



# EPIC: A Journey into Preservation Planning

Barbara Bültmann  
bb342@cam.ac.uk

Anna Collins  
ac281@cam.ac.uk

David Piper  
dp422@cam.ac.uk

Elin Stangeland  
es444@cam.ac.uk

Cambridge University Library

[www.lib.cam.ac.uk/preservation/epic](http://www.lib.cam.ac.uk/preservation/epic)

The JISC-funded *Evaluating Plato in Cambridge* (EPIC) project looked into using Plato, the preservation planning tool in the PLANETS suite, to plan the preservation activities of the institutional repository, DSpace@Cambridge.



## Background Research

Good sources of practical information on implementing preservation planning included:

- Previous JISC projects, eg the **KeepIt** Project.
- **InSPECT**, for guidance on significant characteristics.
- **Plato** and **Planets**, for documentation and case studies.

Many publications contain useful high level guidance to help tailor best practice to the institution.



The literature search is available at:  
[www.lib.cam.ac.uk/preservation/epic/litreview.html](http://www.lib.cam.ac.uk/preservation/epic/litreview.html)



## Needs of the Research Community

*"It is important to keep the existing structure, and that the table of contents and footnotes all correspond and remain coherent."*

*"The most important thing is to preserve the font."*

We asked 11 researchers from different disciplines who had contributed items to DSpace@Cambridge about **important properties of text-based documents**. All agreed that the document should remain "accessible and searchable, stable, uncorrupted, safe" and most importantly "**intelligible**". A document needs to retain its characters and their order, the document coherence, and to be human- and machine-readable.

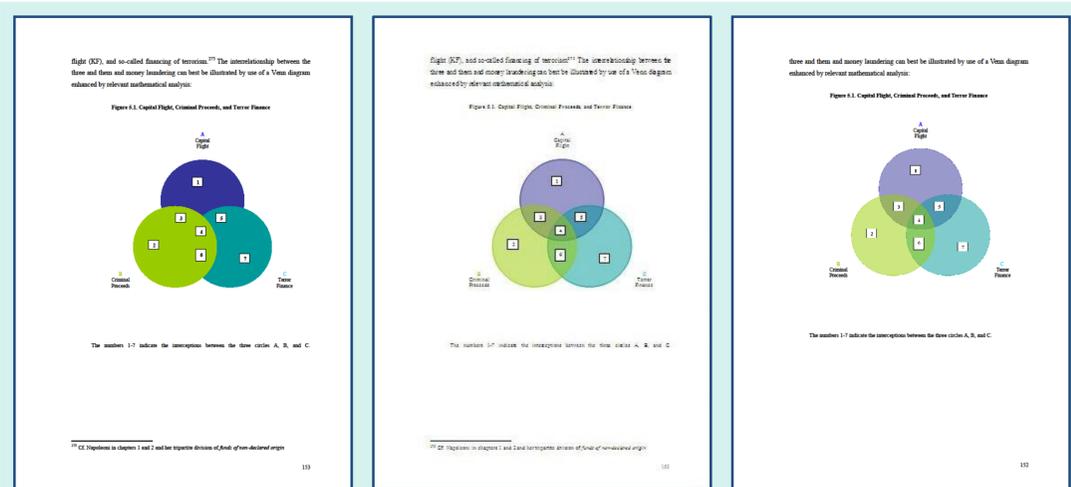
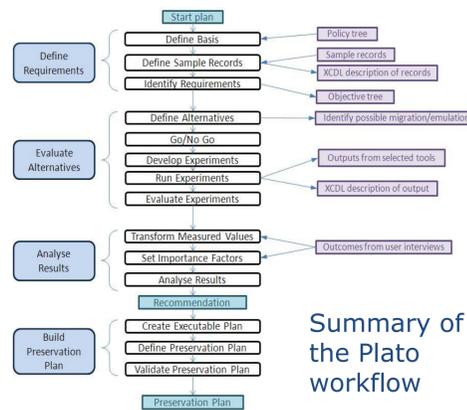
*"In terms of figures, colours are important because colour is used to label things... Resolution needs to stay the same because I probably have the figures as small as I could anyway."*

## Using Plato: Exploring Preservation Planning

We used Plato to look at the migration of PhD theses from MS Word to a suitable archival format. We used a sample of three theses and investigated three tools: **OpenOffice**, **MS Word 2010** and **Adobe Acrobat Pro**. The Adobe tool was discarded during assessment because we had difficulties in using it to produce PDF/A-1a files.

*"[Plato] makes you step through a formal process of defining requirements, examining alternatives, analysing results, and making recommendations. It asks about objects, technology, users and policies. ... Few have answers to all the questions."*

W. Kilbride, 2010  
<http://www.dpconline.org/newsroom/whats-new> (Issue 28)



Page of thesis, showing original (centre) with results of migration to PDF/A using Word (left) and OpenOffice (right).  
Madsen, F.G. 2009, 'International Monetary Flows of Non-Declared Origin', PhD Thesis, University of Cambridge.  
<http://www.dspace.cam.ac.uk/handle/1810/224484>



Issues were encountered with all migration tools tested including:

- **Font changes**, particularly in headers and footers.
- **Changes to figures**, eg colour, transparency.
- **Incorrect page breaks** leading to incorrect indexing.
- **Incorrect numbering** of footnotes.

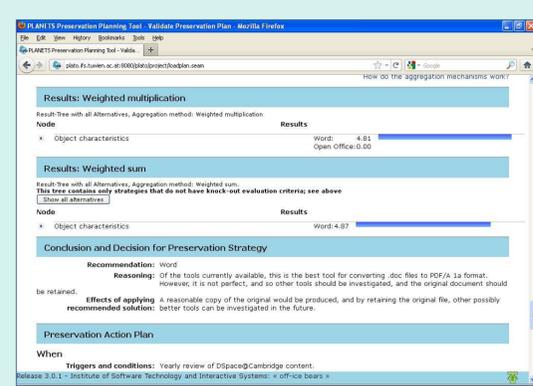
**None of the tools were considered suitable for non-urgent migration of Word files.**

**We found Plato to be useful for encouraging in-depth evaluation of preservation needs. However the process is detailed and preservation planning remained a time-consuming process; overall there was a lack of guidance to help ensure that the planning was appropriate.**

## Future Work

Following on from this project, future activities of DSpace@Cambridge will include:

- Improving awareness of the digital life cycle amongst depositors and the wider research community.
- Developing guidance resources to improve information management.
- Continuing development of preservation planning for different types of collection within the repository.
- Reviewing and updating automated ingest procedures.



Screen shot of the Plato web interface near the end of the planning process.

At the assessment stage OpenOffice scores 0 indicating it fails on at least one essential evaluation criterion. Of the tools under evaluation, Word produced the best result by our assessment criteria.