

PARADIGM

Academic Advisory Board Meeting 2

Monday, 12 December 2005
John Rylands University Library, Manchester

Final version
Susan Thomas
February 2006

Table of Contents

Participants	4
Using Fedora to explore digital private papers.....	5
Introduction.....	5
Hands-on session	5
Evaluation and discussion	7
Tour of the Labour History Archives Study Centre.....	14
Date of next meeting	14
<i>Appendix A: Programme</i>	15
<i>Appendix B: Academic Advisory Board – Terms of Reference</i>	16

Participants

Present were:

- **Dr Lawrence Goldman (Chair)**
Editor in Chief, Oxford Dictionary of National Biography
- **Simon Bailey**
Keeper of Oxford University Archives
- **Dr Stella Butler**
Head of Special Collections, John Rylands University Library of Manchester
- **Jenny Curtis**
Head of Digital Programmes, John Rylands University Library of Manchester
- **Professor Steven Fielding**
Professor of Contemporary Political History, University of Salford
- **Renhart Gittens**
Software Engineer, Oxford University Library Services
- **John Hodgson**
Keeper of Manuscripts and Archives, John Rylands University Library of Manchester
- **Helen Langley**
Curator of Modern Political Papers, Bodleian Library
- **Dr Alex May**
Research Editor, Oxford Dictionary of National Biography
- **Janette Martin**
Part-time Digital Archivist, John Rylands University Library of Manchester
- **Dr Kevin Morgan**
Department of Government, International Politics and Philosophy, University of Manchester
- **Michael Popham**
Head of the Oxford Digital Library and Project Director
- **Susan Thomas**
Project Manager and Digital Archivist, Bodleian Library

Apologies were received from Dr Martin Conway, Dr John Davis and Richard Ovenden.

Using Fedora to explore digital private papers

Introduction

Shortly before midday (after several transport hiccups) the Advisory Board were welcomed to the John Rylands University Library and offered coffee. Susan Thomas, Paradigm Project Manager and Digital Archivist, briefly introduced the aims of the day and outlined how the Paradigm team are testing Fedora software as a repository for digital private papers. The aims of the session were to evaluate the potential historical value of materials accessioned by the project; explore the basic functionality of Fedora; and to think about the functionality that researchers and curators might want in a digital repository.

Susan explained that the Fedora digital repository software is designed and maintained by the Universities of Virginia and Cornell. The software has a growing user-base in Europe, Australia and the United States. Users of Fedora in the UK include the National Library of Wales and the Wellcome Trust.

Fedora's developers intended the software to serve as a foundation for other services; that is to say, Fedora is not an 'off-the-shelf' product. The development priorities of the Fedora community concern the enhancement of the repository's underlying functionality rather than the development of friendly interfaces for end-users. These priorities are reflected in the default web interface (used for the hands-on session) which is really a tool for administrators who understand something of the workings of the repository, not researchers. It is expected that Fedora implementers will develop and adapt interfaces in-line with their own requirements.

Some of Fedora's users also develop and contribute new services, which can be used by other implementers of the software. Examples of these additional services include interfaces, which are more attractive and intuitive, such as Fez, Elated and Vital.

Hands-on Session

Members of the advisory board were asked to act as guinea pigs and test the default Fedora web interface by working through a specially created tutorial *Paradigm: Using Fedora to explore digital private papers*. The core Paradigm team Susan Thomas (Project Manager and Digital Archivist, Oxford), Renhart Gittens (Software engineer, Oxford) and Janette Martin (Digital Archivist,

Manchester) were on hand to answer questions and to assist with any computer hitches.

A sample of around 200 digital objects were available to explore, though only minimal descriptive information was available for each. This descriptive metadata comprised draft collection level descriptions of the archival collections (using EAD) and basic Dublin Core item-level records, including fields such as 'creator', 'title', 'date' and 'description'. Collections and objects were not indexed with personal, organisation, place or subject index terms. There were no multi-level catalogues and the digital objects had not been arranged to mimic original directory structure (i.e. the original order in which they existed on the creators computer).

On completion of the tutorial, Janette Martin gave a quick demonstration of an email inbox accessioned from a constituency office. Many of the emails were circulars from the Party's Resources Centre and not specific to the politician *per se*. However, it is useful that such material is being captured in personal digital collections as it is not, at present, being collected elsewhere.

Evaluation and discussion

The Advisory Board members were asked to complete an evaluation form, which aimed to cover two of the areas established in the Advisory Board's *Terms of Reference*:

1. To advise on the policies relating to the selection of materials, also giving an indication of the relative academic value of different materials.
2. To advise on the policies and applications relating to the usability and access regimes surrounding personal archival material in digital form both from the viewpoint of the researcher, and from the viewpoint of the academic tutor/supervisor.

This information will assist the Paradigm team with future development work. Whilst the sample is too small to offer a representational view, it does give us some valuable information

1. About the respondents

The project team received nine responses in total. Two respondents described themselves as 'Librarian', one as an 'Archivist', four as 'Researchers', one as 'Librarian' and 'Other – Management', and one as 'Archivist and Researcher'. Most respondents were familiar with the use of primary sources, including private papers, official papers and parliamentary debates, etc. The level of experience with digital materials was less consistent, though some of the researchers already used web-resources in their work.

2. Selection of materials and their relative historical value

We asked the Board to rate the value of different record-types on a scale of 1 to 5, where 1 represents 'Not valuable' and 5 represents 'Very valuable'. The records that scored highest were email and personal websites/weblogs. Only a few members rated presentations, spreadsheets and databases; there were only a few of these included in the sample and it is likely that they remained undiscovered by most. General comments about the exercise were that value depends on what interests you, and whether material is unique, or available elsewhere. One member wondered whether images and visuals were as important as 'types' of material, as arguably it is only necessary for a sample to survive; he also felt that this was not the case for records such as email, where it is much more important that specific correspondence survives.

Advisory Board ratings for common digital records

Where 1 is 'Not valuable' and 5 is 'Very valuable'

Images (3)

Of most value to librarians (5); of least value to researchers (2-3).

Email (4.5)

Scored highly with all groups.

Speeches (4)

Scored highly with all groups, though unpublished speeches or drafts of speeches are likely to be of most interest.

Press releases (3)

These generated a very mixed response.

Personal websites and weblogs (4)

Nearly everyone scored these at 4.

Campaign materials (3)

Nearly everyone scored these at 3.

Engagement diaries (3.5)

Engagement diaries generated a mixed response. It is interesting that the archivists and librarians scored these higher than most of the researchers, except one which allocated a 5.

Presentations (3.5)

Only six responses to this, which probably reflects the fact that there were few presentations in the sample and hardly anyone found them in their searches. Most who responded scored these at 4.

Spreadsheets and databases (4)

Only five responses for each record-type, again reflecting the small number in the sample. Most responded with a 4 or 5.

3. The relative historical value of the collections

Many commented that this was not easy to answer based on the sample and the time given to explore the materials. One collection was consistently perceived to be the most valuable; these were the reasons given: 'focuses on ideas', 'glimpses of high politics', 'contains material relating to the making of party policy on important public issues' and 'he is a major figure in his party'. One commented that the 'quantity and range of material' in another archive made it the most valuable, and another felt that two archives containing constituency-related material were useful for those interested in local activity and that public materials were important (e.g. websites and press releases) for those interested in how the parties want to be viewed.

Few ventured which collection they thought to be the weakest, but comments received indicate that range and depth of material is important, as is the 'uniqueness' of the material. One record-type considered to lack depth by some was the press release.

4. What characteristics should we preserve in order to retain the value of digital archival materials?

We asked this question, and the following question, because digital preservation research has not adequately explored what researchers will expect of digital archives. Do researchers want archives with strong evidential qualities, or is maintaining access to the information contained in the original records sufficient? If evidential qualities are important, what will satisfy researchers that a record can be taken as evidence? We may not be able to satisfy all the requirements articulated by researchers. Indeed, it is likely that many users have yet to identify what their requirements are. Beginning to think about these issues with researchers will help digital archivists to decide which characteristics of digital archives are essential, which are desirable and which are unimportant. This information will be useful when drafting digital preservation policy and strategy. The preservation approach(es) adopted will impact on the Library's ability to satisfy criteria relating to characteristics such as 'look and feel', 'original functionality' and 'authenticity'.

I don't care about the original 'look and feel' and 'original functionality' of an object, the content is more important

75% disagreed with this statement, one researcher was unsure and one librarian agreed.

Preserving 'look and feel' is important for textual objects

The same 75% agreed with this statement. The other two were unsure.

Preserving the 'look and feel' is desirable, but the content is the most important thing

This statement attracted a very mixed response. Two librarians and two researchers agreed; three (one researcher, one archivist, one archivist/researcher) disagreed and one librarian was unsure.

Preserving 'original functionality' is more important for function-rich objects like websites

Three agreed (librarian, archivist, researcher), two disagreed (librarian, researcher) and three were unsure (librarian, archivist/researcher, researcher).

I always want to be able to get hold of the object in its original format

Five agreed (one with the caveat 'if possible'); two disagreed.

5. Authenticity: 'What would satisfy you that an object is what it purports to be?'

Archivists and archival institutions are viewed as trusted custodians of archives. Arguably, archives in more traditional materials are as vulnerable to additions, removals and alterations as digital materials. Perhaps the main difference lies in the researcher's ability to judge the authenticity of an item. An original document of paper or vellum can be appraised against a more familiar set of physical characteristics. We can judge whether the ink, handwriting, letterhead and quality of paper is consistent with what we would expect from an object of the era and genre. We can look for signs of tampering. If digital archives must be migrated to new formats to remain accessible, then the researcher's ability to evaluate the authenticity of the object may be undermined by changes in its appearance. At this stage, perhaps we need to offer researchers proof of authenticity, or at least a record of alterations made to an object.

We asked respondents whether they agreed or disagreed with the following statements:

I trust the archive or library to maintain the authenticity of digital objects

Interestingly, all the researchers agreed with this statement. The only person to disagree was an archivist. One agreement came with the proviso 'depending on the library', and one librarian was unsure.

I need technical proof that an object is authentic

Agreement from three (Librarian, Archivist, Librarian/Other), disagreement from two (Researchers) and two were unsure (researcher and archivist/researcher).

I would want to see an audit trail of actions performed on an object

Agreement from three (Librarian, Archivist, Librarian/Other) and disagreement from three (2 researchers, 1 Archivist/Researcher).

6. What contextual information do users want regarding the accession and management of digital objects?

Changes in the nature of archives, and the processes used in their management, may mean that the contextual information we record about them also needs to change. If archivists adopt regular snapshot accessions for digital archives, then the structure of the resulting archive may differ from that found in paper archives. Early capture of records also affords archivists the opportunity to ask creators about the context of the records: we can ask the 'why', 'when', 'how', 'who', 'where' questions, and make the answers available in archival finding aids. Some digital objects may require additional, or different, descriptive metadata to make them meaningful or technical information to ensure they are usable. With this in mind, we suggested some contextual information that archivists could provide for researchers and asked the Advisory Board to rate these on a scale of 1 to 5, where 1 is 'of no value' and 5 is 'extremely valuable'. There was a very mixed response to most of the suggestions.

Information about any problems encountered during capture e.g. error logs from the capture of a website
Rated 2 on average.

Biographical information about office staff who create records on behalf of a politician
Rated 3 on average, though responses varied between 1 and 5. The project archivists feel that this kind of information is useful.

Date of accession
Rated 4 on average, responses varied between 3 and 5. This information was of most value to information professionals.

Information regarding file format transitions e.g. where a file accessioned in a legacy format, e.g. Wordstar, has been migrated to a modern format, e.g. PDF, for the purpose of preservation and/or accessibility
Rated 3 on average with responses between 1 and 5.

Information about the hardware and software environment used to create the digital object
Rated 3 on average with responses between 1 and 5.

Total data size of an archive, e.g. how many megabytes
Rated 3 on average with responses between 1 and 5.

Total number of files and folders in an archive
Rated 4 on average with responses between 3 and 5.

7. Using Fedora

The evaluation moved on to address the Fedora digital repository software, which board members had used to access the sample objects.

How easy to use did you find Fedora

The majority of respondents found the web interface relatively easy to use: on a scale of 1 to 5 where 1 represented 'impossible' and 5 'very easy', the average response was 3.

How easy to use were the search functions?

The search functions were rated 3.75 on average using the scale above.

Did you think any of the search options were particularly relevant or irrelevant to you?

One researcher commented that searching on particular issues such as 'tax' and 'asylum' resulted in little or no results. This is due the size of the sample and the lack of subject indexing and full-text searching.

A librarian commented that the options relating to Fedora's internal workings and terminology were not relevant. The interface is constructed in this way because it is really meant as a tool for curators. The Fedora developers expect adopters to design their own interfaces in-line with their own needs, or to adapt interfaces designed by others with similar needs.

Are there any search options you would like to see added?

Answers were 'combinations of phrases', 'phrases and dates', 'can you search by ranges (e.g. date range)'; 'would have liked button which took you back to search screen'.

Please tell us any other comments you have on the the use and usability of Fedora

Comments included:

'Needs more user friendly specific field search. Too complex for non-technical users.'

'OK so far as it went but more flexibility in searches needed if possible'

'The descriptions of the functions are very obscure. I didn't know what many headings meant. 'Show and tell' would be better – ie something very clear.

'Fedora doesn't seem to be able to display the hierarchical structure of an archive'

'Seems to work very well'

'I would have appreciated full text search'

8. Future functionality: Would you like to see the following access and discovery features in a digital repository?

The Advisory Board was presented with a list of discovery and access features that could be integrated into a future digital repository. Most respondents said 'Yes' to most features, but more than one respondent circled 'Not sure' for 'Links to similar materials' and 'Tools to manipulate digital objects'. The features listed in the evaluation were as follows:

- *Access to full descriptions of archival collections*
- *Links from full archival descriptions to digital objects*
- *Links from digital objects to full archival descriptions*
- *Links to similar materials*
- *The ability to sort digital objects*
- *Full text searching of documents (similar to the way Google can search pdf and other documents containing text)*
- *Browsing objects by subject*
- *Browsing objects by place*
- *Browsing objects by creator*
- *Browsing objects by date*
- *Tools to manipulate digital objects (e.g. resize, convert to another format, extract text, annotate, etc.)*
- *The ability to save searches*
- *The ability to compile a list of references*

9. What other functionality would you find useful?

'For each digital collection it is vital to have some explanatory introductory material on how the collection was made; in what circumstances' etc. etc. This material can appear very random; it needs to be obtained in such a way that the context of an item is clear to a user.'

' 'Google' text-searching most useful.'

The Advisory Board were given the opportunity for further questions before the meeting adjourned for lunch at Christies Bistro. Discussions on digital political collections continued over a convivial lunch.

Tour of the Labour History Archives Study Centre (LHASC)

The final part of the day was an optional trip to the Labour History Archives and Study Centre, which is based at the head offices of the People's History Museum at 103 Princess Street, Manchester. Janette Martin (who in a previous incarnation worked at the LHASC) gave a tour of the reading room and archive stores and showed the advisory board highlights of the collection including letters from George Orwell, Walter Crane, Ramsay MacDonald and some early Chartist letters smuggled from prison.

Participants also learned about the management of the museum's objects. Museum Registrar, Phil Dunn, provided a brief demonstration of the Museum's Ke Emu software, described the Museum's collections and collections policy, and introduced the group to the large Labour Party photographic collection in his care.

The People's History Museum holds the largest collection of historic trade union and political banners in the world and is the UK's leading authority on the conservation and study of banners. The tour also included a visit to the Textile Conservation Studio and a behind the scenes look at the banner store (where trade union banners are stored once they have undergone conservation). This wrapped up the Paradigm's Academic Advisory Board visit to Manchester and the Oxford contingent departed to catch at train.

Next meeting

We hope to hold the next meeting of the advisory board in May/June at Oxford.

*Appendix A***UNIVERSITY OF OXFORD and UNIVERSITY OF MANCHESTER****Personal Archives Accessible in Digital Media project (PARADIGM)
Academic Advisory Board Meeting****Meeting 2, 12th December 2005
Teaching Suite, John Rylands University Library, Manchester****Programme**

11.00-11.45	Arrival and coffee
11.45-13.00	Using Fedora to explore digital private papers <ul style="list-style-type: none"> • Introduction (Susan Thomas, Project Manager and Digital Archivist) • Hands-on session • Demonstration of a political email archive • Evaluation and discussion
13.00-14.00	Lunch at Christies Bistro
14.30-15.30	Tour of the Labour History Archives Study Centre (LHASC) Tour of the LHASC (Janette Martin, Digital Archivist)

Appendix B

Academic Advisory Board - Terms of Reference

To advise the Project Directors and more broadly the project staff as a whole on matters relating to the academic requirements of personal archives in digital form. Specifically:

3. To advise on the policies relating to the selection of materials, also giving an indication of the relative academic value of different materials.
4. To advise on the policies and applications relating to the usability and access regimes surrounding personal archival material in digital form both from the viewpoint of the researcher, and from the viewpoint of the academic tutor/supervisor.
5. To advise on matters relating to promotion of the project, and advocacy relating to the issues surrounding digital archiving.
6. To advise on other matters as requested by the Project Directors.