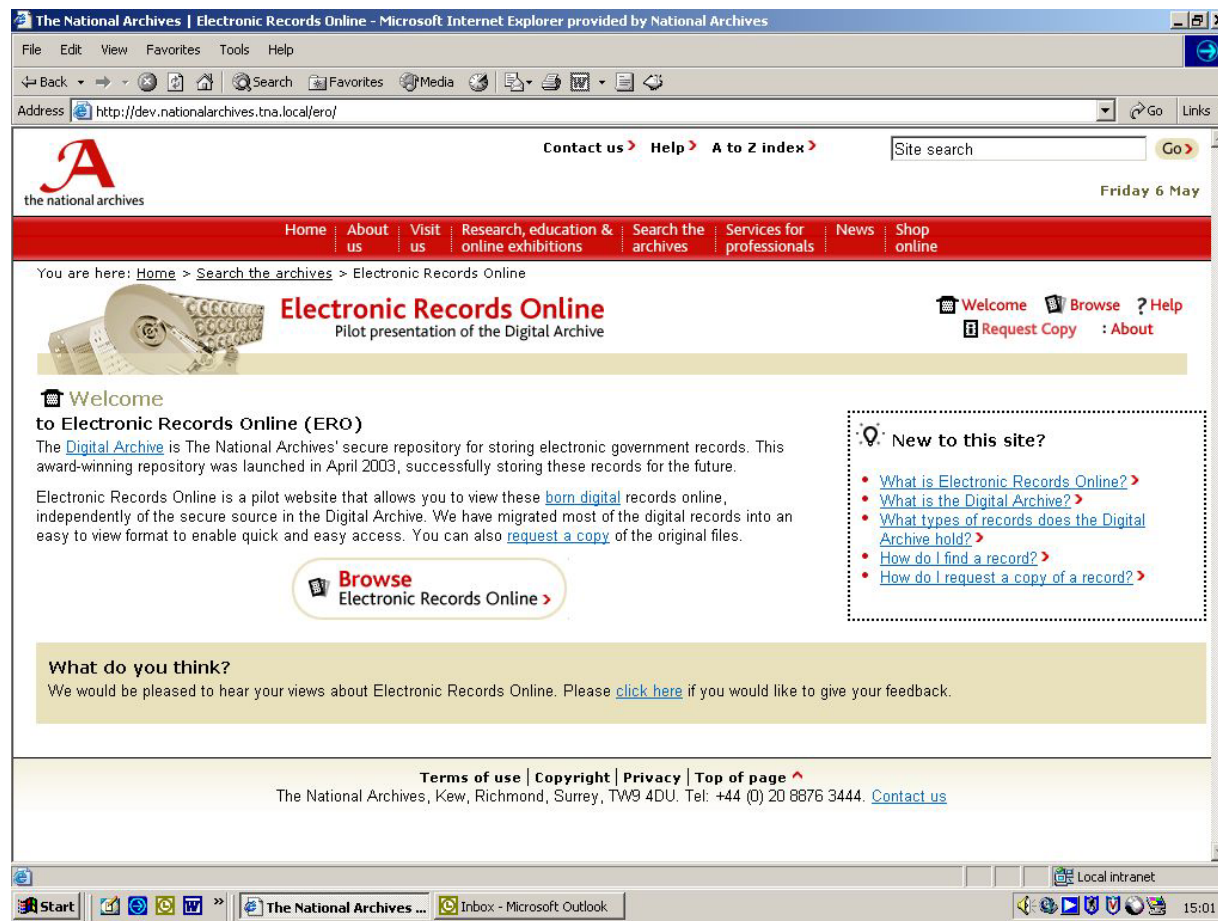
- DA = capital cost of systems
- DPD = annual operational cost of DPD
- NDAD = annual cost of NDAD contract
- VOL = total volume (MB) of records stored
- **Current cost = £3.34 per MB**
- **Current minimum cost = £0.06 per MB**

Data and Storage Management

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

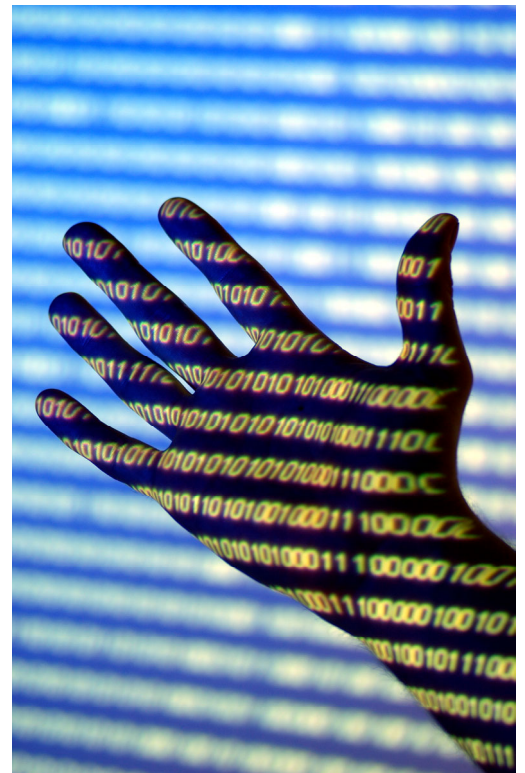


Access



Access

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

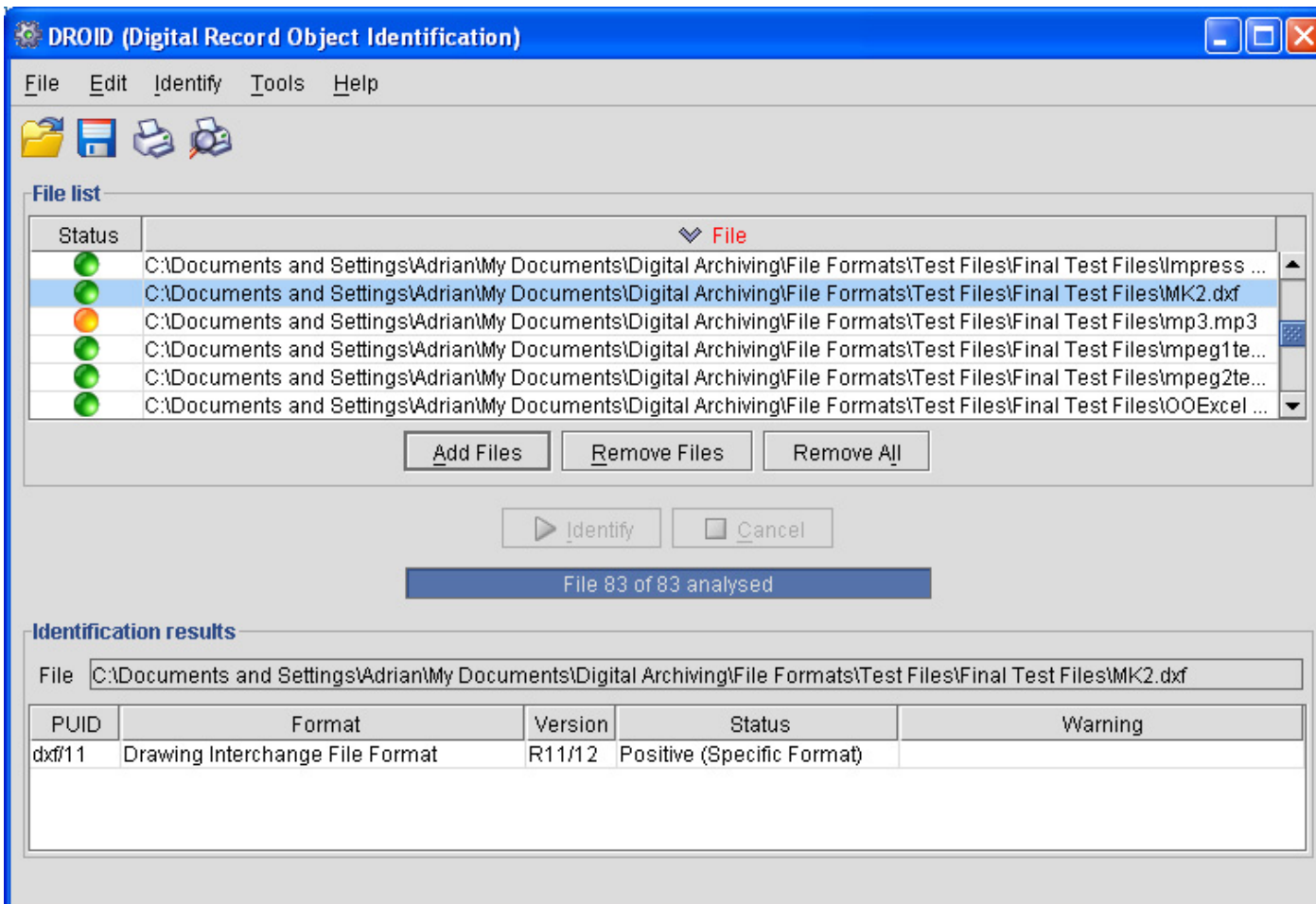


Predicting future costs

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



Characterisation



DROID (Digital Record Object Identification)

File Edit Identify Tools Help

File list

Status	File
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\Impress ...
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\MK2.dxf
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\mp3.mp3
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\mpeg1te...
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\mpeg2te...
	C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\OOExcel ...

Add Files Remove Files Remove All

Identify Cancel

File 83 of 83 analysed

Identification results

File C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\MK2.dxf

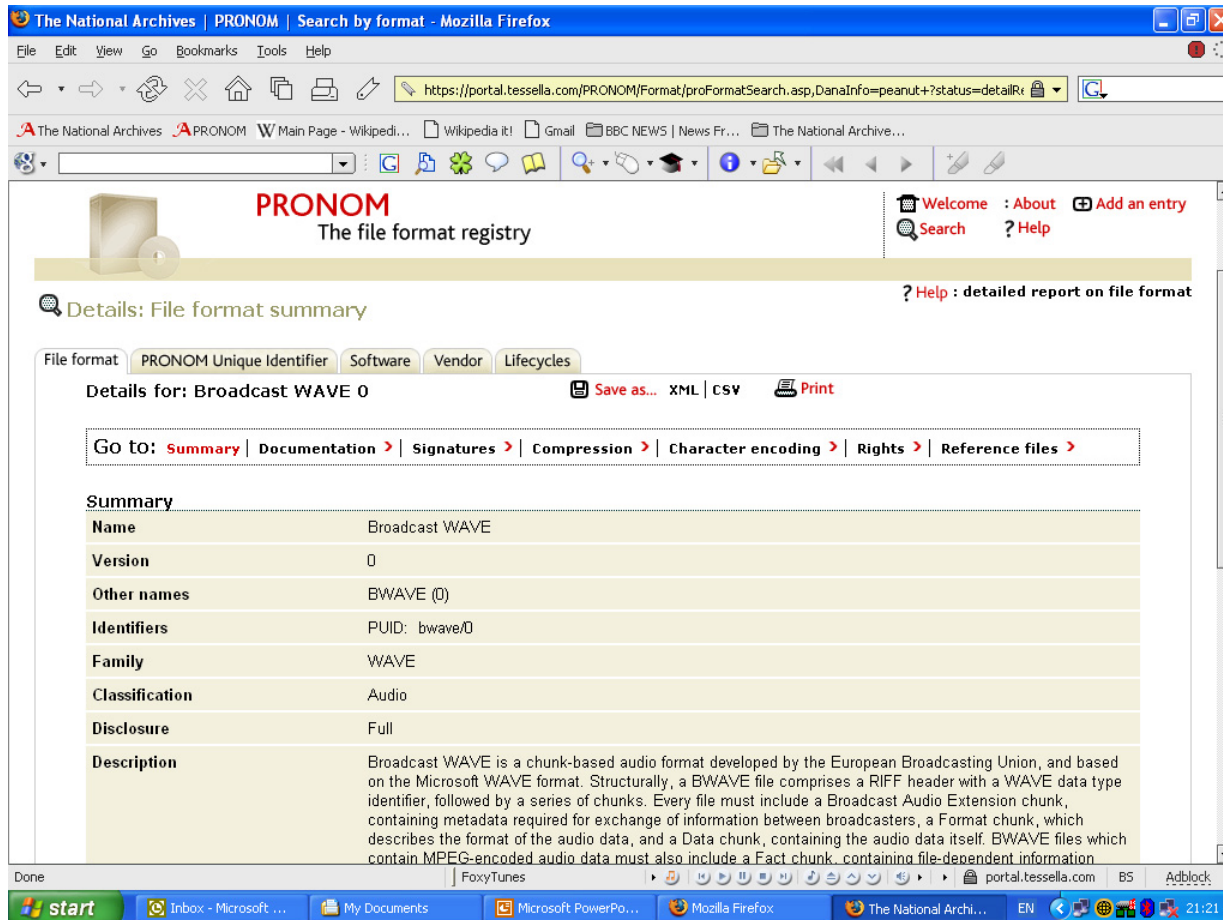
PUID	Format	Version	Status	Warning
dxfr11	Drawing Interchange File Format	R11/12	Positive (Specific Format)	

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



The screenshot shows a Mozilla Firefox browser window displaying the PRONOM website. The page title is "The National Archives | PRONOM | Search by format - Mozilla Firefox". The address bar shows the URL: <https://portal.tessella.com/PRONOM/Format/proFormatSearch.asp?DataInfo=peanut+?status=detailR>.

The PRONOM website header includes the logo "PRONOM The file format registry" and navigation links: Welcome, About, Add an entry, Search, and Help. A link for "? Help : detailed report on file format" is also present.

The main content area is titled "Details: File format summary". It features tabs for "File format", "PRONOM Unique Identifier", "Software", "Vendor", and "Lifecycles". The "File format" tab is selected, showing details for "Broadcast WAVE 0".

Below the tabs, there are links to "Save as...", "XML", "CSV", and "Print". A "Go to:" section includes links to "Summary", "Documentation", "Signatures", "Compression", "Character encoding", "Rights", and "Reference files".

The "Summary" section contains a table with the following information:

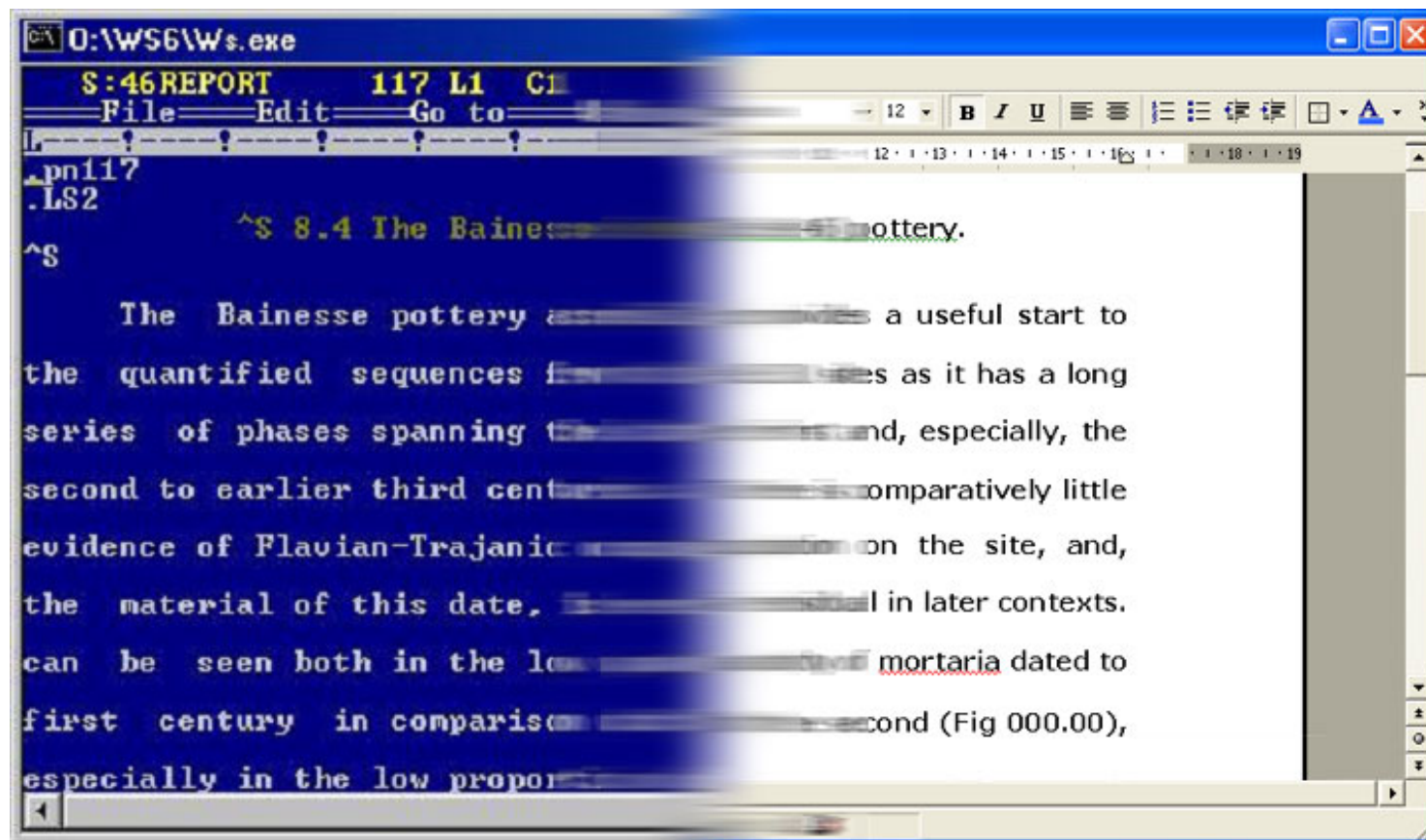
Name	Broadcast WAVE
Version	0
Other names	BWAVE (0)
Identifiers	PUID: bwave/0
Family	WAVE
Classification	Audio
Disclosure	Full
Description	Broadcast WAVE is a chunk-based audio format developed by the European Broadcasting Union, and based on the Microsoft WAVE format. Structurally, a BWAVE file comprises 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. BWAVE files which contain MPEG-encoded audio data must also include a Fact chunk, containing file-dependent information.

The browser's status bar at the bottom shows the taskbar with various applications open, including "start", "Inbox - Microsoft ...", "My Documents", "Microsoft PowerPo...", "Mozilla Firefox", "The National Archi...", and the system clock showing "21:21".

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

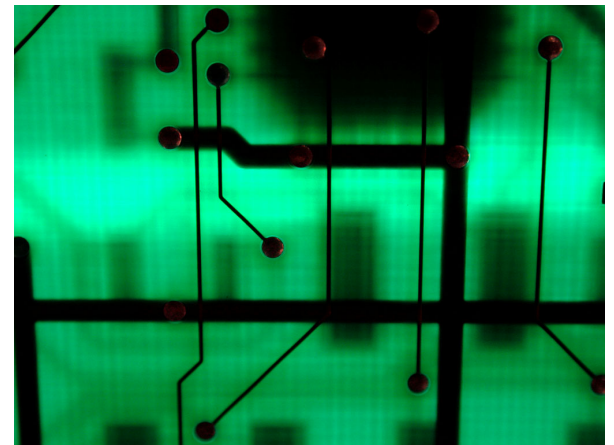


Preservation Intervention

- 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

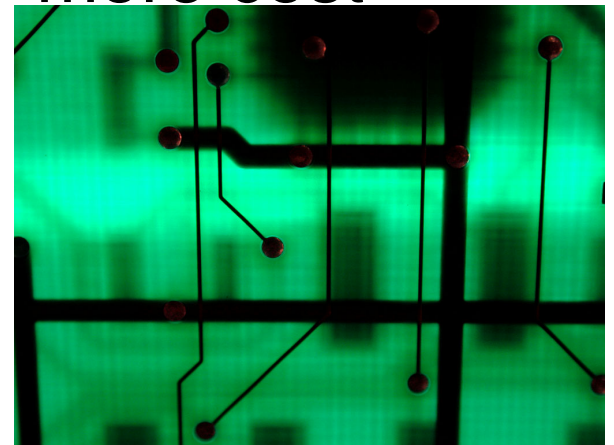
Migration

- 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