

## Digital Lives Research Project

### Case for Support to Arts and Humanities Research Council

Neil Beagrie, British Library, November 2006

#### Research Questions

(i) How are modern personal digital collections of scholarly interest being created, managed and disseminated? What are the significant differences with the past in terms of format, content, or volume and their implications for curators and users?

(ii) What are the needs and views of potential scholarly users of future personal digital collections such as biographers and historians? How can their requirements be factored into new approaches, tools and services?

(iii) How should curators approach selection, preservation and access for these personal digital collections? What aspects of existing practice can be applied? What needs to be changed? What will be the differences between distinct collection areas such as oral history, history of science, or literary manuscripts and correspondence?

(iv) What are the implications of digital obsolescence and ephemeral media for the transfer of personal digital collections from individuals to long-term repositories? Should approaches and timescales for collection development and relationships with potential donors change? Do we need to be more pro-active? Can we develop better guidance, toolkits and services for individuals to ensure preservation before transfer? Should we explore methods for continuous capture of collections over individual lifetimes? What are the views of potential donors on these issues?

(v) Are there existing or emerging tools and services in the public or commercial sector which could assist with management, preservation and access of personal digital collections? Can we identify promising new approaches and potentially transferable practice to the academic sector?

(vi) What are the impacts of legislation such as the Freedom of Information and Data Protection Acts, intellectual property rights, confidentiality and professional ethics and practice (e.g. for scientific and medical information) on personal digital collections and the implications for their dissemination, or acquisition by repositories? Can model deposit agreements for acquisition of personal digital collections be developed?

(vii) How should we address “hybrid” personal collections of digital and traditional media? How can new and established curatorial practice be integrated?

(viii) Are there new organisations acting as intermediaries for managing and publishing personal digital collections? Who are they and what are their services? What could they contribute to preservation? Can curators develop relationships with them as they have in the past with traditional intermediaries such as publishing houses?

(ix) How can we share our experience and outcomes from this research with others and help build a community of repositories and researchers working in this area?

## **Research Context**

Note given limitations of space we have only cited articles and monographs which individually provide more extensive bibliographies and coverage of the research issues and context below.

### **Personal Digital Collections**

The term "personal digital collection" has been used by Beagrie as a working research definition to define the informal, diverse, and expanding memory collections accumulated and maintained by individuals. The term focuses on what is maintained and accumulated by an individual, and excludes, for example, information on individuals that may be held in government sources such as census records or reviews of an individual's work created and maintained by third parties [1].

### **Curatorial Practice and Personal Digital Collections**

To date there has been little related activity in major research repositories on personal digital collections. The PARADIGM project by the universities libraries of Oxford and Manchester provided a best-practice template for establishing long-term access to private digital papers of politicians. A major focus of Paradigm was on building a digital repository: it could only begin to explore other curatorial issues with the six politicians involved [2]. We intend to utilise and go beyond this work, look at other types of donor, larger samples, and new methods and tools.

The British Library itself has several areas of activity directly concerned with personal digital collections and has formed a "Digital Lives Working Group" of curators to help share knowledge across the different teams involved. CVs of curators involved in this proposal are provided in Appendix 3.

### **Digital Preservation**

Digital preservation has been defined as referring to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. [3].

As digital content in personal collections continues to grow, public consciousness of and concerns over digital preservation are increasing. At its most basic level this focuses on better provision and automation for backup of content. It is telling that research on digital data loss has suggested that a substantial amount of personal data is not backed up and that, on average, 6% of data held on all PCs is lost each year (more for laptops because of the higher incidence of theft) [4].

### **Personal Information Management and Archiving**

In a recent small study Marshall et al examined three central questions that would allow them to design a service for personal digital archiving:

- What kinds of digital belongings do people have and what do they value?
- How do people archive their digital belongings now?
- What are the central archiving challenges stemming from current practice, digital genres, and home technology environments that will guide archiving service design?

In particular it focused on three main themes: consumer strategies and the gaps between principles and practice; specific observed challenges for implementing a digital archiving service; and overlooked environmental factors that must also be taken into account [5]. It provides findings and hypotheses which we will explore and develop.

### **Postcustodial and Digital Preservation Lifecycle Methods**

A traditional approach of capture close to or at the end of an individual's life for a collection may pose significant challenges in a digital environment, including obsolete formats and media, and missing data (email, web pages, etc.) or access gateways such as passwords. We intend to explore the boundaries and new ground in postcustodial [6] and digital preservation lifecycle methods [3] and their application to personal collections. For example continuous digital capture or synchronisation of collections with a repository is one approach that we wish to explore.

Several companies now offer online backup of digital data to a remote secure repository using synchronisation software as a safeguard against data loss e.g. British Telecom's Digital Vault service. Others are offering secure web-hosting of selected personal data such as address books and contact details that can then be centrally maintained and accessed from different mobile and fixed devices. The Internet Archive and other partners have established Ourmedia. Individuals can store their content for free in perpetuity on Ourmedia's servers, as long as they're willing to share their works with a global audience.

Intermediaries often can play a significant role in the chain of custody and preservation lifecycle. In the digital environment, new intermediary services and projects have emerged for individuals to publish creative writing, blogs, or digital images such as, Flickr, the TRACE online writing archive, the "One Day in History" mass blog for 40,000 individuals, or the Millennium Memory Bank of

6,000 individual oral histories (both of the latter collections have or will be acquired by the British Library).

### **Computer Forensics**

If collections are not captured close to the point of creation or managed in contemporary formats, any surviving parts of the collections are likely to be transferred to repositories in obsolete file formats and on obsolete storage media. The process of “data archaeology” in such circumstances has been studied by Ross [7]. More recently the British Library has been exploring the use of computer forensics technology in capturing and preserving digital manuscripts in modern papers. With forensic technology it is possible to recover lost files and fragments. It is also possible to extract embedded metadata such as the dates and times when word processed documents have been created, modified and accessed, and to establish provenance.

### **Digital Memory**

In the UK "Memories for Life" has been recognised as a Grand Challenge for Computing Science by the UK Computing Research Committee. An interdisciplinary Memories for Life Research Network has been funded by EPSRC and a review paper of the science and technology published [8]. It spans a wide range of potential applications from health and aging populations (digital memory aids for individuals), to life-caching and personal digital agents and indicates many potential future developments that could influence personal digital collections.

## **Research Methods**

The research consists of a series of linked studies arranged to explore key issues, followed by a process of mapping and assessing impacts on curatorial practice, and finally dissemination of our findings:

### ***Study of personal collections and dissemination practices of researchers and authors (Work Package One)***

This research led by Ian Rowlands and supported by the research assistant at UCL and BL curators, will be undertaken in two stages. The first stage (October 2007-February 2008) is intended to provide deep insights into the issues, concerns, practices and behaviour of representative individuals sufficient to generate appropriate, telling questions for the second stage NOP survey (March-May 2008):

(a) A qualitative survey will first seek to scope and categorise the information objects held in the personal digital collections of selected researchers and authors and explore potential links to research repositories. The study will do this via questionnaires, and interviews of the individuals. The sample will consist of 20 senior/mid/ and early career individuals selected to provide a representative

sample of fields, ages, and gender and approached to participate by the BL and UCL. Short pre-interview questionnaires and briefing papers will be circulated followed by in-depth interviews undertaken by the research assistant at UCL and BL curatorial staff.

(b) The second-stage study would then employ a questionnaire to gather and analyse a quantitative set of data via a survey outsourced to NOP. We would use randomized samples of research-active scholars, drawn from ISI mailing lists and use a web-based survey method to elicit insights into the awareness of the problem, attitudes regarding longer-term preservation and personal information management. We would aim for around 1,000 completions. This would provide some essential behavioural and attitudinal benchmarks as well as highlighting differences (by age, gender, subject, geography).

Deliverables will include:

- in-depth interviews with selected individuals
- NOP survey findings and analysis
- working paper reporting and analysing the findings and potential implications
- a chapter in the final project synthesis.

### ***Legal and Ethical Issues (Work Package Two)***

This research will be undertaken by Andrew Charlesworth between September 2007 and February 2008. It will provide stakeholders - individuals and institutions seeking to develop digital collections and repositories, and to disseminate their contents - with a detailed survey and description of the legal and ethical environment, and an overview and analysis of current thinking as regards practical approaches to legal and ethical issues. It will consider the impact of legislation affecting digital information resources (e.g. freedom of information; privacy, confidentiality and data protection; intellectual property rights; and professional ethics and practice) on personal digital collections and the implications for their dissemination, or acquisition by repositories.

The methodology for this work package will be based primarily on desk research, including a literature review, and on-line research examining the practices and policies of repositories. It will, however, also seek to draw on work packages 1(a) and 3 and BL staff, in terms of obtaining feedback from key user groups on the legal and ethical issues that they see as most problematic for personal digital collections, and eliciting suggestions as to how such problems might be overcome.

Deliverables will include:

- a working paper on legal and ethical issues
- model deposit agreements for acquisition of and access to personal digital collections

- and a chapter on legal and ethical issues in the final project synthesis.

### ***User Focus group (Work Package Three)***

With the support of professional bodies a user focus group will be held at and organised by the British Library. This will be held in October 2007 to provide early input on requirements from key user groups such as historians and biographers including representatives of professional and learned associations, specialist intermediaries, and selected individual scholars with relevant research expertise. The participants in the focus group would also be invited to the project conference near the end of the project for discussion of early outcomes and recommendations from the research.

Deliverables will include:

- engagement of key user groups
- a report of the focus group meeting and key outcomes
- a chapter in the final project synthesis.

### ***Desktop Studies of (i) promising technologies and (ii) services offered by new intermediaries (Work Package Four)***

This research will be led by Neil Beagrie supported by the research assistant and specialist input from Paul Wheatley and Jeremy John over 3 months (March-May 2008). It will have two elements:

(a) desktop literature and online research to assess promising and potentially transferable technologies from elsewhere including the EU PLANETS project selected by BL curators and technical staff. These will include synchronised data capture and backup, peer to peer file sharing, format converters and normalisation tools to support acquisition and collection management, and computer forensics.

(b) Desktop research to assess the approaches to preservation or attitudes to long-term repositories of new digital intermediaries. We will include ten services and projects in the assessment nominated by curators at the BL and researchers at UCL. They will include online intermediaries offering “archiving” services such as Ourmedia and Ark, data management and publishing intermediaries such as Flickr and Google, and collecting projects such as Millennium Memory Bank, “One Day in History”, and Trace. We will develop a categorisation of service components for digital preservation to compare offerings by online services. We will explore attitudes and roles of collecting projects via interviews with key individuals involved in their creation and transfer between organisations.

Deliverables will include:

- a categorisation of service components for digital preservation and comparison of offerings by online services
- working paper reporting and analysing the findings of desktop research and potential implications for digital repositories and services

- a chapter in the final project synthesis.

### ***Mapping the 'personal' with the professional (Work Package Five)***

This research will use the outputs from work packages 1-4 to map against current selection, acquisition, description, activities, and process and workflows at the British Library. It will be undertaken between May-September 2008 by Katrina Dean assisted by feedback and discussion of issues with a focus group of curators and users drawn from the BL, other institutions, and members of the advisory board.

Deliverables will include:

- Identifying gaps and points of convergence, implications for selection policies and workflow and for interoperability and 'joining-up' of distributed content
- Assessing and evaluating potential technical solutions and gaps
- a focus group held at the British Library to obtain feedback from curators
- a report of the focus group meeting and key outcomes
- a chapter in the final project synthesis.

### ***Dissemination and Knowledge Transfer (Work Package Six)***

An active programme of outreach and dissemination to the library, archive, and computer science research community will be a critical component of the project. We wish to share our experience and outcomes with others and help build a community of repositories and researchers working in this area.

Deliverables will include:

- Full Dissemination/Publications Plan (September 2007)
- Creation and updating of the project website
- Participation in Memories for Life, Digital Preservation Coalition and other professional workshops and fora
- Publishing working papers, project synthesis, and reports on project website and articles in relevant journals
- Holding a conference in 2009 to disseminate and discuss project outcomes

### ***Project Management (Work Package Seven)***

The project will be directed by a project board of senior staff consisting of Prof. David Nicholas and Ian Rowlands (UCL), John Tuck, Katrina Dean and Neil Beagrie (BL). The project board will hold four meetings over the course of the project.

The principal investigator Neil Beagrie will be responsible for overall management of the project and first point of contact for all exchanges with the AHRC. The co-investigators Katrina Dean and Ian Rowlands will co-ordinate and supervise respectively the curatorial work and investigative survey.

Staff will attend quarterly project team meetings and quarterly reports will be circulated to all participants.

A project advisory board will provide input to development of the project and act as external “advocates” of its work. The board will meet twice over the life of the project.

A GANTT chart is provided as Appendix 5 as a visual illustration of the project timetable and selected milestones.

### **Statement for Speculative Research**

The partners believe this project is highly appropriate to the speculative research route for three reasons:

- This is a largely unexplored area of research and curatorial practice where we are seeking significant breakthroughs in knowledge and approaches by evaluating new techniques, methods and tools;
- Some of the approaches we wish to explore are highly experimental and the outcomes uncertain. However they have the potential to make a substantial impact on the future development of research collections and to provide the future foundation for larger scale research projects;
- The research is interdisciplinary and establishes important relationships between research skills and cutting edge curatorial practice.

### **References**

[1] N. Beagrie (2005). “Plenty of Room at the Bottom? Personal Digital Libraries and Collections”, *D-Lib Magazine*, Volume 11 Number 6, June 2005, <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>.

[2] *Paradigm Workbook on Digital Private Papers* retrieved 23 October 2006 from: <http://www.paradigm.ac.uk/workbook/index.html>

[3] N. Beagrie and M. Jones (2001) *Preservation Management of Digital Materials: a Handbook*, British Library 2001.

[4] D. M. Smith (2003) *The cost of lost data*. Graziadio Business Report: Journal of Contemporary Business Practice, vol.6. Los Angeles, CA: Pepperdine University.

[5] C. Marshall et al (2006) “The Long Term Fate of Our Digital Belongings: Toward a Service Model for Personal Archives” *Proceedings of IS&T’s Archiving 2006 Conference*. Springfield, VA: Society for Imaging Science and Technology, pp. 25-30

[6] F. Upward (2000) "Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond: a personal reflection", *Records Management Journal* volume 10 Issue 3, Dec 2000, pp.115-139

[7] S. Ross and A. Gow (1999) *Digital Archaeology: Rescuing Neglected and Damaged Data Resources*, British Library Research and Innovation Report 108.

[8] K. O'Hara et al (2006), "Memories for life: a review of the science and technology", *Royal Society Interface Journal*, Volume 3, Number 8, June 2006, pp 351 - 365.