University of London Computer Centre

Some notes on the "Future-Proofing" project at the UoL

Ed Pinsent, Digital Archivist

University of London Computer Centre

What was it about?

- *Can we intervene in the life-cycle and make our fragile digital records safer?*
- Two-man project – Kit Good (records manager) and Ed Pinsent (digital archivist)
- We tested open source tools on common documents
- Wrote acceptance criteria for success
- We converted (normalised) test files
- We built basic archival information packages
- We collected technical metadata and significant properties
- We wanted to empower the records manager and meet preservation standards
- No support from IT required!

JISC FUNDED PROJECT: FUTURE PROOFING: ENABLING PRACTICAL PRESERVATION OF BORN-DIGITAL RECORDS

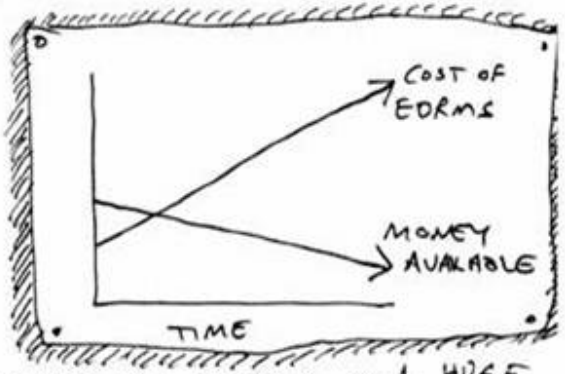
ED PINSENT
DIGITAL ARCHIVIST / PROJECT MANAGER

KIT GOOD
UNIVERSITY RECORDS MANAGER

(1)

MOST OF THE INFORMATION NOW CREATED IN THE UNIVERSITY IS 'BORN DIGITAL'. STAFF CREATE DOCUMENTS IN FILE FORMATS THAT HAVE MUCH SHORTER LIFE SPANS THAN PAPER. WE NEED TO MITIGATE THE RISK THAT THESE FILES BECOME UNREADABLE.

(2)



BUYING AN EDRMS IS A HUGE COST FOR AN ORGANISATION IN TERMS OF LICENCES, IMPLEMENTATION AND TRAINING. IS THERE A WAY THAT ELECTRONIC RECORDS CAN BE EFFECTIVELY MANAGED AND PRESERVED WITHOUT THIS COST?

(3)

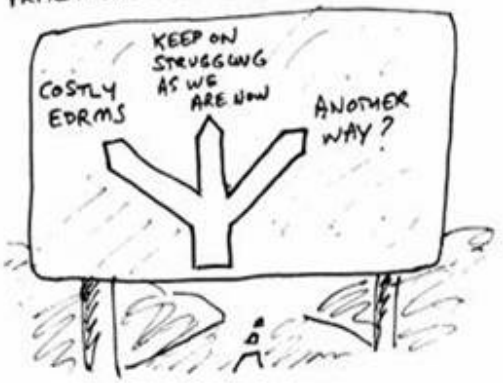
WE THOUGHT:
"CAN WE USE OUR EXISTING INFRASTRUCTURE, WITH SOME OPEN SOURCE TOOLS, TO BUILD A PRACTICAL, COST-EFFECTIVE SOLUTION TO LONG-TERM MANAGEMENT OF OUR KEY ELECTRONIC RECORDS?

(4)

THIS PROJECT WILL BUILD A TEST ENVIRONMENT FOR CONVERTING FILES INTO PRESERVABLE FORMATS. WILL THEY BE FIT-FOR-PURPOSE AS UNIVERSITY RECORDS?

(5)

THE AIM OF THE PROJECT IS TO IDENTIFY THE OPPORTUNITIES AND CHALLENGES TO THIS APPROACH TO ELECTRONIC RECORDS MANAGEMENT. IS IT A VIABLE ALTERNATIVE? IS IT PRACTICAL FOR RECORDS MANAGERS?



University of London Computer Centre

Five tools and what they do

Name	Function	Site
DROID (Digital Record Object Identification)	Automatically profiles file formats.	http://sourceforge.net/projects/droid/ and http://www.nationalarchives.gov.uk/information-management/our-services/dc-file-profiling-tool.htm
Xena (XML Electronic Normalising for Archives)	Converts digital objects into open formats for preservation.	http://sourceforge.net/projects/xena/ or http://xena.sourceforge.net/
JHOVE	Performs format-specific identification, validation, and characterization of digital objects.	http://sourceforge.net/projects/jhove/ and http://hul.harvard.edu/jhove/
NZ Metadata Extraction Tool	Extracts preservation metadata from file formats	http://meta-extractor.sourceforge.net/
The Digital Preservation Software Platform	Bundle of applications which support the goal of digital preservation.	http://sourceforge.net/projects/dpsp/

Acceptance criteria

Stage 4: Metadata extraction
Tool: DROID, NZ Metadata Extractor and JHOVE
Expected results: as per table below

Element	**Descriptive metadata**	**Technical metadata**	**Significant properties**
Definition	• original author • title • date of creation • date last modified • keywords • comments	• objectIdentifierValue • objectCategory • size • formatName • formatVersion • CreatingApplicationName • CreatingApplicationVersion • MessageDigestAlgorithm • MessageDigest	• PageCount • WordCount • CharacterCount • ParagraphCount • LineCount • TableCount • GraphicsCount • Language • Fonts • FontName • IsEmbedded • Features

Stage 5: Conversion
Tool: DPSP or XENA
Expected results: as per table below
Note that the aspects are not necessarily in increasing order of complexity or quality, in particular the functionality aspect may have little correlation with the other aspects.

Element	**Readability**	**Comprehensibility**	**Presentation**	**Functionality**	**Look and Feel**
Definition	Text is legible	Text with some markup	All markup and graphics display correctly	Links work	Appearance and quality identical to original

Benefits of the tools

Tool	Records Manager	Digital Archivist
DROID	Analyses a drive	File format IDs, technical metadata, checksum
Xena	Normalised records = more preservable digital objects	Basic AIP and some identification of file formats
JHOVE	No measurable benefit	Rich technical metadata for (some) objects
NZ MET	No measurable benefit	Less rich technical metadata
DPSP	Automated manifest (transfer list), assurances of quality and peer-checking in deposit process	Credible deposit workflow with automated preservation steps

University of London Computer Centre

Xena

- Produces Open Office equivalents for MS Office – signed off as authentic and preservable
- Open Office > PDF/A also possible
- Xena also works for images, audio and emails
- Process produces an AIP in XML (but not much metadata)
- Xena could have a role in RM processes

University of London Computer Centre

Metadata extraction tools

- DROID, JHOVE and NZMET yield technical metadata (with limits, and some conflicts)
- In our project, we failed to integrate metadata outputs with their objects
- DROID could have a place in RM scenarios – audit a record store, create a disposal log

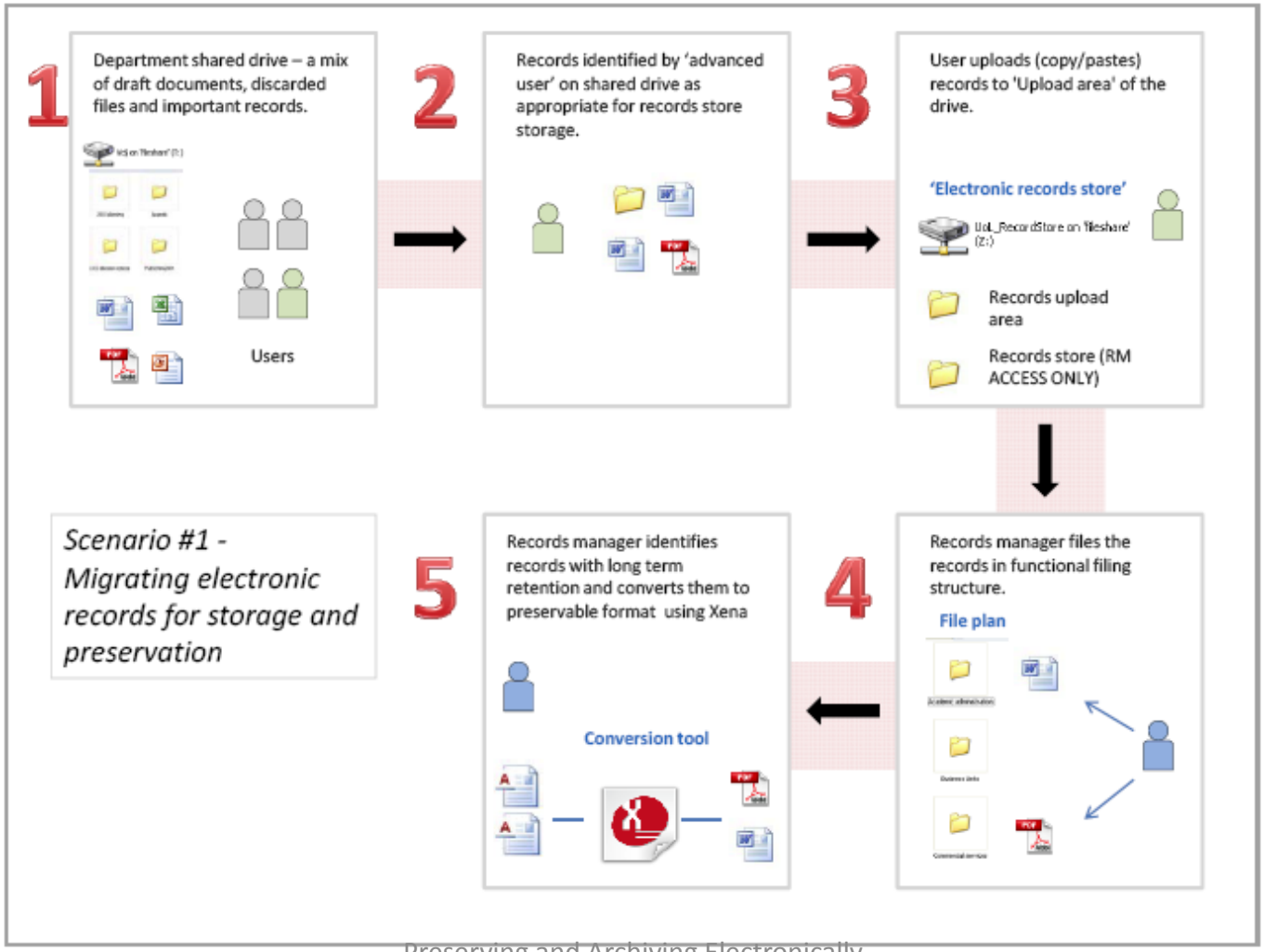
University of London Computer Centre

DPSP

- PRO
  - Automated process with checksums and virus checks built in
  - Creates audit trails
  - Strong on quality-assurance
  - Has its own database
- CON
  - Too many stages; quarantine, normalisation, repository ingest
  - Creates unique IDS and folders for all stages
  - Requires numerous logins (and logouts)

University of London Computer Centre

An RM scenario



University of London Computer Centre

Conclusions

- Open source tools + self-sufficiency is possible
- Xena + Open Office brings us closer to preservation
- Technical metadata gap in our method – where to store it?
- Highlighted the records management > digital archives gap
- Make early decisions about permanence

University of London Computer Centre

Links

- Future-Proofing blog: http://11kitbid.jiscinvolve.org/wp/
- Xena: http://xena.sourceforge.net/
- Chris Prom's assessment of Xena: http://e-records.chrisprom.com/review-of-xena-normalization-software/