BRITISH LIBRARY

DIGITAL >>LIVES

An Initial Synthesis

DIGITAL LIVES

Personal Digital Archives for the 21st Century

>> An Initial Synthesis

by

Jeremy Leighton John
with
Ian Rowlands
Peter Williams
Katrina Dean

THIS IS VERSION 0.1A

WELCOMING SUGGESTIONS & FEEDBACK

A Digital Lives Research Paper

AN ACCOUNT OF
THE DIGITAL LIVES RESEARCH PROJECT 2009
A research project funded by the
Arts and Humanities Research Council

Grant BLRC 8669 (Research Grants: Speculative)

A synthesis of the Digital Lives research project, led by the British Library, in partnership with University College London and the University of Bristol, and funded by the Arts and Humanities Research Council

http://www.bl.uk/digital-lives/index.html

The synthesis considers the curation of personal digital archives across the whole archival lifecycle

Beta Version 0.1A, 22 February 2010

Possible citation style: John, J. L.; I. Rowlands; P. Williams; and K. Dean (2010) *Digital Lives. Personal digital archives for the 21st century >> an initial synthesis*, Digital Lives Research Paper, 22 February 2010, Beta Version 0.1A

All hyperlinks will be checked in subsequent versions

A Note on the Beta Version

In the spirit of the times we are making the Initial Synthesis available first as a Beta Version. Readers are invited to send their thoughts and suggestions to the email address: jeremy.john@bl.uk

We may not be able to reply to all email messages: for one thing we do not know how many there will be, but we shall read all of them. Please identify yourself, and if available please use your institutional email address. Please do not send any attachments

With many thanks in anticipation

Please take care to cite this document as Beta Version 0.1A

"Turned aside into the meadow to look at the great stone of Cross Ffordd. It is a long time since I stood beside it, and I had forgotten that the stone was so large. I suppose no one will ever know now what the grey silent mysterious witness means, or why it was set there. Perhaps it could tell some strange wild tales and many generations have flowed and ebbed round it. There is something very solemn about these great solitary stones which stand about the country, monuments of some one or something, but the memory has perished and the history is forgotten"

29 March 1871, Diary of Reverend Francis Kilvert
living in the Borders of Wales

Summary

Abstract

Digital Lives: Personal Digital Archives for the 21st Century
>> An Initial Synthesis
Beta version 0.1, 22 February 2010
Jeremy Leighton John, Ian Rowlands, Peter Williams and Katrina Dean
A Digital Lives Research Paper

The digital era has changed the nature and scope of personal archiving forever. The Digital Lives research project accordingly has examined both theoretical and practical aspects of curating personal digital objects, or eMANUSCRIPTS, over the entire archival life cycle. This initial synthesis offers an overview of the emerging field of personal informatics and personal curation. It contemplates three audiences: the individual who is leading a digital life and creating a personal digital archive, the practicing professional archivist and curator, and the scholar and scientist who is accessing the contents of personal archives for research purposes.

Methods and procedures included person-to-person interviews, online questionnaires, workshops and focus groups, online and library research and literature reviews, online interviews, a research conference with real and virtual world presentations and attendance, and technology testing and implementation. Creators, archivists and scholars have all made important contributions as have digital specialists and media researchers.

It is concluded that the role of personal archives in daily life and their research value have never been more profound. The potential benefits to society and to individuals are both deep and far reaching in their capacity to empower research and human well being and advancement.

With society and humanity facing some very demanding challenges and changes in the coming years and decades it is imperative that analysis of the way people live their lives in relation to their cultural, social and natural environments can be conducted regularly and systematically, based on reliable sources of personal information - obtained and used ethically and legally with the full and ongoing support of participating individuals. Life information promises to be an invaluable resource in monitoring the natural environment, in capturing historical events and precedents as witnessed by people everywhere, and in comprehensive recording of literary, artistic, sociopolitical and scientific endeavour and enlightenment. At the same time it represents a possible emancipation of people generally, allowing interested individuals of the digital populace to have their personal and family memories, creativity and unfolding lives, acknowledged as a persistent personal archive and through lasting personal digital objects.

A prominent perspective highlighted by the project is that of recognising the diversity of attitudes and inclinations among people. Not everyone will want to engage in these activities at first; and people will always vary in the extent to which they do so. Some, however, will be early adopters, and already many individuals are holding varying quantities of personal life information; and it is anticipated that this phenomenon will increase rapidly. A critical motivation for research into the personal information that will be increasingly held in personal archives by means of digital technologies is - slightly paradoxically - the need to understand the impact of these technologies on people's lives.

With this in mind it is possible to envisage a network of repositories, large and small, cooperating towards two common goals: (i) the building up of holdings of personal archives for institutional, local, regional and national collections and scholarship, and (ii) the promotion and facilitation of sustainable personal archives retained by members of the digital public and their families, as archives in the wild.

It is postulated that roles of repositories may lie sometime in the future not only (i) in garnering selections of personal archives and personal digital objects, and (ii) in providing enhanced security and access for the most culturally and scientifically valuable personal archives, but also (iii) in serving as guardians of the authenticity of the originals including digital objects in the wild, (iv) in providing services that strengthen and assist the effectiveness of archives in the wild, and (v) in developing an expertise in the content of archives in the wild as well as the repositories own digital holdings.

For sure there are major challenges ahead. Digital capture and preservation are two key activities central to the sustainability of personal archives and the holding of life information by individuals and repositories. The project emphasises, however, a third interrelated activity of digital function or utility, which seeks to understand, improve and diversify the actual and potential usefulness and value of the personal digital object as far as it is possible. It is in some sense the most fundamental activity for it ultimately motivates the expenditure of resources and effort on stewardship and provision and warrants much more attention.

The project has outlined the concept of Personal Informatics to encapsulate these three concerns of digital capture, preservation and utility in the context of personal digital objects, and to embrace the study of digital personal information in all its manifestations.

In addition to highlighting gaps and necessary research for the future, the synthesis shows how many of the technical challenges are being met with new, modified and integrated techniques emerging from the digital curation, preservation and resource discovery and interpretation communities.

In particular this study has brought to attention forensic capture for authentication, to be combined with file format conversion or migration for longlasting interoperability, and exact replication of files with high fidelity emulation for historically reliable perception and interaction.

The notion of enhanced curation has been promoted with exemplar activities directed at capturing comprehensive contextual information has been proposed, along with an adaptive approach to curatorial challenges.

The synthesis highlights crucial and influential work being done in the digital preservation community. The Planets Framework is delivering a coherent focal place for key preservation activities. Among the most exciting developments are a series of modular emulation projects and the ongoing practical refinement of the universal virtual machine concept. Many of these tools are nonproprietary or open source, and will transform the situation for hard pressed archivists.

The shape of an archival lifecycle is outlined at the conclusion of the synthesis with illustrative tools for each phase.

The recent progress in digital curation and preservation is encouraging and heartening. Yet there is clearly much more to be done before contemporary archivists are universally applying these processes as everyday activities in their professional lives.

To this end, five strategic modules each comprising a pair of closely related activities are highlighted. One of these activities is research: crucial in stimulating, informing and consolidating the practical progress being made, and most especially in allowing for an ongoing advancement and innovation in the face of a continual evolution of digital and modern life.

Twelve Sets of Observations

>> INFOETHICS

Closely allied to digital utility or functional value is the interpretation and assessment of the ethics and regulation of personal information, infoethics. This is one of the most demanding of challenges due to the diversity and changeability of opinions, the uncertainty of consequences, and the unceasing appearance of new and powerful technologies. Yet these concerns are not unique to personal archives. Emerging technologies will continue to test many aspects of personal, social, cultural and political life far into the future, and acceptable solutions will need to be found time and time again. Dynamic and mediated access to numbers of personal digital archives could provide researchers with a timely and continuous means of assessing the impact of technologies in detail, supporting for example the development of effective privacy enhancing technologies and policy instruments.

>> AUTHENTICITY, FORENSICS AND EMULATION

Research depends on the authenticity of the digital object and its presentation. Forensics provides essential tools for the modern curator as does high fidelity emulation that ensures that digital objects can be perceived in a manner close to the original 'look and feel'. Authentication will be a fundamental role of future archival repositories and libraries. Document analysis, provenance, digital materiality, critical textual analysis, iconography and even palaeography: these all demonstrate continuity in digital scholarship.

>> ENHANCED CURATORSHIP

The research value of an archival object can be greatly enhanced through the collection of contextual and complementary information. Examples include video conversational tours; panoramic photography and immersive graphics of creative environments, from studio to laboratory; and compilation of life inventories, the documenting of personal libraries and artefacts, with selective digitisation in 3D. A key aspect of the approach is its association with the original archival objects - digital and analogue - as well as the person and the person's landscape: interviews in the presence of the letters, diaries, field notebooks, ciné films and emails that belong to the lives of writers, reformers and scientists. Remote viewing of critical experiments in a prestigious laboratory, online master classes with renowned writers discussing their archival materials, and frank and private political conversations on historical developments or contemporary events as these happen - mediated by a virtual curator - are possibilities that would enrich the archival experience.

>> TRAINING AND GUIDELINES

As repositories receive personal computer media, there is an urgent need for a network of programmes directed at training existing and future curators and archivists. Many archival repositories are likely to continue to rely on a modest number of individuals who will

increasingly have some digital training and experience and be expected to process both digital and analogue materials. Personal digital archiving is a paragon of the multidisciplinary and interdisciplinary field founded on multitasking archivists who possess a wide range of skills and knowledge, with each individual also maintaining one or two specialisms that contribute to a team's overall operations. Wide reach can be enabled through audio and video podcasts and virtual training and conference centres.

>> iCURATION

With more and more of digital life taking place online, a need for curatorial and archival activities to follow suit is identified. The advance of iCURATION would allow archivists to undertake many of their processes such as previewing, acquisition and supply of archival objects remotely. Concerning the digital objects of people generally, repositories could investigate the feasibility of helping to guard longterm provenance through authenticating and registration services involving hash values, persistent identifiers and handles of digital objects. The technical aspects of tracking and measuring the use and value of personal objects could be explored along with infoethical implications. An openly available authentication facility might be combined with preservation services of the kind developed by the Planets project: with remote characterisation and validation of digital objects and emulation, as well as remote visualisation and analysis services, and - within limits - digital object storage services.

>> COLLABORATION AND INTEGRATION

Another core aspect of iCURATION is the networked integration of repositories and of analogue with digital materials. An obviously powerful property of digital archives is the capability of integrating content across repositories, with the chronological sequencing of digital trails of influence. This requires careful agreements and interoperability standards. Fast digitisation of analogue materials combined with inventory on reception makes it possible for born analogue objects to be interpreted alongside born digital objects. The network of archival repositories as a whole is encouraged to safeguard comprehensive instances of digital lives from across the diversity of the social spectrum as well as focussing on the most influential and creatively renowned individuals. An outstanding question concerns the relationship between analogue processing and its digital counterpart. Analogue archives tend to be more severely constrained by space, potentially meaning that vastly greater numbers of digital archives could be captured and retained. Highly advanced encryption and security processes may be critical to the scale and viability of institutional sharing activities.

>> PARTICIPATION AND ADVANCED CATALOGUING

Digital archives are inherently receptive to automated or quasi-automated procedures that potentially allow large volumes of digital objects to be processed. Nonetheless, considerable value can be added to these collections through handcrafted metadata and description as well as experienced supervision of quasi-automated cataloguing. It is recommended that metadata creation and cataloguing content be opened to creators and families, expert researchers and others. The use of metadata icons and visual relationship schemes, annotation jamborees, and wiki-style systems with detailed authorship-tracking should be explored systematically.

>> FROM LIFETRACKING TO PERSONALISED USABILITY

With lives becoming increasingly mediated through digital technologies and with increasing digital opportunities and benefits for the individual, there is an apparent trend towards holding and dynamically engaging with personal information by individuals. Life tracking, life

caching, personalised medicine, personalised usability, biometric and individualised security, context aware ubiquitous computing and digital portfolios and learning: all rely to a significant extent on personal information such as identity, activity, creativity and sociability patterns, which will need to be safe and authentic. It is more than conceivable therefore that personal digital archives in some configuration and at some location will underly digital life in a fundamental and universal way.

>> ARCHIVAL PIM AND ARCHIVES IN THE WILD

Effective, versatile and robust personal information management (PIM) can be expected to emerge in time as demand for efficient handling of personal information mounts; but at present such a comprehensive capability is sorely missing. There is a specific need to promote an archivally-oriented form of PIM that embraces the entire information life cycle, and is directed at securing authentic personal digital objects and making them readily available for use and reuse by the individual creators and owners beyond the immediate future. Notwithstanding increasing capacity for storage, it seems likely that some selection of digital objects will continue to be necessary and desirable for many people and for the time being. A key requirement will be a process for ascertaining which objects are to be favoured, based on a combination of identifiable future value, available resources, serendipity, future possibilities and even random choice, according to the wishes and sentiments of the individual.

>> ADAPTIVE CURATION AND MANAGEMENT

A rapidly intensifying challenge of the 21st century is the organisational structure and management of information systems and institutions, small and large, and the ability to respond effectively to the ongoing and expanding emergence of technologies that will lead to sustained change and occasional upheavel. This will be demanding enough for organisations that are assessed according to a relatively straightforward metric such as profit. It is even more so for those that are answerable to diverse stakeholders and have longterm responsibilities. In recent years novel techniques have been sought and explored in management including the use of agile and evolutionary procedures that can be contrasted to techniques that attempt, for example, to predict in advance, detailed requirements and specifications. Research provides an important way to anticipate change as far as sensibly possible, and in combination with gradualistic, flexible and iterative development processes, offers a possible way to adapt to unremitting change and uncertainty. Pertinent techniques will no doubt continue to be produced and refined. Critical to the adaptability of an organsiation will be its system of internal communication and innovation.

>> EVOLUTIONARY DYNAMICS AND COMPLEX NETWORKS

A unique analogue object can only be passed on as a single entity. One of the most consequential changes in the digital era is the process of personal digital objects being replicated and spread across a population of personal digital archives existing in the wild and within repositories. In being united by the investigation of the past, history, geology and palaeontology, evolutionary biology and forensic science can and already do benefit from fruitful cross fertilisation of ideas, techniques and approaches to research. This can be extended much further and offers significant theoretical avenues too. In particular, evolutionary and complexity perspectives offer the prospect of advancing the theory and practice of personal informatics in fundamental ways.

>> ADVOCACY

The most crucial requirement for the immediate future is for advocacy on behalf of personal

digital archives, both in motivating individuals to keep and develop them and in inspiring decision makers to provide policy and funding support to secure and maximise the effectiveness of this essential resource for research and for individual well being. It would be useful in the coming years to explore and measure the research value of personal digital objects.

Some of the Key Findings

>> *Digital preservation and the longterm*

Among creators the current situation overall regarding longterm preservation is unpropitious. Beyond a sense of digital vulnerability manifested partly through the printing out of valued documents, the concept of backing up files and the storing of copies at more than one site, there is little effective awareness of the technical concerns that relate to digital preservation among members of the digital public and academics.

In response to an online questionnaire, the backing up of files was reported as a preservation measure by 72% and 54% of academics and the digital public respectively. Yet, over a third of computer users indicated that they do not possess backup software, and of those that do so, less than half of them use the software regularly at least once a month.

Arrangements made in case of sudden death or incapacity are limited: about a quarter or less of academics set aside details of webmail services and of passwords.

>> *Organising and finding*

It was apparent from interviews that personal digital archives grow organically with new files and folders being created alongside existing ones with little appetite for systematic and regular deletion in many individuals. The hierarchical folder system was widely perceived as information in its own right, and there is a mistrust of the automated naming and dating of files and folders.

The questionnaire revealed a widespread effort and desire to organise files but there is considerable diversity. Approximately 50% of people either tidy their folders and files regularly or do so at the end of projects. There is a widespread tendency to give names to self-created files (typically alluding to subject): for instance nearly 95% of academic respondents give a file name to a self-created file deemed to be of great importance - ca 46% subject descriptive, ca 18% date-or-version descriptive, and ca 31% subject and date-or-version.

Of academics, 31% make several copies of files at different locations in order to ‘make sure that you can subsequently find files on your computer’; less than 5% do nothing for this purpose.

Memory reportedly plays a key role in finding files: for example, 71% and 85% of academics are ‘quite or highly dependent’ on memory of filename and of folder or location, respectively.

More than 95% of individuals reported that it was ‘quite easy’ or ‘very easy’ to find files when needed.

>> *Serious loss*

On the other hand between one quarter and one third of the two populations (academics and digital public) reported a serious loss of computerised information at home. In nearly 70% of

these cases this loss manifested itself as an inability to find the information; by comparison, 4% & 8% of cases were due to hard drive failure.

One interpretation is that this might reflect a diminishing ability to find files as more time passes since they were used or created, a view that concurs with the general dependence on personal memory for finding files. It is suggested that an archivally-oriented system could help to sustain the use and reuse of personal digital objects for much longer.

>> *Attitudes*

Following the purchase of a new computer, academics are more inclined than the digital public to transfer some or all of their data to their new machine: 30% of the digital public respondents did nothing with the data, it being lost or deemed unnecessary.

Most people do not retain interim versions, drafts, of 'precious files', either because there is only one version or because interim versions are actively deleted; approximately 15% and 20% of people actively deleted drafts. Just over 40% of academics reported keeping all versions of a special file, whereas 26% of members of digital public did so.

>> *Value*

For academics, the primary value of the file of great personal importance was in 46% of instances its 'interest to future historians' and in 23% of instances its 'sensitive, personal or financial' nature. For the digital public, its primary value lay in its 'sensitive, personal or financial' nature and its 'interest to future historians' in 32% and 20% of instances respectively.

In choosing a file of great personal importance, 90% and 77% of academics and digital public chose a file which they had created themselves rather than one that they had acquired.

The main reason for archiving computer files is 'as a witness to creativity' according to 63% and 45% of the academics and digital public samples, followed by 'sentimental reasons, personal memory' by 15% and 26% respectively. There is, however, a noticeable diversity of reasons selected as the primary explanation by respondents.

The literature review and interviews emphasised that the exact identification of the future value and usefulness of personal digital objects is not trivial. A strong and prevalent rationale for keeping files is the feeling of 'just in case'. Useful objects are commonly mixed with less useful. Some files which have never been used are still retained.

>> Digital lives in the cloud

That personal objects and content are being made widely and increasingly valuable through digital technology is demonstrated every day, hour and minute by online service providers such as Facebook, MySpace, Flickr and YouTube.

Cloud computing and the use of web 2.0 services are manifestly of very great appeal to people. The online service providers are able to attract many users because they provide compelling services that allow people to create and share files and content either globally or to a restricted set of people, and to record and share events in their lives and the lives of family and friends. The success appears to tap into and demonstrate with little doubt, human desire for personal creativity, social validation, recognition, connectivity and legacy; and yet in offering these services, many online service providers have avoided any commitment to the longterm security and preservation of the information, to demonstrating its authenticity in

the future, and any guarantee that a set of information will remain an integral whole. Recent tales of woe in the newspaper and magazine press reveal all too clearly the limited preservation nature of these services.

It is recommended that consumer watchdogs and ombudsmen are engaged in evaluating personal information services and tools for their archival flexibility and sustainability, legal and ethical transparency and accountability, and capture and preservation attributes. For example: Is it possible for an individual to download readily and regularly the entirety of their own contents and profile to their own computer?

To this extent there is an open niche available albeit a challenging one. Public, and perhaps commercial, repositories acting together and possibly in collaboration with amenable online service providers might offer longterm preservation and retention, authenticity, integrity of an archival whole, a high and sustained ethical standard, standarised legal annotation system, and ultimate receptivity to public and longterm interest.

>> *Technical aspects of personal curation*

Significant advances have emerged in the technologies of digital curation and preservation which need to be conveyed to curators and archivists widely.

It is necessary both to retain digital replicates of the personal digital objects for ensuring authenticity and full informational content, and to create digital facsimiles (of varying fidelity) that are interoperable and comply with modern expectations for longterm preservation.

Along with this conversion of file format, known within the preservation community as ‘migration’, there is agreement that emulation is an essential approach for the future, and may ultimately be the preferred access route for many eMSS scholars, warranting much more research and development. Among the most exciting developments are the emulation projects of Dioscuri and KEEP, and the ongoing refinement of the Universal Virtual Machine.

Forensic technologies such as write blockers and forensic ‘imaging’ software provide the ability to authenticate the capture of computer media. Existing forensic software offers a model for a curatorial examination environment, with an audit trail system that tenders a chain of custody along with examination and action histories. A variation or derivative form of environment could be created for archival and curatorial purposes with more appropriate terminology and inclusion of specific archival functionalities. Such a tool could be developed by the digital curation community or existing producers of forensic software could be coaxed into creating archival versions.

The use of hash libraries for identifying files including application software is an approach that should be explored further by the personal archive curation and preservation communities. Hash libraries can be used to identify actual known system, application and documentation files along with file type, and a comparison with the output of characterisation and validation processes would provide a test for identifying discrepancies.

It is in any case crucial that software is archived by and for the curation and preservation communities. The options of voluntary or legal deposit for software should be considered further. An understanding and appreciation of ancestral computers, their workings and manner of use, can be anticipated in both curators and scholars of the future, and is already evident in digital humanities scholarship.

The contemporary development of reconfigurable and modern hardware for use with ancestral computer media and software is an exciting and important progression, especially as a means of enabling digital capture and informing development of emulation.

>> *Curators*

There was a perception among curators and archivists who contributed to the project that much of the activity within the field of digital preservation has been of little practical relevance to them. Some archivists anticipated that the necessary digital technology would require resources that are unavailable and would be oriented towards major operations by large institutions. There was a desire for solutions that can be scaled down as well as up.

This perception based on experience to date will hopefully change with the completion of the Planets project for example. The Planets Interoperability Framework, Planets Testbed and Plato Planning tool are implementations and services that are relevant and extensible to small archives as well as large archives, and the overall package promises ready installation with some clicks of the mouse. Many of the emerging solutions are open source, and some of them (eg Hoopla) are being developed specifically for small institutions.

Instruments for assessing risks (eg Drambora) and costs (eg LIFE3) have been produced or are on course for delivery.

Nonetheless, there is clearly a need for ongoing and extended outreach and training directed at multitasking curators and archivists working at small archives, and for a significant effort in tailoring the techniques and services for large numbers of small archives. Many archivists will need to be trained in the use of planning tools such as Plato. The more user-friendly the tools and the more comprehensive the training programmes, the sooner a digitally active and extensive archival community will exist.

With regard to capture, curators expressly want to engage in a hands on way with processes of digital archive curation, and would welcome training in the practicalities of forensically authenticated capture and metadata extraction, file conversion, use of emulators, and in the basics of ancestral computing and digital object analysis.

Archivists were highly receptive to the participation of creators, family members and researchers in metadata creation with contributions being tagged according to origin.

Curatorial processing and cataloguing can be and should be advanced with automated and supervised technologies and with semantic information and ontologies.

Comprehensive and convincing open source cataloguing tools and platforms suitable for archival digital objects in personal context remain thin on the ground but the Archivists' Toolkit is a very promising development.

>> *Computer use and the nature of digital personal archives*

Over 35% of people responding to the questionnaire kept a hard copy, ie a printout, of the file which the respondent had identified to be of great personal importance. This emphasises the dual nature of personal archives.

Archivists widely agreed that the seamless integration of the analogue and digital components of a hybrid archive could be best realised through digitisation of the analogue objects (eg

paper letters), a process that can be combined with inventory on reception (although there was a concern about available resources).

Although 95% of the digital public respondents use Microsoft Windows - with only 3.5% using Apple Macintosh computers - just over 16% of academic respondents reported the use of Macs, with nearly 80% using Windows. This finding can be further contemplated in the light of Apple's prominence in consumer use of laptops and handheld devices, and the various operating systems in mobile devices including new ones such as Google's Android. Even when dealing with contemporary computers, archival repositories can expect to encounter significant system diversity.

The role of the operating systems (eg Windows vs Apple vs Linux) in shaping personal information management warrants further detailed study.

More than 90% of the populations sampled are using a home computer that was purchased no more than five years ago, with very roughly 5-10% of people sampled using their first purchased computer.

Flash media have quickly emerged as a popular form of information storage at home, along with optical disks (recordable CDs and DVDs) for both academics and members of the digital public. The external hard drive is an important option that is based on relatively reliable and high volume magnetic media but is used by just under 40% of the digital public respondents and by 60% of the academic ones. Floppy disks remain in use with just over 10% of the digital public.

On being asked to consider a file of great personal importance, the file category was: for academics, a 'word processed document' in 41% of cases and 'photographs or digital art' in 21% of cases; for the digital public, 'photographs or digital art' in 34% of cases and 'text-based documents other than word processed ones (such as PDFs)' in 16% of cases. Moving image featured relatively little but this will likely change rapidly in coming years. As might be expected a wide range of file types and dates-of-origin are to be found in personal digital archives.

>> *Users*

Users indicated that archives arriving at a repository should be processed promptly applying as much automation as possible.

There is still a perception that repositories form a closed community. To some extent this reflects the role of the curator as a mediator between the not entirely identical interests of the creator, third parties and the consumer. To help counter this perception, regular and genuine communication between curators and users is vital.

Users and researchers should be encouraged to contribute to the location and selection of objects for archiving and also to the creation of complementary contextual information and catalogue entries and metadata.

Together with timely access, the primary concern of the users was the authenticity and provenance of the digital objects. This can be met to a significant extent by the use of computer forensic technologies, as shown by this project.

Creators and users could be brought together directly where appropriate - for oral history

interviews, metadata creation, and also for greater understanding of need for sensitivity in personal matters - through careful mediation by curators.

An important aspect of access provision that requires research is the level of fidelity of the user's experience with the digital object, and most especially the mode of referencing this experience. The digital objects presented will be access versions of digital replicates, and of digital facsimiles, and these may be of varying fidelity and resolution; moreover, these may be accessed *via* various routes (eg online on remote computer or onsite with one of a number of diverse repository computers); a language and means are required for characterising the hardware and software that determine the exact experience and for providing a record for the user.

>> *Legal and social environment*

Repositories, particularly public repositories, are required to meet and balance the wishes and interests of a very diverse set of stakeholders. With emerging digital technologies and their social and economic consequences, legal systems are presented with new, often unanticipated, scenarios for which they were not designed. This has led to some uncertainty in the interpretation of legal requirements, and not inconsiderable divergence in the application of procedures for addressing them.

It is concluded that archival repositories should adopt a pragmatic, multifaceted approach: (i) pass some of the onus to creators and originators of the archive for tagging with suitable metadata instances where issues of privacy, copyright and liability exist; (ii) use automated and supervised searching and filtering for identifying objects that require the attention of curator (from relatively straightforward regular expression searching for telephone and credit card numbers through to more advanced text and image mining); (iii) curatorial examination and annotation; (iv) solicit and address promptly the views and wishes of third party individuals; and (v) inform and continually remind researchers of their legal obligations.

It is suggested that highly transparent and simple guidelines regarding repository policy towards legal aspects need to be established. Inordinately complex and convoluted routes to compliance should be avoided. Widely applied and recognisable licenses with layered explanations and corresponding icons may be helpful.

Archival copies are essential but are currently made at some legal risk, a situation that serves to highlight one of the many quandaries facing the world's digital heritage[1]. Where a conservative interpretation of copyright seems appropriate, access to the digital object can be limited to a single onsite access point with all copying and printing functionalities securely disabled[2].

For archives or archival elements or research practices that are of a sensitive nature, an ethical committee attached to a repository may be sensible along with seminars on ethical research practice, specifically in the context of personal digital archives. For certain research purposes (eg epidemiological and lifestyle analysis) information could be anonymised by the repository for mediated access.

---

[1] There are recent indications that helpful changes may be underway in the UK to amend the copyright law, potentially enabling the functions of archives and libraries; see http://www.ipo.gov.uk/consult-gowers2.pdf

[2] This approach is one of those that have been adopted for the Access to eMSS subproject within the British Library, with access to a small selection of literary eMSS initiated in February 2010 in the Manuscripts Reading Room through a single machine

An important aspect of the ethics is an understanding of the potential benefits as well as the potential costs. At the same time there is a primary duty to care for individual rights especially with regard to highly personal matters.

The field of infoethics is potentially very wide ranging, and it is anticipated that specialist infoethical committees will expand in a manner akin to those that address diverse and intricate bioethical challenges and opportunities. These could provide up-to-date guidance for repositories and curators (and others) in, for example, identifying universal distinguishing features where personal digital objects should be handled in a particular way; and advocate any necessary legal amendments and technical processes that support these procedures[3].

>> *High volumes and archive duality*

The great opportunity offered by digital technology is that of a greater breadth and depth to the holding of personal archives.

The continuing existence of hybrid personal archives with both analogue and digital objects presents a dilemma since space is much more limiting for paper, meaning that it is possible to hold, preserve and make available digital objects on a greater scale. Important decisions therefore circle around the extent to which individual repositories will be constrained by analogue space, bearing in mind that the quantities of paper in personal archives are often as great if not greater than ever. Will some repositories opt to collect both analogue and digital allowing a greater intake of digital content, or even to collect the digital components of personal archives alone? Will some repositories choose to embrace both digital and analogue, while others opt to specialise in one or the other? One determinant will be the success of fast digitisation and of systems for accessing the digitised surrogates.

>> *Archives in the wild, archives in concert*

It is suggested that the community of repositories as a whole endeavour to ensure that at least some elements of the repository network take into their digital vaults a sample of personal archives from people generally in order to help provide detailed snapshots of everyday digital life and private lives. While it remains impractical for even the largest repositories to be able to extend this goal to all individuals at large, it is conceivable that everyone is encouraged and supported in looking after their archives in the wild: for themselves and their families and friends - indeed this can be seen as an essential element of furthering digital literacy and inclusiveness, both nationally and internationally (and including the developing world), and would concur with the outlook of the Digital Britain programme.

A key need will be an advanced form of archival personal information management geared to the whole lifecycle that makes it possible for people to grow sustainable personal archives based on continuing use and reuse of their contents and good digital preservation practices.

Reflection on the nature of digital living in the future suggests that even those individuals who do deposit their personal archive in a repository will want to retain copies of the digital archive for themselves and their families. In the immediate future it is likely that individuals will deposit the personal digital archive incrementally during the life of the individual, as a

[3] In the UK, would the Office of the Information Commissioner provide an existing platform, to be furthered with the extensive involvement of human rights and privacy organisations, information and library organisations, scientific bodies such as the Royal Society, and humanities organisations such as the British Academy? In any case, international understanding and agreement is also essential

living, growing archive, and indeed this is already happening for prominent scientists and writers. For everyone who is digital, personal information holdings will be essential.

Repositories may serve and engage with archives in the wild (i) by aiding digital preservation, capture and conservation (eg migration, emulation, planning, recovery); (ii) by serving as guardians of the authenticity of the originals, and providing registries for digital objects in the wild; (iii) by providing services that strengthen and assist the effectiveness of the collections of personal digital objects; (iv) by mediating careful and accountable access (eg protecting individuals' privacy according to the wishes of the participant while enabling monitored research); and (v) by making available more powerful research and use techniques and technologies (eg providing access to the latest tools that would be too expensive for the individual researcher).

Archival repository institutions might also conduct their own research into archives in the wild, or do so in collaboration, capturing content according to research themes and longterm trends, and thereby gaining and offering expertise in the content and nature of the digital universe, especially the more personal aspects that are not otherwise available, with discussions and interpretations of findings, suitably anonymised as necessary.

>> *Strategy and research*

Five strategic areas of activity are (i) research and tools; (ii) professional collaboration and integration; (iii) collections and participation; (iv) promotion of skills; and (v) public and institutional engagement.

It is essential that digital archiving is underpinned by good research. Key topics for the immediate future are: (i) authenticity and forensics; (ii) sustainable personal information management incorporating personal digital preservation and enriched digital utility; (iii) usability and personalisation; (iv) infoethics, digital rights and digital value; (v) advanced cataloguing and context; (vi) emerging research techniques, visualisation, networks and metrics; (vii) evolutionary dynamics, phylogenetics and stemmatics; (viii) adaptive curatorial systems for complex archives; (ix) history of personal information technologies; and (x) review and analysis of the use of personal archives by researchers, as well as an understanding and appreciation of the content and nature of archives in the wild.

Collaborations and agreed standards are essential in order to take full advantage of the networking potential of digital objects. Participation of creators, curators and consumers acting together could enable and greatly enrich the collection and description of personal archives, in significant volumes.

With more and more archival repositories beginning to receive personal computer media, there is an urgent need for a network of training programmes directed at multitasking archivists.

In the near future the single most pressing requirement is that of advocacy at the highest levels: advocacy that seeks to improve and elevate understanding of the enormous potential of personal archives. It is hoped that this synthesis goes some way towards this objective.

Aims of the Initial Synthesis

(1) This synthesis aims to provide an informal and reasonably detailed overview of the project. An attempt has been made to allow each chapter to stand more or less on its own. This has led inevitably to some repetition.

(2) The project has directed much of its effort towards producing a series of formal and - as far as possible - peer-reviewed research publications. The process of publication continues. The emerging publications represent the considered output of the project, and should be consulted for details and fuller analysis and discussion. Please see Chapter 15 for an outline of publication plans and a list of existing publications. The present report is not fully referenced although it does contain a bibliography and footnote references.

(3) This account does not attempt to report all of the research. The aim is to characterise the boundaries of the research, to provide some of the research findings, to highlight some emerging recommendations, and to consider some future steps. Perhaps as much as anything the synthesis sketches the emerging landscape of personal curation, ranging widely across topics.

(4) This synthesis represents some of the output of a research project, and so it does not necessarily reflect the official policy of the British Library. Equally it is not a formal report to the Research Council.

(5) It is directed at a wide audience.

(6) It is intended to be an optimistic and hopefully readable account. It has tried to show that there are ways forward, that genuinely useful and applicable tools are emerging, that pragmatic insights into legal and ethical quandaries can be found, that there are crucial roles to be played by traditional as well as specialist digital archivists, that researchers, humanities scholars and scientists all have important contributions to make to the future of personal archives.

(7) Everyone is encouraged to maintain a personal archive.

Authors, Team Members, Contributors

Authors
Jeremy Leighton John, Principal Investigator
Ian Rowlands, Coinvestigator
Peter Williams, Senior Researcher
Katrina Dean, Coinvestigator

Team Members
Jamie Andrews
Andrew Charlesworth
Alison Hill
Kristian Jensen
Rory McLeod
David Nicholas
Robert Perks
John Tuck
Paul Wheatley
Lynn Young

Initiator of the Project
Neil Beagrie

External Advisors
Sheila Anderson
Professor Dame Wendy Hall
Christian Lindholm
Clifford Lynch
David Thomas
Susan Thomas
Professor Jonathan Zittrain

Budget
Elfrida Roberts
Martin Reagan
Philip Michel

Workshops and Conference
Gareth Burfoot
Alison Faraday
Jennie Patrice

Conference Design
John Overeem

Website
Rob Ainsley
Adrian Arthur
Andrew MacCalman
Adrian Turner
Colin Wight

Video
Matt Casswell
Lois Froud

CONTENTS

Chapter 1: Introduction

1.1 Background

(1) For centuries and indeed millennia individuals have used physical artefacts as devices for personal memory and reference. In recent centuries these material objects of symbolic storage[4] have manifested themselves as personal journals and correspondence, draft compositions, letter and journal writing, note taking, travel sketches, accounts of an individual's informal reading, bibliographic records, appointment diaries, financial receipts, bank statements and personal compilations of books, serials, clippings, offprints and observational data, all in close association with individuals.

(2) In the 20th century, these 'memory references' embraced even more materials, from photographs, sound recordings and ciné films and home videos through to airline tickets and telephone bills.

(3) With people leading increasingly digital lives in the 21st Century, creating, acquiring and sharing highly diverse and numerous personal digital objects (or eMANUSCRIPTS), there is a critical need to understand the implications both for archival repositories and for individuals themselves. Elements of the personal digital archive - the digital equivalent of 'personal papers' - include word-processed documents, emails, blogs, digital photos and audio and video, websites restricted to family and friends, and even the output of twittering and texting.

(4) Aspects of these personal archives and collections are often (i) of profound importance to individuals and to their descendants, and (ii) of immense value to research in a broad range of arts and humanities such as literary criticism, biography and history as well as in the social, human and natural sciences.

(5) Personal archives support histories of literary, scientific, political and socioeconomic practice by capturing creative processes, witnessing shared historic events, documenting individual activities, personal thinking and sentiment, and revealing social and professional networks, not least in the production and dissemination of artistic and scientific innovation, knowledge and understanding.

(6) While offering researchers nuanced contexts for understanding wider scientific, literary and cultural developments, personal archives provide individuals and descendants with ancestral history, family memory and a personal sense of origin.

(7) Personal collections of scholarly and research significance vary in the extent to which they are structured, ranging from a 'freestyle' nature at the organic end of the spectrum through to a more systematic nature - though still not a rigidly structured one - at the other end of the spectrum.

(8) Four examples illustrate the variety of sources of personal information derived from the individuals themselves:

---

[4] F. d'Errico, C. Henshilwood, G. Lawson, M. Vanhaeren, A.-M. Tillier, M. Soressi, F. Bresson, B. Maureille, A. Nowell, J. Lakarra, L. Backwell and M. Julien (2003) Archaeological evidence for the emergence of language, symbolism, and music - an alternative multidisciplinary perspective, Journal of World Prehistory 17:1-70

- archives of influential and eminent individuals (for example the correspondence and other papers of literary figures such as Arthur Conan Doyle or of celebrated scientists such as Alexander Fleming);
- compilations and collations of naturalistic and social observations extracted from diverse and widely distributed personal archives (eg observations of meteorological, astronomical and geological events and of botanical and zoological properties and occurrences);
- contributions of large numbers of individuals systematically approached by researchers (for example the diaries collated in the Mass Observation Archive and the 6,000 oral history interviews in the Millennium Memory Bank);
- documented responses to interviews and questionnaires arising from longitudinal studies of individuals belonging to a cohort as with the surveys of the National Child Development Study.

1.2 Organisation

(1) The mission of the Digital Lives research project is to help (i) to enable personal digital archives to attain their far-reaching research potential, (ii) to understand the way both academics and people generally engage with and use the personal computer, (iii) to progress the capture, holding and use of personal archives throughout an individual's life, and (iv) to impart to the individual the option of passing the personal archive to a repository or to other individuals such as family members.

(2) The Digital Lives research project brought together archivists and curators of diverse expertise, digital specialists and academic researchers from the British Library, the Department of Information Studies at University College London, and the Centre for Information Technology and Law at the University of Bristol.

(3) The project explored the whole archival lifecycle, from the behaviour of creators (writers, scientists and others), the expectations of researchers and users generally, and the perceptions of curators and archivists. It examined legal and ethical issues; identified suitable technologies for securing and making available the personal digital archives of both academics and members of the digital public in the future; and considered the increasing significance of cloud computing[5] for personal archives.

(4) The research was structured as five components directed at: (i) characterising the personal information behaviour of the creator (component 1), (ii) outlining the legal and ethical environment (component 2), (iii) documenting the opinions and experiences of users, researchers (component 3), (iv) exploring the technical capabilities emerging (component 4),

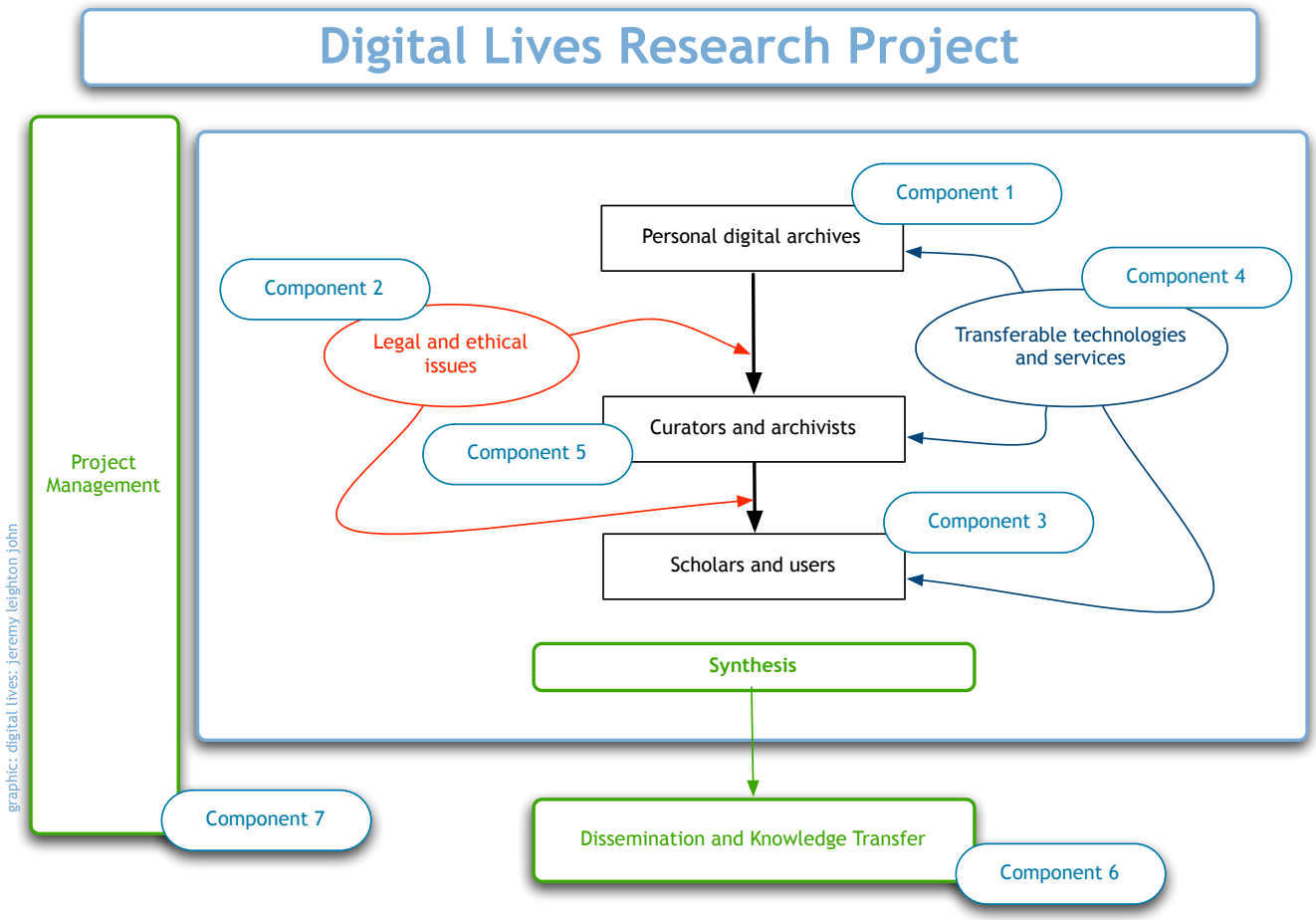
[5] With the term 'cloud computing' still evolving, it may be a relief to find a definition from NIST, National Institute of Standards and Technology, which confirms that cloud computing embraces a variety of service and deployment models: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models", P. Mell and T. Grance (2009) The NIST Definition of Cloud Computing, Version 15, 7 October 2009, 3 pp; http://csrc.nist.gov/groups/SNS/cloud-computing/. The five characteristics are: on-demand self-service; broad network access; resource pooling; rapid elasticity; and measured service. In short, the user uses as much or as little as is needed and pays for what is used.

and (v) feeding in the professional observations of curators and archivists (component 5).

(5) Methods employed included reviews of research papers, personal interviews (face-to-face and over the internet), online surveys, email questionnaires, requests for public participation *via* email, a series of workshops and a three-day conference.

(6) During the course of the research the emphasis changed in four ways: (i) the digital public was examined along with the academics; (ii) the scientific uses of personal archives were contemplated alongside the uses by humanities scholars; (iii) an archival perspective was firmly embraced; and (iv) the goal of publications was reinforced.

(7) The initial synthesis document is structured accordingly with Chapters 3, 4, 5, 6 and 7 reporting on the components 1, 2, 3, 4 and 5 respectively. Chapters 8, 9, 10, 11 and 12 provide some synthesis and overview. Chapter 13 provides a bibliography of example references. Chapters 14 and 15 conclude with some details of the achievements to date of the Digital Lives research project.



Chapter 2: Terminology and Research Context

(1) Archives of 'personal papers' contain letters, notebooks, diaries, draft essays, family photographs and travel ciné films; and in 2000 the British Library adopted the term eMANUSCRIPTS (eMSS) for personal digital objects, the digital equivalent of these 'personal papers', having begun accepting diverse computer media as part of its manuscript holdings[6].

(2) Personal computer media include punched cards, paper tapes, magnetic tapes, program cards, floppy disks of several sizes (8", 5.25", 3.5" and 3"), zip disks, optical disks (eg CDRs and DVDRs) and various hard drives, both internal and external. The files that might be contained therein are of a diverse range: word documents, PDFs, digital photos and graphical images, spreadsheets, travel video, audio recordings, emails with miscellaneous attachments, and so on.

(3) The eMSS that make up a personal digital archive are typically born digital but individuals are increasingly digitising their analogue personal objects - from paper letters to tape recordings and ciné film. These are also eMSS and are accordingly part of the personal digital archive.

(4) A distinction is drawn between (on the one hand) this personal digitisation carried out by individuals and families and (on the other hand) the digitisation performed by institutions. The term 'digital manuscripts' is used at the British Library in the context of personal archives to embrace both eMSS and the digitised manuscripts that arise from institutional digitisation within the library or repository.

(5) The personal library (personally accumulated publications) can be separated conceptually from the personal archive. Both components are of interest but the project is centred primarily around the personal archive. The distinction becomes less clear when publications have been annotated by the individual. Together with physical artefacts (ornaments, paintings) the library and archive make up an individual's personal collections.

(6) The precise boundaries and definition of a personal digital archive and indeed any personal archive are subject to opinion, and changing opinion too. A personal archive may contain links to online publications but these remote objects would not be considered part of the archive by the project: unless the objects are owned by the individual (eg on a remote but personal server space) the links have more in common with a personal list of bibliographic references.

(7) However, any digital objects and content of an individual that are online (eg a social networking account that is restricted to family and friends) would be regarded as a component of the individual's personal archive. In general if the personal digital object or eMANUSCRIPT is in some sense possessed by an individual it would be deemed by the project to be part of the personal archive (or personal library). Clearly, these considerations are of crucial relevance to curators taking individual archives into the holdings of longterm repositories.

---

[6] Among the first papers on the subject of personal digital archives to emerge from the archival community is that of A. Cunningham (1994) The archival management of personal records in electronic form: some suggestions, Archives and Manuscripts 22(1): 94-105

(8) Sometimes the word ‘archive’ is used by people in the context of computers as a verb, ‘to archive’, essentially meaning to set aside in a store (from which an item will occasionally be retrieved). For many professional archivists, an archive is a highly dynamic entity and its curation concerns the whole lifecycle from appraisal and acquisition through to access, resource discovery, reuse and dissemination[7]. Even in the case where the originator of an archive has long since deceased, material continues to be added to that person’s archive within a repository (eg a series of letters that the originator wrote to a friend) and curators may create new material (perhaps by recording an interview with that same friend of the originator).

(9) To highlight this dynamic aspect of a personal archive, it may be referred to as a ‘living archive’. The phrase is most apt perhaps when it reflects an ongoing relationship between a repository and the individual, *with the archive remaining in active use by the originator*.

(10) The personal digital archive can be placed in the context of a longstanding and accomplished heritage of personal archives and historical and literary manuscripts, while promoting an enhanced curation that adopts digital technologies for capturing rich contextual and complementary information in the form of video conversations, landscape panoramic photography and graphical images of the environment alongside the firmly established culture of oral history.

(11) The phrase ‘archives in the wild’ refers to the personal digital archives that exist outside an official longterm repository, including the personal archives of academics, writers and members of the digital public.

(12) There appear to be four key factors contributing to the increase in numbers and quantities of personal digital archives: (i) the growth of digital storage capacity; (ii) the availability of content-creating and sharing tools and devices; (iii) the accessibility of information on the world wide web and other networks; and (iv) the creative and social instincts of people.

(13) It was estimated in one report[8] by the International Data Corporation (IDC) that the quantities of information that are attributable to individuals would reach nearly 70% of the digital universe by 2010: not all of it being expected to find its way into a personal digital archive. A subsequent report[9] by IDC noted that some of the burgeoning information is directly created by individuals (eg user-created content, photos taken, YouTube videos uploaded, emails sent, VoIP phone calls) while some of it emanates indirectly from their actions and presence in the form of a ‘digital shadow’ (eg records in Amazon book store, eBay reputation metrics, banking, airline, telephone and heath databases, surveillance images and

---

[7] It is important, too, to be aware of the entirely different ethos of curating the archive of an individual compared with an archive of an organisation - such as a governmental archive - that addresses the assets of a ‘designated community’ and may be underpinned by legal obligations with practices following community regulations. “Working with the creators of personal archives is entirely different: it entails working with a host of diverse people, cultures, and systems. We collect material which individuals have no obligation to give us; we cannot impose standards...”, S. Thomas and J. Martin (2006) Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives, Journal of the Society of Archivists 27(1): 29-56

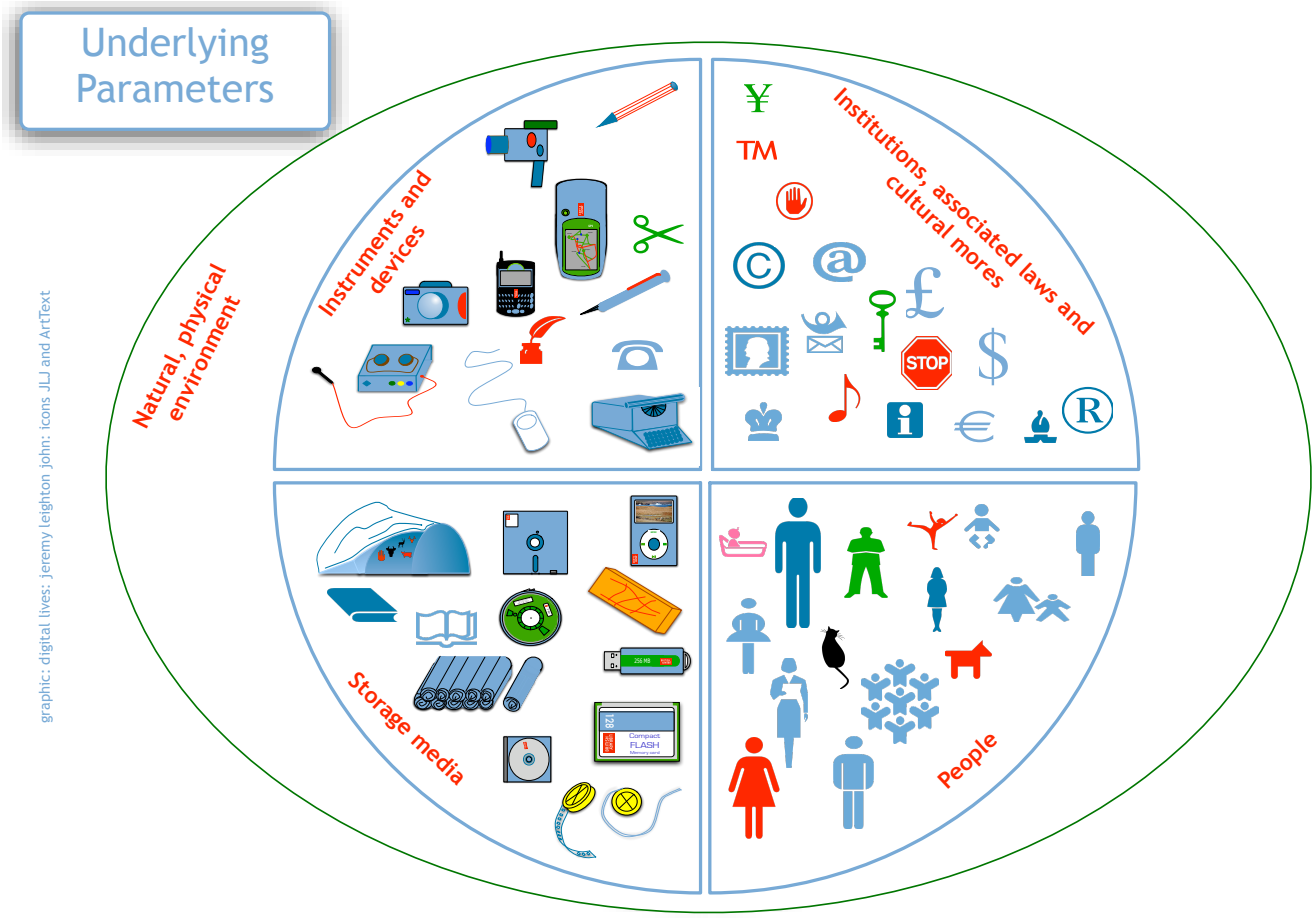
[8] J. F. Gantz (2007) The expanding digital universe. A forecast of worldwide information growth through 2010, an IDC White Paper sponsored by EMC, Framingham, Massachusetts

[9] J. F. Gantz (2008) The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011, an IDC White Paper sponsored by EMC, Framingham, Massachusetts

hospital scans).

(14) The personal digital archive is a multidisciplinary resource: for individuals themselves, for biographers, for historians and socio-political and economic researchers, for literary scholars and for scientists, medical, social and natural. Its increasing value will mean that there will be potentially far more information retained than ever before for scholarship, historical research and scientific analysis alike.

(15) Five key parameters relevant to the personal archive: (i) information storage media; (ii) devices and instruments for creating and transferring information; (iii) people and social networks; (iv) legal, ethical, cultural, and economic environment; and (v) the natural and physical environment.



Chapter 3: Organising, Storing and Finding Personal Information

3.1 Aims

The aim of this component of the project was to gain a better understanding of the way in which individuals behave in creating, acquiring, sharing, storing and retrieving their information, and in actively or passively building up and organising their personal digital archives or collections.

The key concept is that of *personal information management* (PIM). There is a large body of research literature about how information and knowledge resources are managed in corporate environments, and yet surprisingly little is available about how individuals manage their own information assets – especially over a long time. This critical gap in our understanding needs to be addressed in order to devise curatorial strategies and practices that can deal effectively with ever-expanding personal collections of digital objects.

This element of the project is directed at the creator, the originator of the personal digital archive. It concerns the digital behaviours of individuals, and the development and safeguarding of individuals' digital archives.

3.2 Methods

Three techniques were adopted: (i) a series of face-to-face interviews with creators were conducted by researchers and curators; (ii) a review of the literature was undertaken; and (iii) a questionnaire was subsequently developed and made available in the form of two online surveys directed at academics and the digital public, respectively.

3.3 Interviews[10]

Interview preliminaries

(1) In all 25 interviewees participated in this research and could be informally classified either as 'established' or 'high profile' individuals whose lives and works would already be of interest to repositories, or as 'emerging' individuals whose lives might be of specific interest in the future.

(2) 'Established' individuals comprised: an architect, authors (including web author and playwright), a crystallographer, a geophysicist, a knowledge management specialist, a molecular biologist, a photographer, a politician, and a web designer.

(3) 'Emerging' individuals comprised: a digital artist, lecturers in cultural studies, education, and in participatory media, a theatre director, and a music publisher, along with postdoctoral researchers and PhD students in the fields of archaeology, cultural studies, English literature, human-computer interfaces, information science, and psychology.

(4) Key questions directed at the interviewees related to:

- how and why people use computers and other information and communication technologies;

[10] P. Williams, K. Dean, I. Rowlands and J. L. John (2008) Digital lives: report of interviews with the creators of personal digital collections, Ariadne 55, April 2008

- the manner in which these activities are yielding a personal digital archive or collection;

- the way individuals learned to use various software and hardware systems, and the extent of any training; and

- the integration of analogue elements of personal archives (eg paper letters, diaries and notebooks) and their digital counterparts.

Interview findings

(1) The interviews clearly demonstrated the existence of highly diverse relationships between individuals, their computer and communication systems, and their personal digital objects or eMSS.

(2) Most people seem to be self-taught in their use of computer software and hardware, and where this was not entirely so training had often been informal and *ad hoc*.

(3) Many misconceptions about digital technologies were apparent, most significantly – from the archival point of view – in the realm of backing up, storage and longterm preservation.

(4) Personal digital archives or collections appear to grow organically, with new files and folders being created alongside existing ones and with little appetite for systematic and regular deletion in many individuals.

(5) On the other hand, reliance purely on search functionality was rare at this time, and most interviewees made some effort at filing personal digital objects in folders in various ways, most especially based on chronology or topic.

(6) The hierarchical folder system was widely perceived as information in its own right.

(7) There were some interesting findings with regard to the naming of files and folders and the management of versions: most especially relating to inconvenient display, misplaced significance attached to minor changes[11], and a mistrust of the automated naming and dating system.

(8) A fairly universal distinction was seen in the separation of personal digital objects according to leisure and work, even though the same computer space tends to be used for both activities by an individual. Family members frequently had separate computers or computer spaces. The creation and adoption of several distinct email accounts by an individual is also pertinent in this context, each account being used to serve different roles or conveniences.

(9) It was evident from the interviews that academics are embracing cloud computing more and more. This was matched by a general – and a somewhat unanticipated - desire to keep up with advancing technology. There was also – as has been reported by other researchers - a

[11] For instance: an automated change of the date of a file attributed to a modification that has little or no creative relevance

demonstration of the capacity of individuals to appropriate technologies in innovative ways, manifested most strikingly in the case of email[12].

(10) The reluctance to dispose of files and email messages that are no longer of immediate use was in part explained by the effort required in actively selecting and deleting particular objects. Moreover, the future value of an object is often uncertain. Retention therefore does not necessarily imply that an object has been invested with significant value. Nonetheless there was plenty of evidence that interviewees do perceive a future use for many of their personal digital objects.

(11) Hard copy printouts of documents continue to be retained by interviewees with or without the corresponding eMSS; significantly, the stored analogue and digital objects did not always represent the same versions, emphasising the need for analogue and digital to be carefully integrated when reaching a repository.

(12) Personal digital objects stored for the future may be valued for their ability (i) to serve as reference information, (ii) to provide a source of creative work that can be reused, (iii) to evoke personal memories and context, (iv) to promote self-esteem, (v) to meet sentimental and memorial needs, and (vi) to witness an individual's past effort and creativity.

3.4 Perspectives Emerging from a Literature Review[13]

Towards an archival personal information management

(1) Much of the literature on personal information management is motivated by the desire to produce successful software and hardware that meet very specific functional objectives; and in many cases is commercially driven.

(2) Personal information management from the perspective of personal digital archives is a surprisingly under researched area. The project has proposed a model that adopts the lifecycle approach to personal archives of professional archivists, and has transferred it to the context of personal information management. It looks at the personal archives of individuals away from any repository (archives in the wild) and draws a parallel with the activities within a repository.

(3) It is a holistic approach where the management of information is considered from the perspective of a 'lifecycle'. This entails examining the stages of the life histories of eMSS from creation, acquisition and manipulation continuing with retention and organisation with backing up and disposal through to longterm storage and accommodation. These activities mirror the formal activities of archiving but correspondingly are conducted by individuals during their lifetime rather than by institutions.

(4) The personal archive of a living person is, of course, a dynamic entity: a 'living archive' with new objects being created, others being acquired, amended, and discarded. Correspondingly, professional curators of contemporary personal archives held in

---

[12] This concurs with other reports such as V. Bellotti, N. Ducheneaut, M. A. Howard and I. E. Smith (2003) Taking e-mail to task: the design and evaluation of a task management centered e-mail tool, ACM Conference on Human Factors in Computing Systems (CHI 2003), Fort Lauderdale, Florida, 5-10 April 2003, pp 345-352

[13] This section of the chapter is based almost entirely on: P. Williams, J. L. John and I. Rowland[s] (2009) The personal curation of digital objects: a lifecycle approach, Aslib Proceedings New Information Perspectives 61(4): 340-363

public repositories are engaging in enhanced curation activities that yield new digital objects such as panoramic images that complement the original contents of an archive. This perspective goes some way towards unifying the activities of individuals and repositories[14]. A key element of the personal lifecycle is deletion, and there is a parallel within the repository lifecycle known slightly euphemistically as 'deaccessioning'.

(5) Much research into personal information management has sought to improve or understand functionality based on present conditions; but an archivally-oriented personal information management is concerned not only with the use of currently active digital objects but also with the retrieval of digital objects that were created, acquired, amended or organised in the past, and subsequently put aside.

(6) In seeking to use and reuse accurate information from the past, an archivally-oriented form of personal information management has objectives that are more akin to those of the professional archivist or curator; for example, literary curators seek to meet the requirements of literary scholars who want to study the way a piece of writing was created, while a historian of science might want to understand the sequence of events and insights that led to a discovery or new theoretical perspective and a curator of scientific manuscripts takes this into account. The desire from the archival perspective is - as far as is practical - for information to be interpretable, authentic and maximally useful for future as well as current generations and immediate requirements.

(7) The great diversity of practices by individuals can be contrasted with the much more consistent practices that exist when documents are passed to a repository, where a more formal and uniform system applies although repositories do vary among themselves. Diversity in the way individuals manage their archives is to be expected but consistency in specific crucial areas such as interoperability would be beneficial and is to be encouraged.

(8) On the other hand, research into usability of devices or software applications has employed aspects of an individual's past behaviour and activity (ie elements of personal information) as a means of tuning the product so that it matches the individual, thereby improving its performance for the person.

(9) There is a recognition in the usability field that different individuals do things in different ways, and that the same individual may do different things at different times. Institutions might benefit from following suit, and in particular consider allowing people to work in different ways even when working towards the same end. In designing curatorial and access systems institutional repositories could aim to be receptive to individual variations among their users and even their curators.

Personal information lifecycles: from obtaining to holding and using

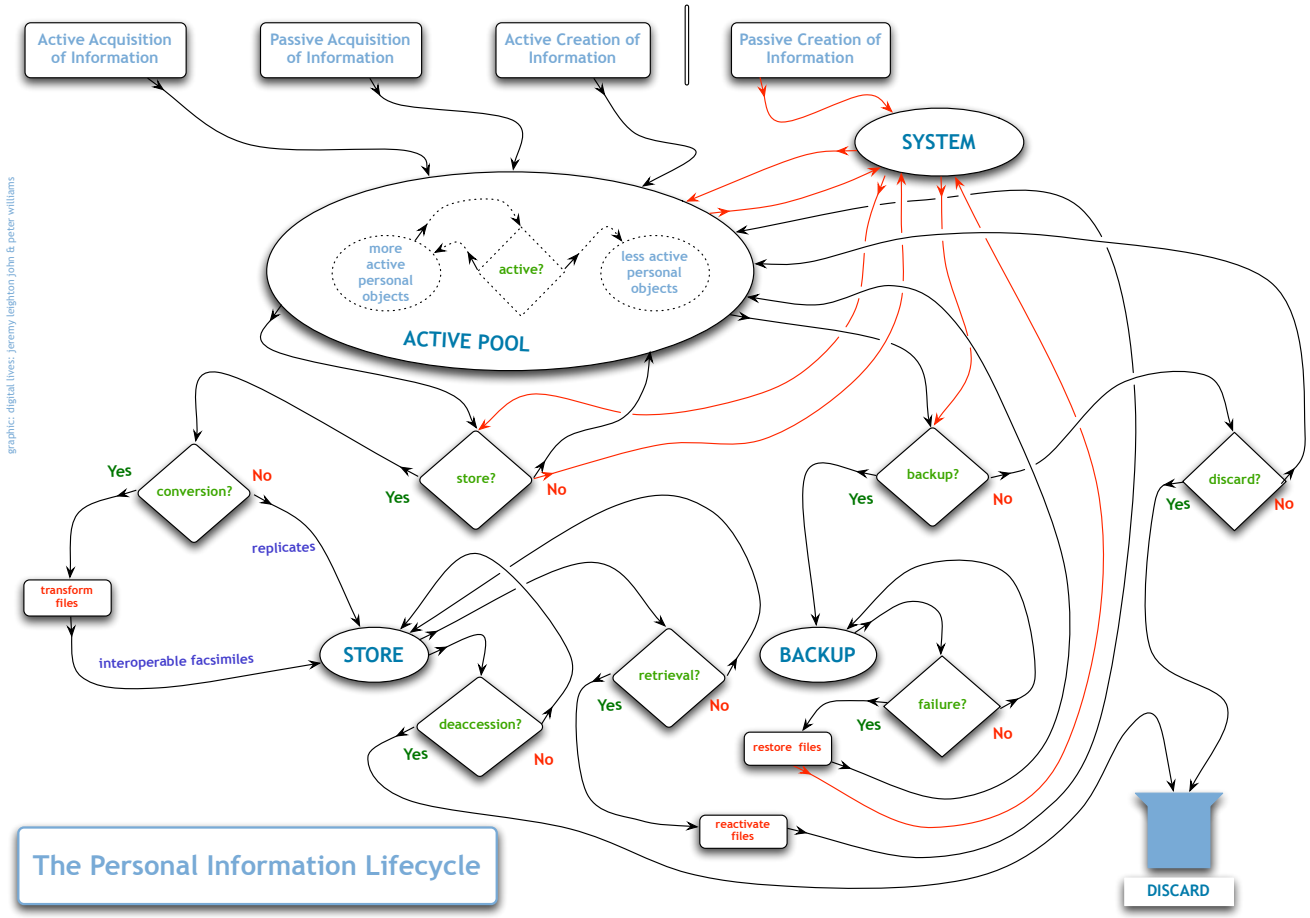
(1) The personal digital object cycle (or eMANUSCRIPT cycle) consists of three principal phases: (i) obtaining information; (ii) short term information use and management; and (iii) longterm information preservation, use and reuse.

(2) There are four ways in which individuals obtain and build up information: (i) actively sought acquisition of information (eg searching for and downloading a PDF from the web); (ii)

[14] It might be noted that individuals do not generally concern themselves with issues of digital rights, authenticity and provenance but if personal curation becomes more widely taken up these will be aspects that everyone will want or need to contemplate

passive acquisition (eg receiving and retaining a spreadsheet attachment with an email, or encountering desirable information unexpectedly); (iii) active creation (eg writing a novel, taking a photo); and (iv) passive creation (eg automatic creation of digital objects and metadata on your computer hard drive by software including web histories, caches, lists of recently opened files).

(3) The two passive phenomena are the least self-evident. In the case of a passive acquisition (and retention) there is an implicit identification of value (potential if not actual), although some objects may be retained because an active decision and deletion would be required as in the case of an email attachment. Passive creation reflects the fact that much personal information is created automatically on the computer. Much of this information would pertain to a personal digital archive, and some of it might be of interest to a historian or social scientist.



(4) On acquiring or creating information there is a decision to retain it or not. If later availability of the information is desired, it may be necessary to decide whether to keep the information itself or to leave it *in situ* on a website (for instance) and retain information about its existence and its whereabouts or about how to search for it again.

(5) Notwithstanding the increasing efficiency and effectiveness of search tools applied to a person's own files, it is clear that there is strong preference (seen in the literature and in the project's online surveys) for classifying files and for arranging folders in hierarchies often based – as the first classificatory rule - on the intended use or purpose of the personal digital

object[15]. The well known, frequently reinvented and not-entirely-appropriate use of email by many people as a means of organising files of diverse types was observed in the present study too. It points emphatically to a clear need for more sophisticated personal information management systems that individuals feel that they can trust, control and understand.

(6) Why would or do people keep files for the longterm? Why should they keep them? What should be kept? This is a complex area of research ranging as it does from personal psychology to professional necessity. Potential answers considered in the present project's online surveys (see §3.5) include: (i) for recording past activity or events, (ii) for sentimental reasons, personal memory, (iii) for witnessing creativity, (iv) for sharing with colleagues, (v) for leaving to a successor or posterity, and (vi) for future reference. A previous study of the personal files of 48 academics identified five categories: (i) finding it later, (ii) building a legacy, (iii) sharing resources, (iv) fears of loss, and (v) identity construction[16].

(7) The identification of potential usefulness is not trivial. Professional archivists and curators are trained or experienced in this aspect but even they make mistakes and can find it problematic.

(8) It is now possible to build up a digital archive without needing to delete anything, and so it should not be assumed that documents that are not needed (serving no foreseeable purpose) will actually be deleted. Even with paper personal archives, some people - 'hoarders' - are inclined to keep everything: at any rate far more than might be ostensibly needed or useful, but a powerful motivation can be the desire to store the objects 'just in case' they are needed.

(9) It is also possible to accumulate files that are never properly examined: unread and unprocessed data. Thus it is not simply a matter of discarding once-valuable information, it can be a matter of discarding information that was never used or evaluated[17].

(10) Information of widely varying degrees of potential usefulness is often stored in the same directory or medium. Some information is retained because it cannot be easily identified as useless or cannot be easily separated from the useful or likely-to-be-useful categories of objects.

(11) An automated prioritising of eMSS based on an individual's past preferences, values, selections and usages, would allow objects to be ranked according to importance with the total quantities retained for storage determined by the financial resources of the individual.
(12) The more careful the original capture of the information, fortified by contextual information, the more effective and far-reaching will be any improvements in reuse potential.

---

[15] For example: W. Jones, C. Munat and H. Bruce (2005) The Universal Labeler: plan the project and let your information follow, 6th Annual Meeting of the American Society for Information Science and Technology, Charlotte, North Carolina, http://kftf.ischool.washington.edu/docs/UL_ASIST05.pdf; W. Jones (2007) Personal information management, Annual Review of Information Science and Technology, edited by B. Cronin, Information Today, Medford, New Jersey, pp 453-504

[16] J. Kaye, J. Vetis, A. Avery, A. Dafoe, S. David, L. Onaga, I. Rosero and T. Pinch (2006) To have and to hold: exploring the personal archive, CHI Proceedings, Personal Information Management, Montreal, 22-27 April 2006, pp 275-285

[17] S. Whittaker and J. Hirschberg (2001) The character, value, and management of personal paper archives, ACM Transactions on Computer-Human Interaction 8(2): 150-170

3.5 Surveys[18]

Survey preliminaries

(1) The online surveys attempted to probe the following topics:

- levels of IT skills and competencies
- attitudes and behaviour towards computer security
- experiences of catastrophic data loss
- strategies for organising digital collections
- tactics for finding files
- approaches towards longterm preservation
- arrangements in case of death or incapacity
- reasons for archiving personal digital objects

(2) The design of the questionnaire was developed by members of University College London and curators at the British Library. It was extensively piloted before being hosted.

(3) Two surveys were carried out using essentially the same questionnaire: (i) one was directed at academics using SurveyMonkey Professional obtaining 1,507 complete responses of which 1,002 were academics; and (ii) one was directed at members of the digital public through the services of eDigital Research obtaining 1,961 completed responses of which 1,910 were classified as members of the digital public (51 being classified as academic).

(4) The same questions and the same structure were used for both surveys. The definition of an 'academic' is based on an 'opt-in' question in the questionnaire: "If your current role includes significant academic or scholarly interests, please indicate your main subject area" (followed by a choice of 18 disciplines).

(5) In the case of SurveyMonkey, links to the survey questionnaire were embedded in a number of websites, notably the British Library itself. In addition, emails were sent to a very wide range of university departments and other target organisations in an attempt to obtain a wide representation of different disciplines.

(6) The academic population includes postgraduate students, researchers and university teachers. The academic survey is not a formal stratified sample and the findings should not be taken to be representative of the whole Higher Education sector; however, a very broad mix of academic subjects and ages was achieved.

(7) The survey of the digital public was conducted on behalf of the project by eDigital Research, a market research company. eDigital has a number of consumer panels that it uses to conduct snap polls for various clients, mostly major high street brands. Members of the panels are people who have self-selected themselves because of an interest in electronic shopping. This sample is not fully representative of the UK public at large, but it is likely to be sufficiently representative of that segment of the population that is engaged with the computer and its applications in daily life.

(8) A series of analyses of these survey data are being conducted. In the first analyses - reported in the present document - two datasets were identified. The digital public dataset comprises those respondents of the eDigitalResearch survey that were not classified as

[18] This section represents the project's the first paper that reports the findings of the online surveys, and was prepared and written by Ian Rowlands, Peter Williams and Jeremy Leighton John

'academic' (1,910 in total). The academic dataset comprises 'academics' of the SurveyMonkey survey pooled with the 'academics' of the eDigitalResearch survey (1,002 plus 51 respectively, yielding 1,053 in total). Analysis was conducted using the Statistical Package for the Social Sciences (SPSS).

Survey results

(1) The purpose of these first analyses is to examine generally these two populations in their own right (in order to understand better how individuals manage their personal information, and the implications for curatorial practice), and to consider similarities and differences between our samples of academics and members of the general public.

(2) The statistical method of Chi-squared analysis has been used throughout for these analyses to highlight differences that are significant at the 5 per cent level of probability, and these are indicated by the symbol: *

Survey demographics

The following five tables provide a demographic view of the respondents from the two populations.

Table 1: Academic and digital public respondents

Column percentages

		n	*%*
Respondents	Academics	1053	35.5%
	Digital public	1910	64.5%
	Total	**2963**	**100.0%**

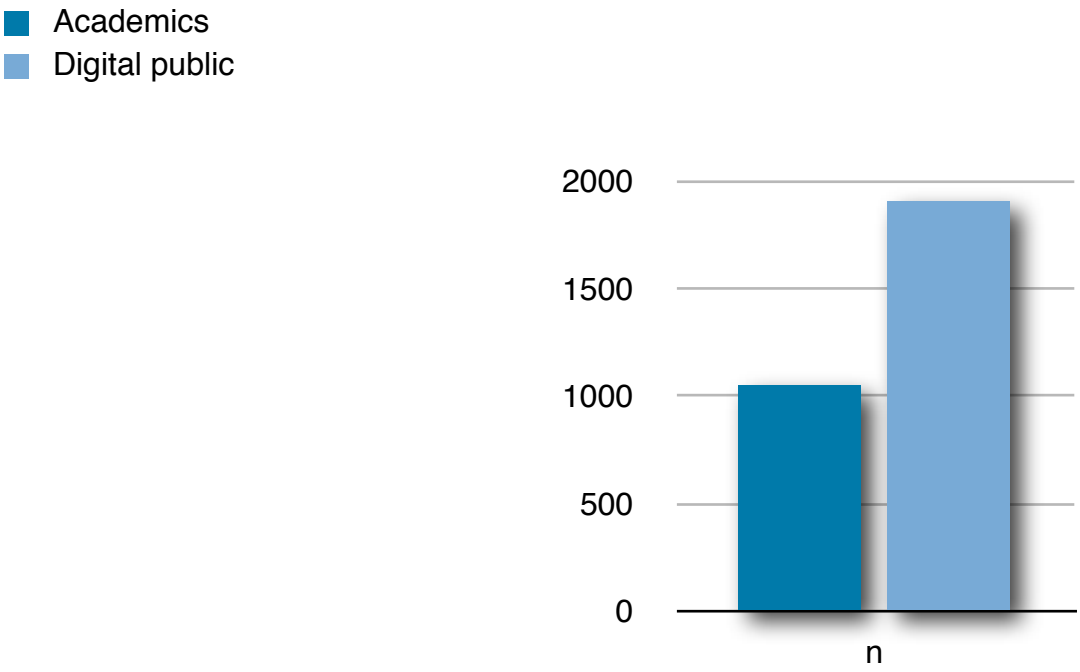


Table 2: Academics by broad subject area

Column percentages

		n	*%*
Academics by broad subject discipline	Arts and humanities	411	39.0%
	Computer & information sciences	155	14.7%
	Socio-economic sciences	197	18.7%
	Natural sciences	290	27.5%
	Total	**1053**	**100.0%**

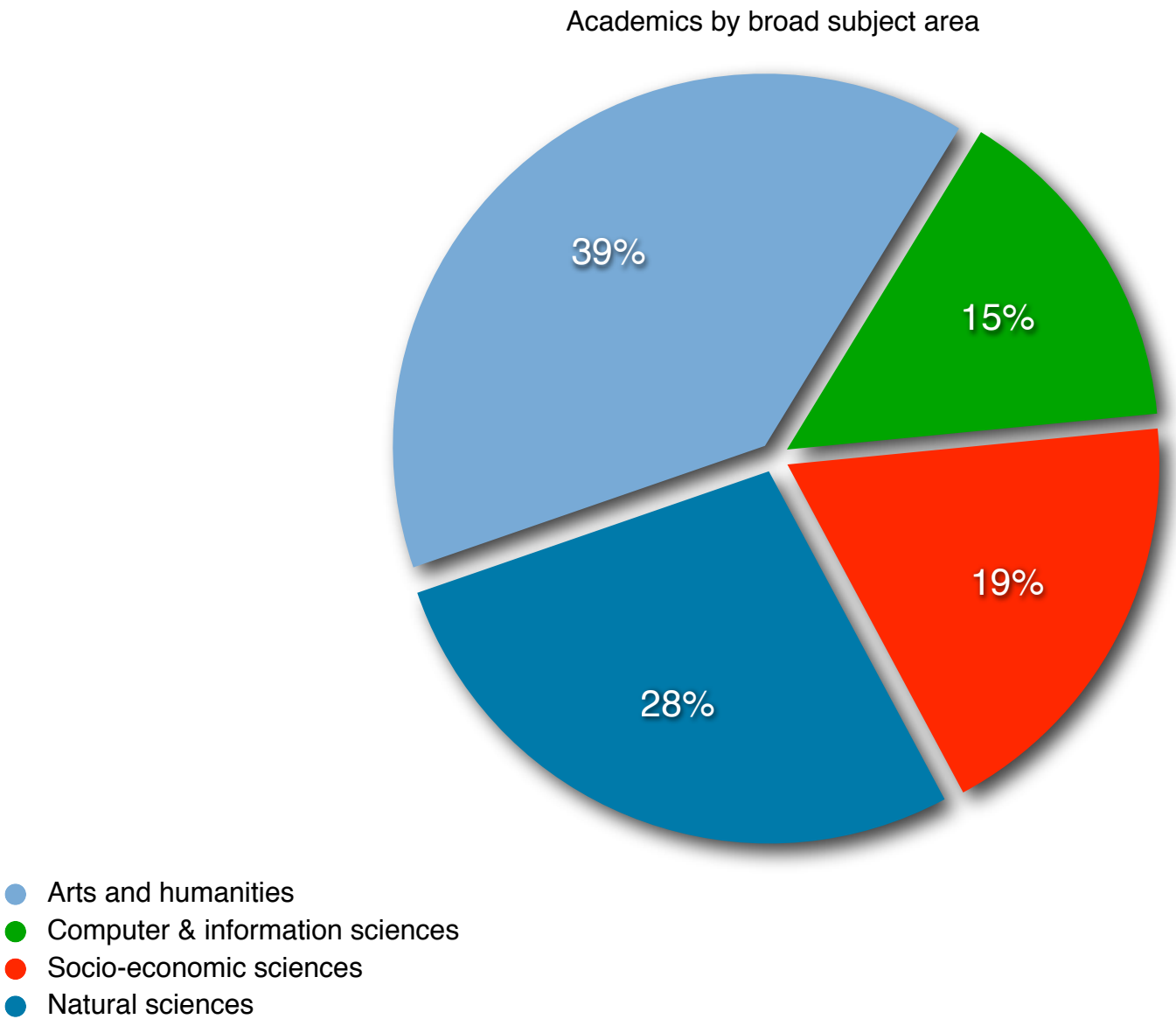


Table 3: Digital public by employment status

Column percentages

		n	*%*
What best describes your employment status?	Employed full time	893	46.8%
	Employed part time	206	10.8%
	Self employed	208	10.9%
	Full time student	41	2.1%
	Part time student	17	0.9%
	Retired	306	16.0%
	Looking for work	51	2.7%
	Not looking for work	188	9.8%
	Total	**1910**	**100.0%**

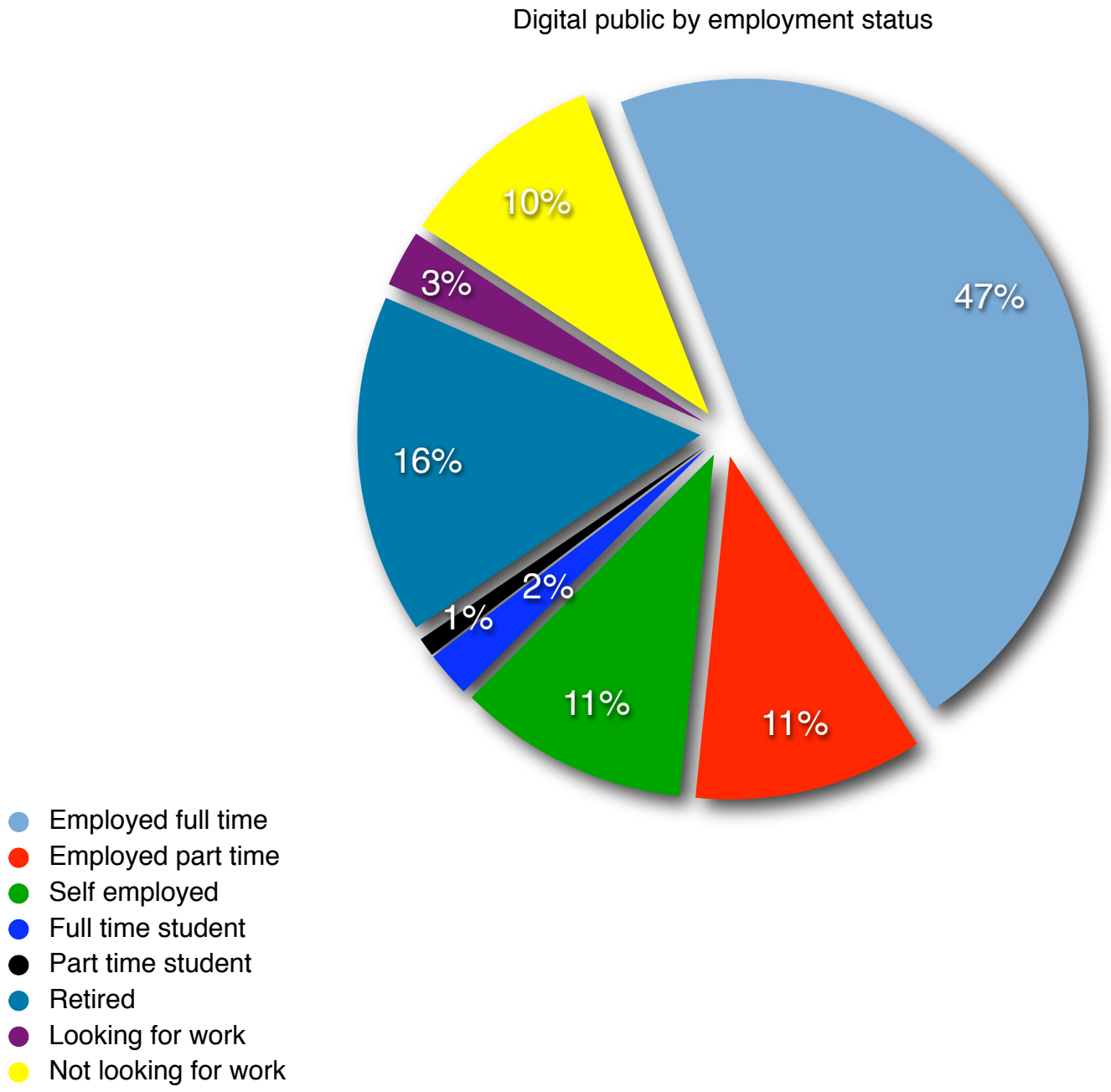
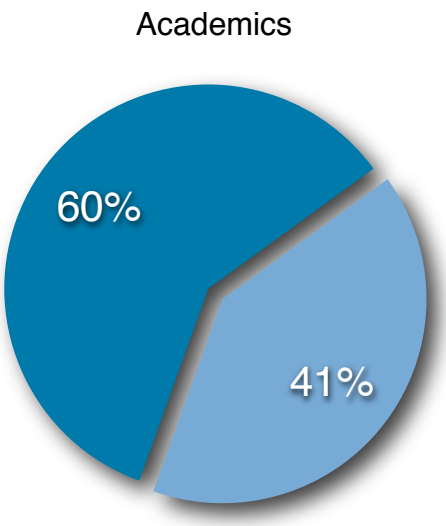
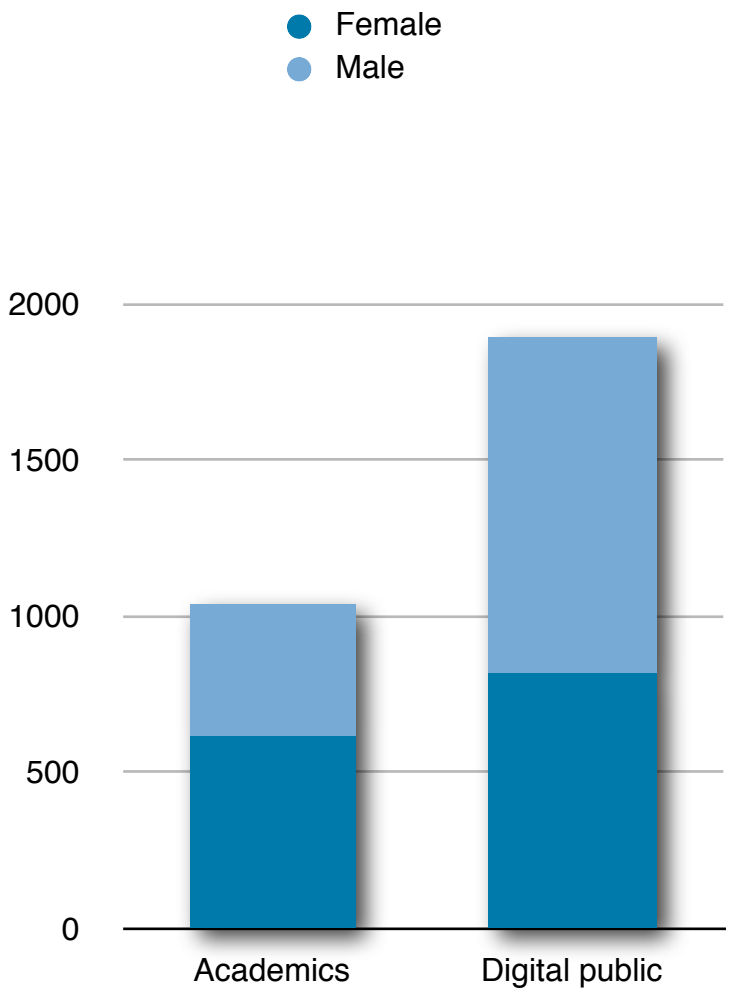


Table 4: Gender*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
What is your sex?	Female	619	59.5%	820	43.2%
	Male	422	40.5%	1076	56.8%
	Total	**1041**	**100.0%**	**1896**	**100.0%**



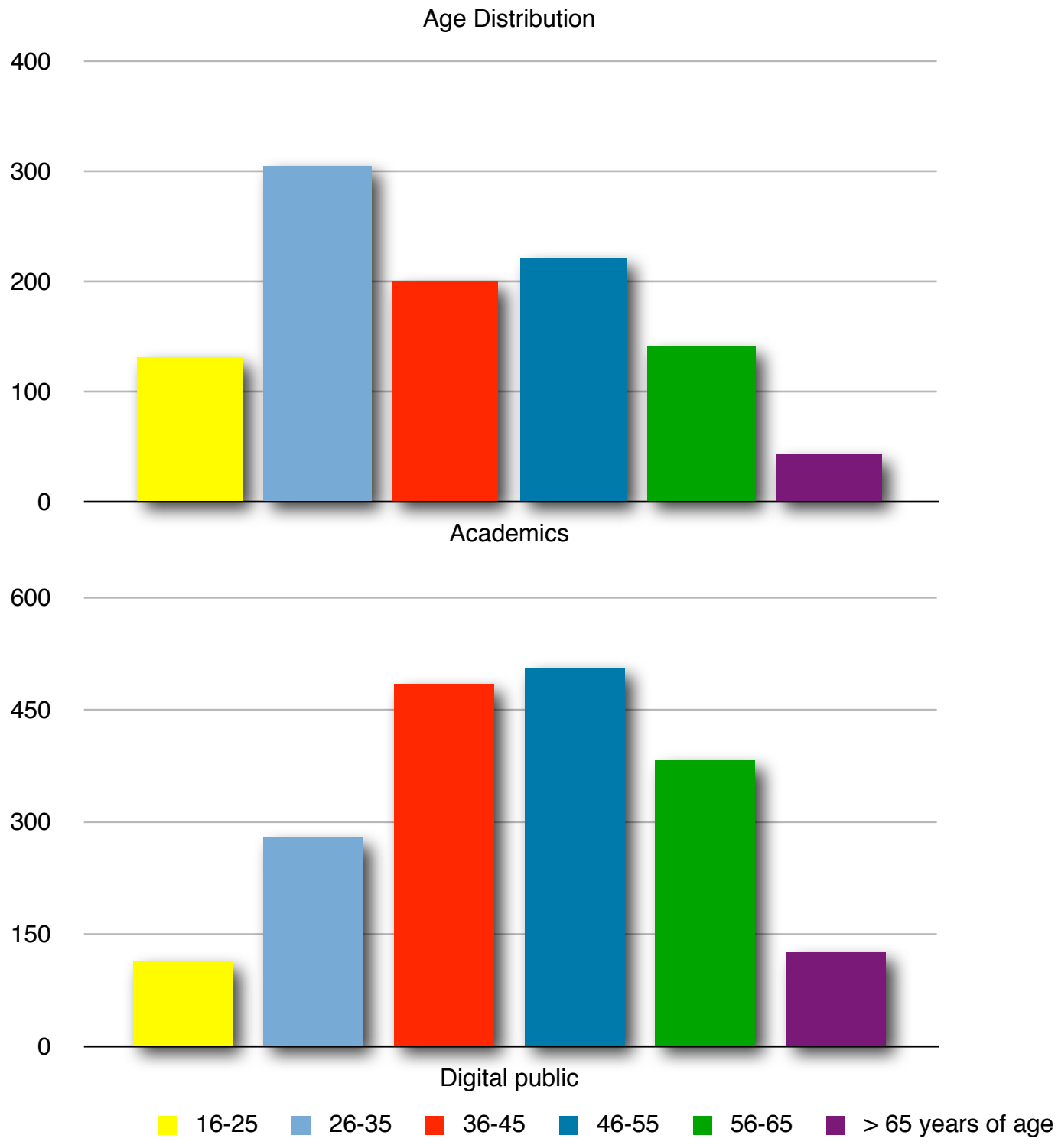


As already noted, both surveys are convenience samples rather than truly random snapshots of their respective wider populations. Women are represented more in the academic sample, which is partly but not entirely reflected by the larger than expected showing of arts and humanities disciplines. Men are predominant in the digital public sample, as a consequence of the panel sampling methodology used by eDigital Research.

Table 5: Age*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How old are you?	16-25	131	12.6%	114	6.0%
	26-35	305	29.3%	279	14.7%
	36-45	200	19.2%	484	25.6%
	46-55	221	21.2%	508	26.8%
	56-65	141	13.5%	382	20.2%
	>65	43	4.1%	126	6.7%
	Total	**1041**	**100.0%**	**1893**	**100.0%**



These tables are provided here for reference and as a framework within which to interpret the remaining findings. It should be noted that only 12.6% and 6.0%, respectively, of the academic and public respondents belong to the youngest age class, 16-25 years. Youth is slightly better represented in the academic sample. On the other hand, Table 6 shows that a greater percentage of academics have more than ten years of experience using computers and the self-perceived level of IT skills is marginally greater for academics too.

Respondents and their computers

The following five tables provide an overview of the relationship of participants with computer technology.

Table 6: Years of computer experience*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
How long have you been using computers?	<1 year	1	0.1%	6	0.3%
	1-2 years	0	0.0%	18	0.9%
	3-5 years	11	1.1%	114	6.0%
	6-10 years	119	11.4%	492	25.8%
	11-20 years	538	51.4%	798	41.8%
	21-30 years	308	29.4%	368	19.3%
	>30 years	69	6.6%	113	5.9%
	Total	**1046**	**100.0%**	**1909**	**100.0%**

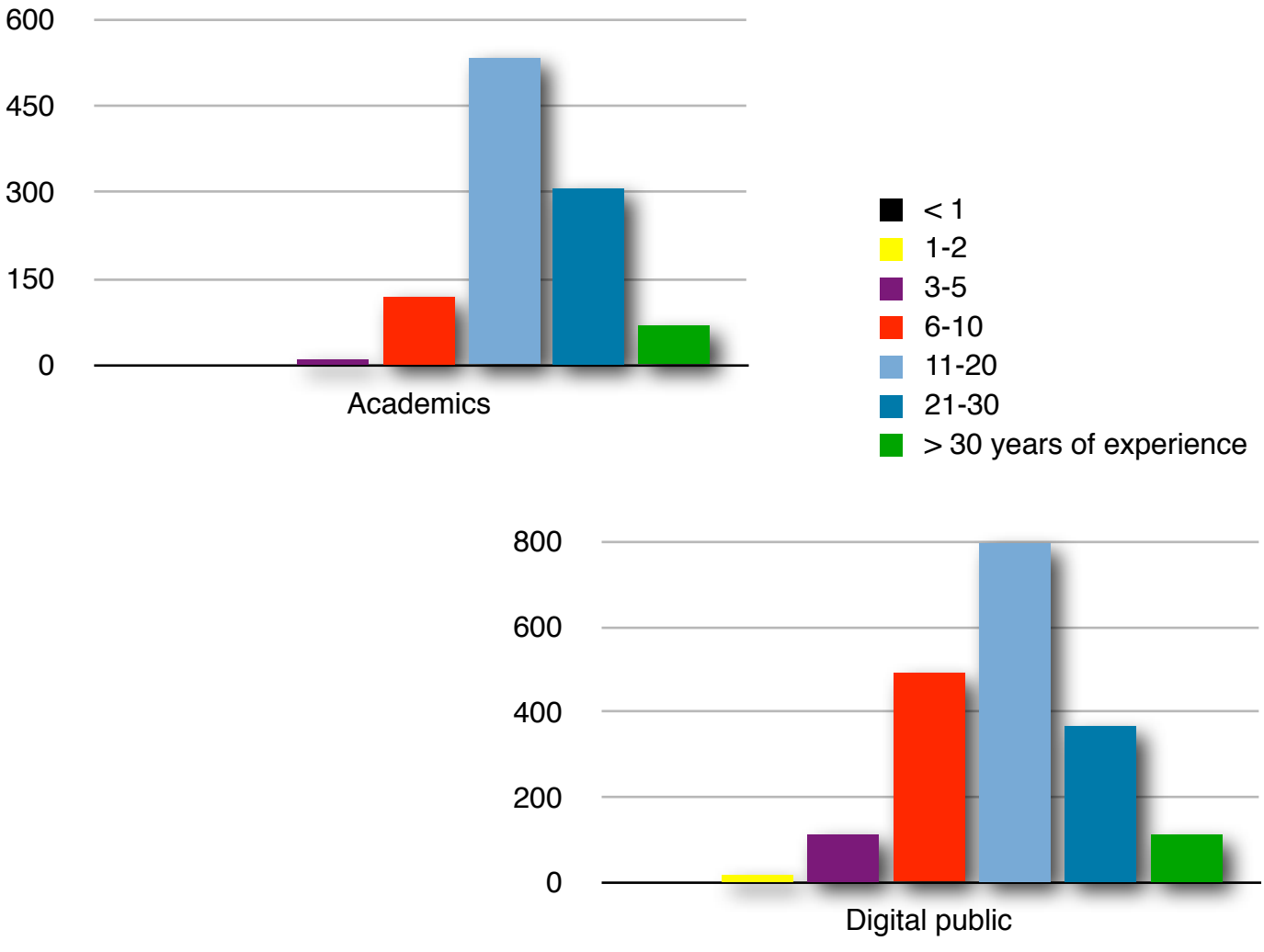


Table 7: Level of IT skills*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	*%*	*n*	*%*
How would you rate your IT skills?	Learner	2	0.2%	30	1.6%
	Capable	240	23.5%	604	31.7%
	Good	520	50.9%	936	49.1%
	Advanced	260	25.4%	336	17.6%
	Total	**1022**	**100.0%**	**1906**	**100.0%**

Table 8: Use of computer at home

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	*%*	*n*	*%*
Do you use a computer at home?	Yes	1024	97.2%	1910	100.0%
	No	25	2.4%	0	0.0%
	Total	**1053**	**100.0%**	**1910**	**100.0%**

Table 9: Operating system(s) used at home

More than one operating system can be selected by respondent

		Academic or digital public			
		Academic		Digital public	
		n	*%*	*n*	*%*
What operating system(s) do you use at home?	Windows	828	78.9%	1264	94.2%
	Mac	171	16.2%	67	3.5%

It is apparent from this question that Apple Macs are much more popular with academics than the general population and this fact should be borne in mind when reading this report. Given the structure of this question, it is not possible to provide separate analyses of 'Mac users' and 'Windows' users (or indeed other platforms such as Linux) since the categories are overlapping. This does however suggest the need for further research into the role (likely to be substantial) that operating systems play in shaping personal information management.

Table 10: Survey by file richness*

Means and 95% confidence intervals

	Academic or digital public			
	Academic		Digital public	
	mean	*CI95*	*mean*	*CI95*
File richness (create)	6.57	6.39-6.76	5.59	5.45-5.73
File richness (acquire)	5.40	5.17-5.62	4.54	4.39-4.70

Respondents were asked to indicate what kinds of broad file types (from a list of 12 types) they have on their home computer, and whether these were created by them or acquired from elsewhere. The 12 types of file: computer programs and models; databases; emails; moving images; music or sound recordings; online calendar or diary; photographs or digital art; presentations, eg Powerpoint; spreadsheets; web pages; word processed documents; other text documents, eg PDFs. File richness is a derived variable[19]. Academics hold a wider range of types: the differences are statistically significant, although not large.

[19] Explain file richness

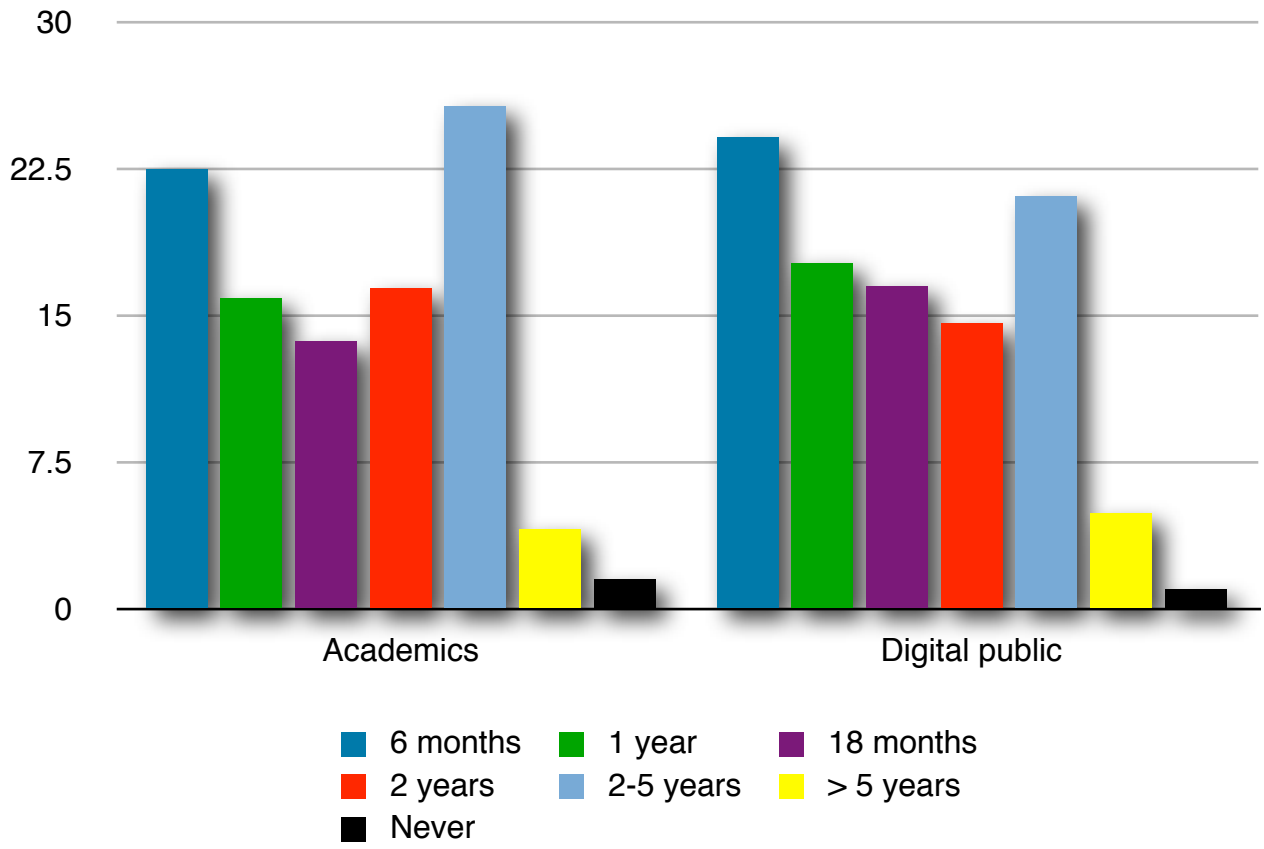
You and computer risk (all respondents)

The next 13 tables address attitudes of the participants towards various aspects of computers as well as risks associated with their use.

Table 11: Age of home computer or laptop

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
How long ago did you purchase your most recent personal computer or laptop for home use?	6 months	225	22.5%	453	24.1%
	1 year	159	15.9%	333	17.7%
	18 months	137	13.7%	311	16.5%
	2 years	164	16.4%	275	14.6%
	2-5 years	257	25.8%	397	21.1%
	>5 years	41	4.1%	93	4.9%
	Never	15	1.5%	18	1.0%
	Total	**998**	**100.0%**	**1880**	**100.0%**

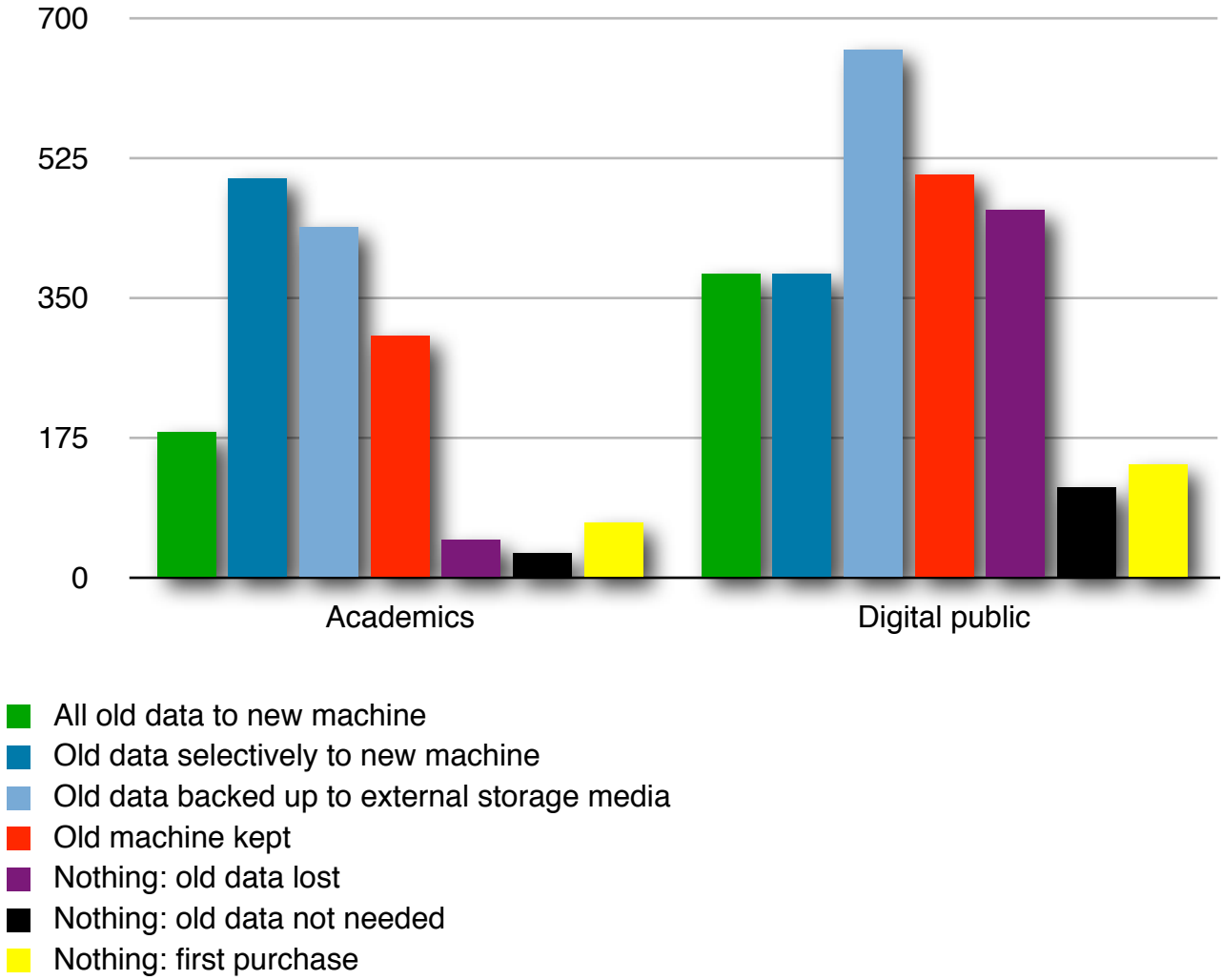


There is a hint of a bimodal distribution, with an apparent tailing off after five years. For very few respondents is the most recently purchased machine more than five years old. The vast majority of the individuals surveyed are almost certainly using current or near-current operating systems. Nearly a quarter of people have a computer purchased about six months before the survey.

Table 12: Steps to safeguard data following purchase of a new home computer*

Overlapping categories, more than one response possible

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
With respect to that purchase, what did you do about the data on your old machine?	All old data to new machine	182	17.3%	380	19.9%
	Old data selectively to new machine	500	47.5%	380	19.9%
	Old data backed up to external storage media	439	41.7%	663	34.7%
	Old machine kept	303	28.8%	504	26.4%
	Nothing, old data lost	47	4.5%	460	24.1%
	Nothing, old data not needed	31	2.9%	113	5.9%
	Nothing, first purchase	69	6.6%	142	7.4%
	Number of respondents	**1053**		**1910**	



Academics are much more likely to have taken steps to retain the computer files on their previous home computer than members of the digital public: 64.8% transferred all or some of their data to the new machine (compared with 39.8%). Similarly academics are more inclined to backup data to external storage media (41.7% compared with 34.7%). This may reflect the nature of academic work, much of it taking place at home. Of the digital public respondents, 30% did nothing with the data – it being lost or deemed unnecessary. Between 6% and 8% of respondents were using a computer that represented their first purchase of a computer. More than a quarter of people held on to their older computer.

Table 13: Serious loss of computer data

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		N	*%*	*n*	*%*
Have you ever suffered a serious loss of computerised information at home?	Yes	274	27.6%	562	29.8%
	No, but had a near miss	332	33.5%	571	30.3%
	No, never	386	38.9%	752	39.9%
	Total	**992**	**100.0%**	**1885**	**100.0%**

As might have been expected, there is no or little difference between the populations in terms of their experience of serious loss of computer data at home, although the absolute levels of incidence are distressingly high: over a quarter of respondents in both samples.

Table 14: Cause of serious data loss

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
How did you lose that data?	Computer lost/stolen	32	12.7%	88	15.8%
	File wouldn't open	10	4.0%	8	1.4%
	Virus attack	1	0.4%	0	0.0%
	Did not have necessary software	5	2.0%	7	1.3%
	Hard disk crashed	11	4.4%	43	7.7%
	Lost password	2	0.8%	2	0.4%
	Couldn't find file	169	67.1%	386	69.2%
	Deleted it in error	22	8.7%	24	4.3%
	Total	**252**	**100.0%**	**558**	**100.0%**

The main cause of serious data loss for both populations is an inability to find files (67.1% academics, 69.2% digital public) rather than, for instance, a hard disk failure (just 4.4% and 7.7% for academics and digital public, respectively). It is conceivable that viruses - of which individuals are unaware - might be responsible in some way for the difficulty in finding the file but it is unlikely - for one thing around 90% and more of individuals possess antivirus software (see Table 24).

It is also notable that 12.7% and 15.8% of the cases of data loss occurred through a computer being lost or stolen, with these events being the second most frequent cause. Loss due to a file not opening or the absence of the necessary software was 6% and 2.7% for academics and the public, respectively. No meaningful difference between the populations is evident. Overall the findings suggest a need for a greater awareness of the real risks and also for more effective personal information management systems.

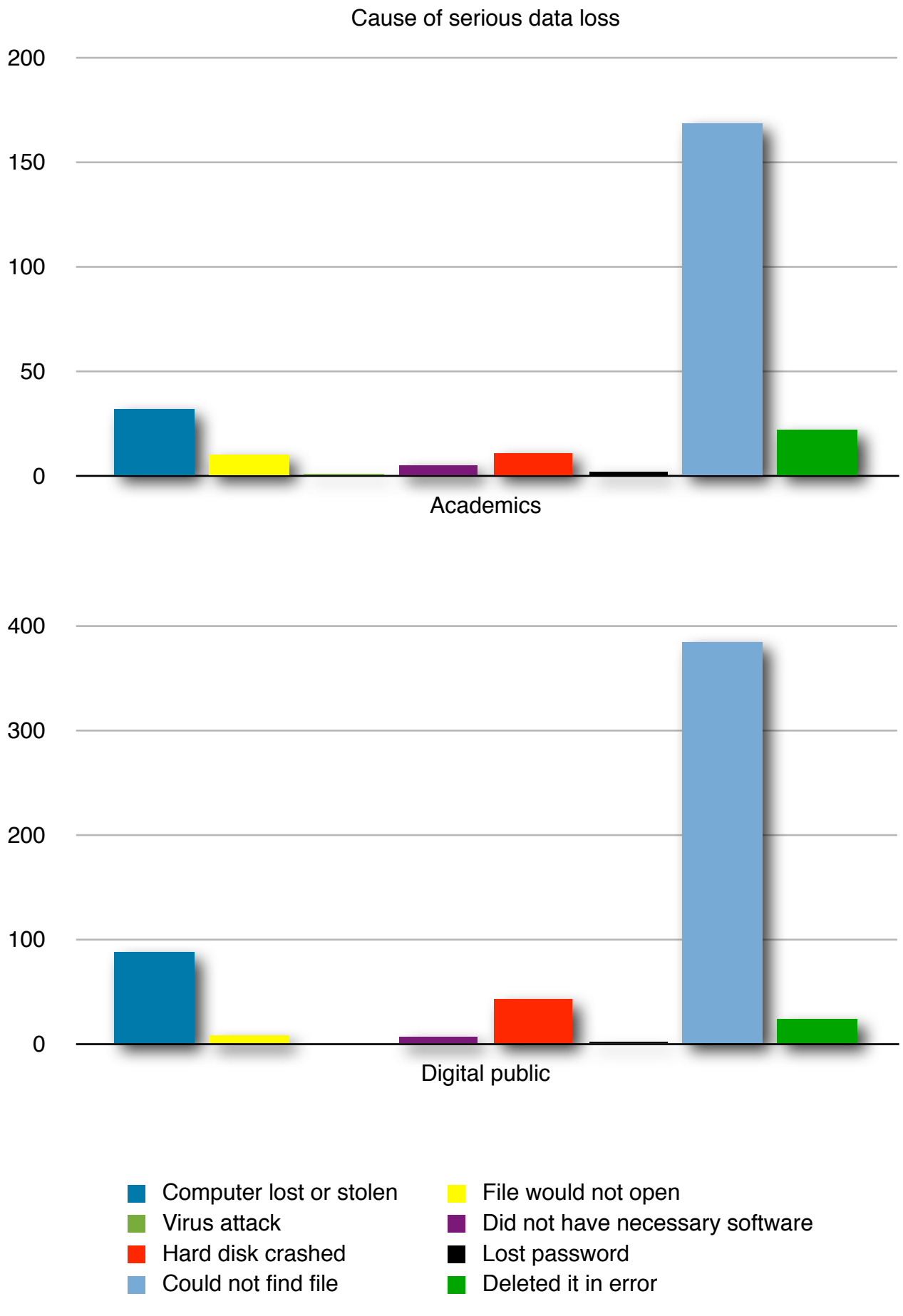

Cause of serious data loss
200
150
100
50
0
Academics
400
300
200
100
0
Digital public
Computer lost or stolen
Virus attack
Hard disk crashed
Could not find file
File would not open
Did not have necessary software
Lost password
Deleted it in error

Table 15: Single or multiple home passwords

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
Do you use the same password for everything at home?	Yes	181	18.3%	322	17.1%
	No, different	809	81.7%	1563	82.9%
	Total	**990**	**100.0%**	**1885**	**100.0%**

Table 16: Number of home passwords

Means and 95% confidence intervals

	Academic or digital public			
	Academic		Digital public	
	mean	*CI95*	*mean*	*CI95*
Roughly how many passwords do you have for home use?	5.55	4.69-6.41	4.97	3.38-6.11

A large majority of both academics and the digital public use multiple passwords at home, around 5.0 to 5.6 on average.

Tables 17 through to 21 explore levels of perception of five sources of computer risk: prying eyes, backup failure, burglary, virus attack and identity theft. In all cases there is a statistically significant difference between the two populations. With the exception of 'backup failure', members of the digital public exhibit higher levels of anxiety than academics. This warrants further investigation[20].

Table 17: Level of concern over 'prying eyes'*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How concerned are you about 'prying eyes'?	Not at all concerned	535	51.2%	688	37.0%
	Slightly concerned	311	29.8%	561	30.2%
	Quite concerned	132	12.6%	342	18.4%
	Very concerned	67	6.4%	269	14.5%
	Total	**1045**	**100.0%**	**1860**	**100.0%**

Respondents who are 'quite' or 'very concerned': academics 19.0%; digital public 32.9%. Thus more than 80% of academics are relatively unconcerned about 'prying eyes'.

Table 18: Level of concern over 'backup failure'*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How concerned are you about 'backup failure'?	Not at all concerned	164	15.7%	284	15.2%
	Slightly concerned	353	33.8%	806	43.2%
	Quite concerned	328	31.4%	492	26.4%
	Very concerned	200	19.1%	285	15.3%
	Total	**1045**	**100.0%**	**1867**	**100.0%**

Respondents who are 'quite' or 'very concerned': academics 50.5%; digital public 41.7%.

[20] A future paper will examine these findings and data in more detail.

Table 19: Level of concern over 'burglary'*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How concerned are you about 'burglary'?	Not at all concerned	231	22.1%	310	16.6%
	Slightly concerned	478	45.8%	707	37.9%
	Quite concerned	220	21.1%	473	25.4%
	Very concerned	114	10.9%	373	20.0%
	Total	**1043**	**100.0%**	**1863**	**100.0%**

Respondents who are 'quite' or 'very concerned': academics 32.0%; digital public 45.4%. It is worth comparing this relatively and apparently low concern with the finding in Table 14 that revealed computer loss or theft as the second most common cause of a serious loss of data. Again, there may be a need for greater assessment and awareness of true risks.

Table 20: Level of concern over 'computer viruses or other malicious software'*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How concerned are you about 'computer viruses or other malicious software'?	Not at all concerned	149	14.4%	147	7.8%
	Slightly concerned	297	28.7%	527	28.1%
	Quite concerned	349	33.7%	609	32.4%
	Very concerned	240	23.2%	594	31.6%
	Total	**1035**	**100.0%**	**1887**	**100.0%**

Respondents who are 'quite' or 'very concerned': academics 56.9%; digital public 64.0%.

Table 21: Level of concern over 'identity theft'*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
How concerned are you about 'identity theft'?	Not at all concerned	156	15.1%	131	7.0%
	Slightly concerned	368	35.7%	516	27.5%
	Quite concerned	318	30.8%	515	27.5%
	Very concerned	190	18.4%	712	38.0%
	Total	**1032**	**100.0%**	**1874**	**100.0%**

Respondents who are 'quite' or 'very concerned': academics 49.2%; digital public 65.5%.

Table 22: Ownership of home backup software

Column percentages

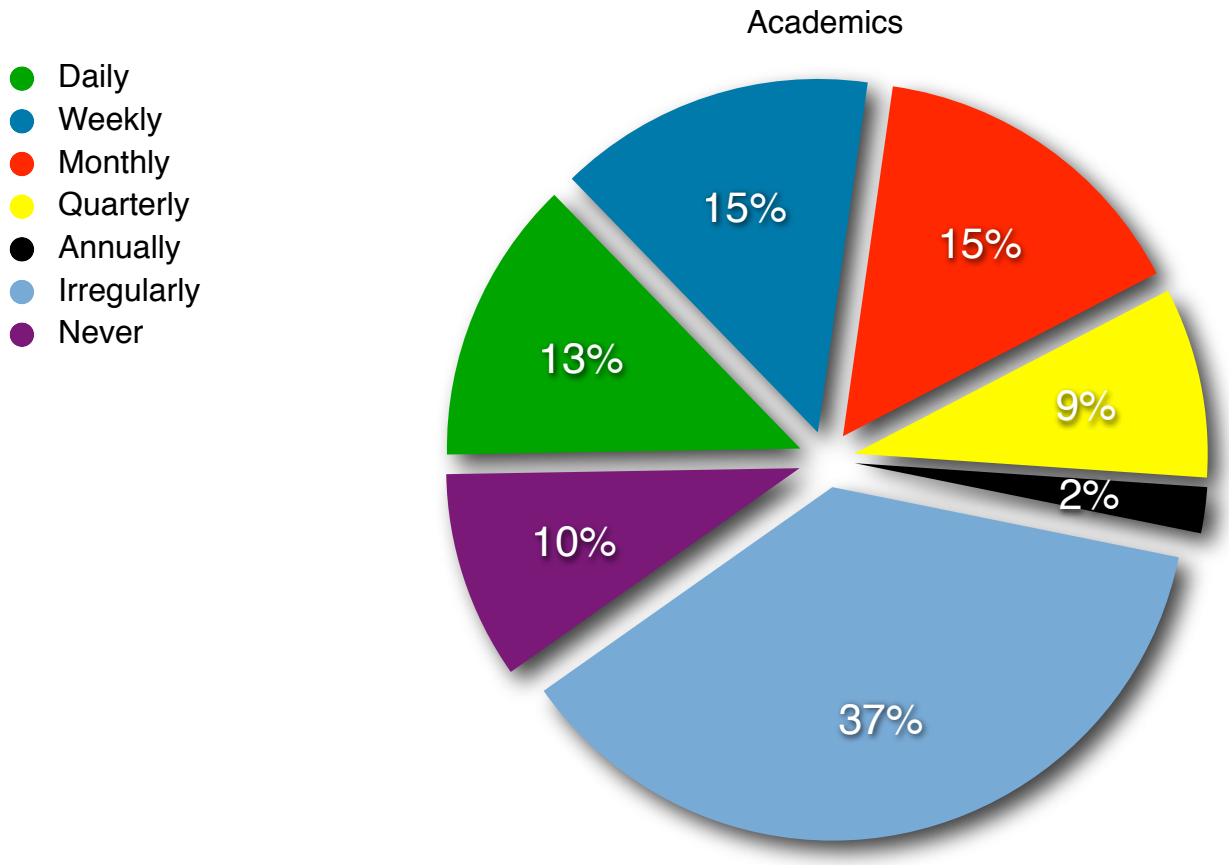
		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
Do you possess backup software for use at home?	Yes	542	55.8%	1151	61.1%
	No	333	34.3%	528	28.0%
	I'm not sure	96	9.9%	204	10.8%
	Total	**971**	**100.0%**	**1883**	**100.0%**

Ownership of backup software appears consistent across the two populations. Notably, well over a third of home computer users do not have such software (or are unaware of owning it), and are therefore putting their digital belongings at risk.

Table 23: Frequency of home backup*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
How often do you backup data at home?	Daily	124	13.0%	195	10.4%
	Weekly	138	14.5%	341	18.3%
	Monthly	144	15.1%	321	17.2%
	Quarterly	83	8.7%	131	7.0%
	Annually	20	2.1%	29	1.6%
	Irregularly	353	37.0%	487	26.1%
	Never	91	9.5%	363	19.4%
	Total	**953**	**100.0%**	**1867**	**100.0%**



Academics appear to be more unlikely than members of the digital public **never** to backup their computer files (9.5% vs 19.4%). Otherwise the differences in frequency are unremarkable. The term 'irregularly' needs to be interpreted carefully. The ordering of the questions might suggest low frequency but this should not be assumed. It might reflect a tendency to backup when the need to do so is perceived and felt more strongly, eg at stages or at the end of a project or when an extensive piece of writing is completed. It may reflect a difference in strategy. In any event it is apparent that only 42.6% and 45.9% use the software that they possess both regularly and at least once every month.

Table 24: Ownership of antivirus software*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Do you possess antivirus software for use at home?	Yes	866	89.2%	1807	96.2%
	No	88	9.1%	47	2.5%
	I'm not sure	17	1.8%	24	1.3%
	Total	**971**	**100.0%**	**1878**	**100.0%**

While members of the digital public report significantly higher levels of ownership of antivirus software, this finding has to be tempered with the earlier observation (Table 9) that they are more likely to be using a Windows personal computer rather than an Apple Mac.

A precious computer file (critical incident)

In this section of the questionnaire, respondents were asked to think about 'a recent example where you have created or acquired a computer file that is of great importance to you in your personal or working life'. This is a 'critical incident' and this methodological approach has two great benefits: it focuses on a specific concrete event and, by virtue of that fact, captures a pseudo-random snapshot of the population at large. The drawback is that the critical event described may or may not represent typical behaviour.

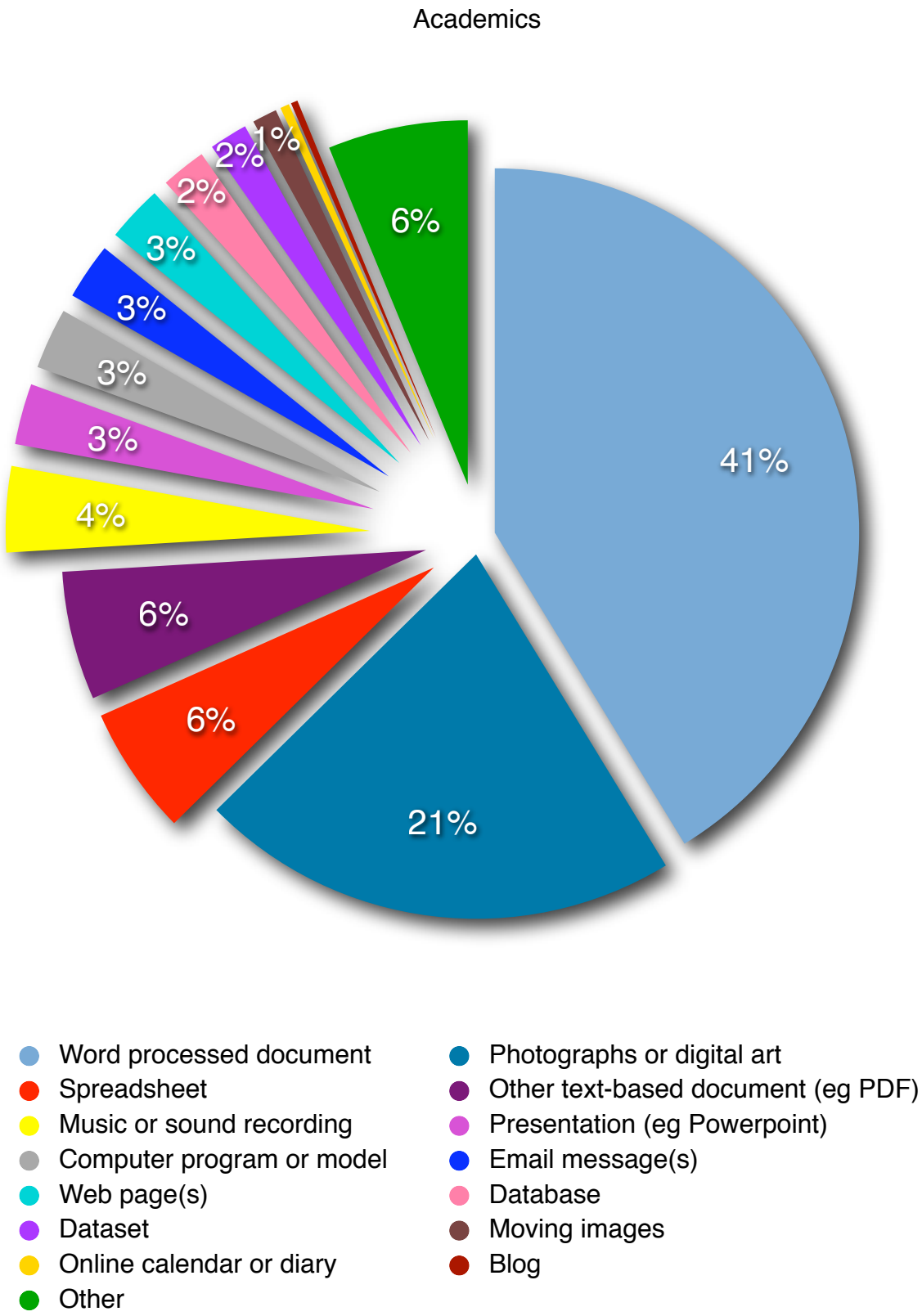
Table 25: A recent computer file of great personal importance, by filetype

Column percentages

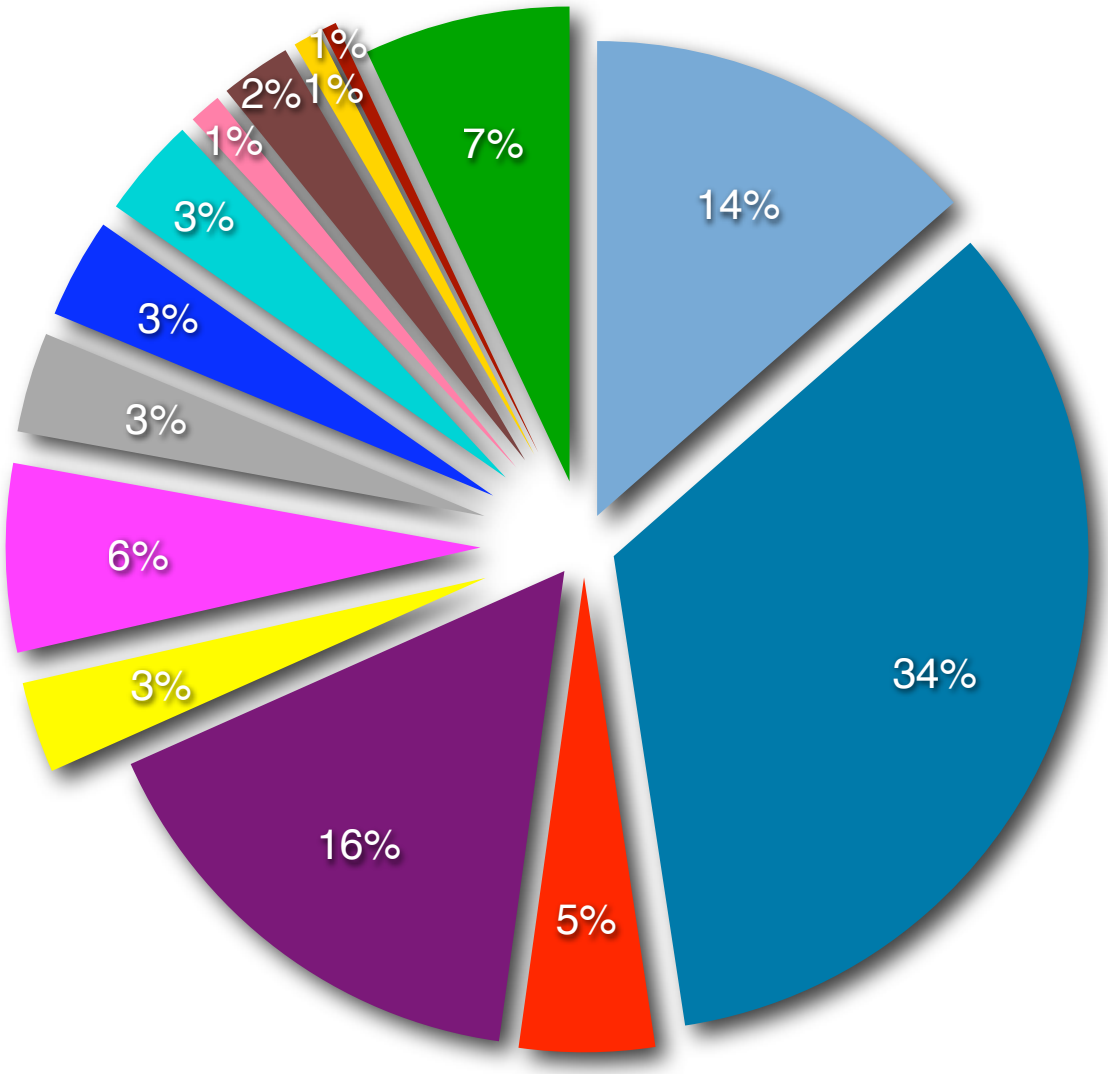
		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
What kind of file are we talking about?	Word processed document	390	41.2	224	13.5
	Photographs or digital art	202	21.3	565	34.0
	Spreadsheet	54	5.7	77	4.6
	Other text-based document (eg PDF)	54	5.7	268	16.1
	Music or sound recording	36	3.8	51	3.1
	Presentation (eg Powerpoint)	26	2.7	107	6.4
	Computer program or model	26	2.7	56	3.4
	Email message(s)	24	2.5	57	3.4
	Web page(s)	24	2.5	57	3.4
	Database	19	2.0	18	1.1
	Dataset	16	1.7	0	0.0
	Moving images	10	1.1	40	2.4
	Online calendar or diary	4	0.4	15	0.9
	Blog	3	0.3	9	0.5
	Other	59	6.2	116	7.0
	Total	**947**	**100.0%**	**1660**	**100.0%**

For academics the top four categories are: word processed documents, photographs and digital art, other text based documents such as PDFs, and spreadsheets. For the digital public the corresponding four categories are: photographs and digital art, other text based documents such as PDFs, word processed documents, and presentations. The differences probably reflect different patterns in the balance of home and work activities, with academics notably reporting a highest incidence for word processed documents. Interestingly email messages score relatively lowly for both populations. Music and sound recordings also have relatively low proportions but this might simply reflect the fact that for most people such recordings are published and can be replaced. The status of (personal) moving images

can be expected to change over the coming years. The diversity of choices selected is interesting and appears to be slightly more evenly distributed in the digital public sample.



Digital public



- Word processed document
- Spreadsheet
- Music or sound recording
- Computer program or model
- Web page(s)
- Dataset
- Online calendar or diary
- Other
- Photographs or digital art
- Other text-based document (eg PDF)
- Presentation (eg Powerpoint)
- Email message(s)
- Database
- Moving images
- Blog

Table 26: A recent computer file of great personal importance, by primary value*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
What was the *primary* value to you of that particular computer file?	Colleagues	45	5.0%	234	13.6%
	Final version	95	10.6%	62	3.6%
	Interest to future historians	406	45.5%	347	20.2%
	Sentimental	91	10.2%	304	17.7%
	Working document	29	3.2%	82	4.8%
	Sensitive, personal or financial information	201	22.5%	541	31.5%
	Personal reference	21	2.4%	36	2.1%
	Other	5	0.6%	110	6.4%
	Total	**893**	**100.0%**	**1716**	**100.0%**

Academics clearly have a much greater sense of the value of these 'precious files' for posterity (45.5% think they will be of interest to future historians, compared with 20.2% of the digital public). The relatively low value attached to 'sentimental' reasons seems at first glance a little surprising. This may reflect the central role that home computers now have in most people's lives for handling material of a sensitive nature (31.5% of the digital public), perhaps related to financial, legal or otherwise private records.

An interesting question is the relationship with analogue items: people are still keeping analogue artefacts in a sentimental capacity. The relatively high level of value attending to sensitive information emphasises the need for confidentiality and privacy issues to be addressed carefully. In examining the findings it is important to bear in mind that these data are *relative*, focussing simply on the *primary* value. The strength of the diversity of reasons for attaching value is very interesting and important. In this respect the responses of the digital public seem to be more balanced, with less emphasis on a single category.

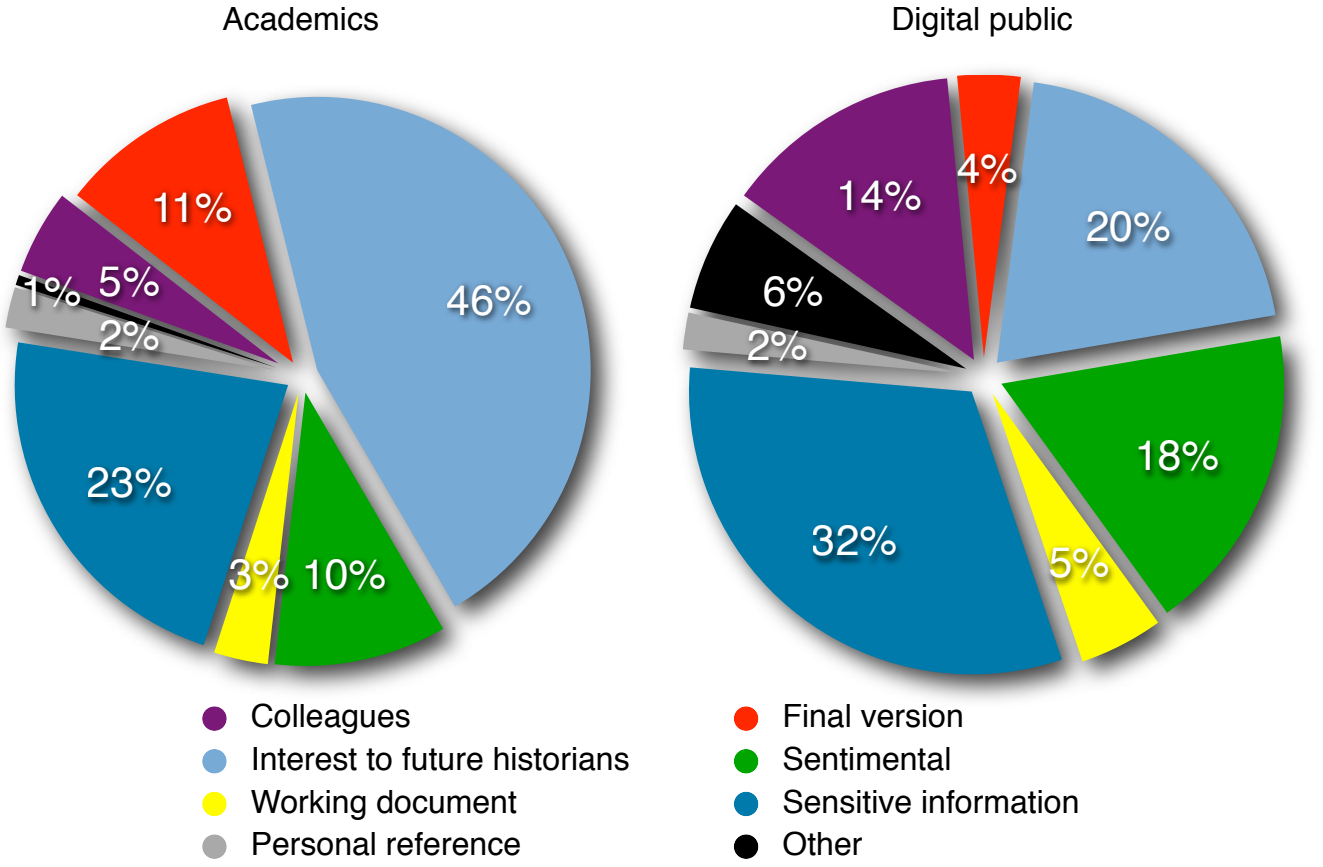


Table 27: A recent computer file of great personal importance, by hard copy

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Did you keep a hard copy (ie a printout) of that computer file?	Yes	358	38.7%	656	36.7%
	No	513	55.5%	895	50.0%
	Not relevant	46	5.0%	188	10.5%
	I'm not sure	8	0.9%	50	2.8%
	Total	**925**	**100.0%**	**1789**	**100.0%**

This finding emphasises again the hybrid nature of many personal archives. The low number of respondents reporting that they 'weren't sure' is interesting, hinting perhaps at clear dichotomy of attitude or necessity with regard to printouts.

The option of printing out a hard copy is, of course, not applicable in all cases (eg computer programs or videos).

Table 28: A recent computer file of great personal importance, working life or personal life*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Does this file relate *primarily* to your working life or to your personal life?	Working life	487	52.0%	472	26.4%
	Personal life	449	48.0%	1314	73.6%
	Total	**936**	**100.0%**	**1786**	**100.0%**

With more than a quarter of members of the digital public choosing a computer file that relates to their working life, it is clear that the computer lies right at the crossroads where the personal and working lives converge and overlap. As throughout this report, the split between these two worlds is very hard to delineate clearly, if at all. With the growing predominance of the laptop, this distinction becomes even more hazy.

Table 29: A recent computer file of great personal importance, created or acquired

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Did you create that file or acquire it from elsewhere?	Created	842	89.9%	1456	77.2%
	Acquired	95	10.1%	429	22.8%
	Total	**937**	**100.0%**	**1885**	**100.0%**

A very large majority of respondents in both populations (89.9% and 77.2%) chose a computer file that they had created themselves rather than one that they had acquired.

A precious computer file *that you created* (critical incident)

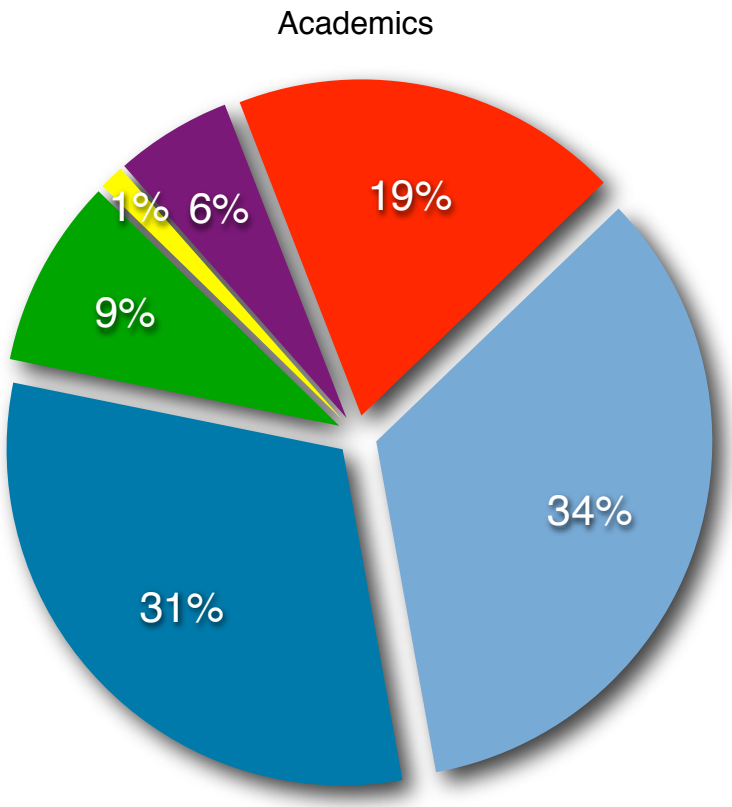
The next three tables relate to the instances where the single 'precious' computer file was created by the respondent himself or herself.

Table 30: A precious created file, version control*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Thinking again about that computer file, how did you manage different versions while you were working on it?	I printed out some versions and saved others	47	5.6%	89	4.9%
	Every time I worked on it, I saved it under a different name	156	18.7%	220	12.1%
	I saved some versions under a different name, but not all	286	34.3%	360	19.9%
	I just saved under one filename, so I only ever had the final version	258	31.0%	696	38.4%
	I only worked on it once	76	9.1%	190	10.5%
	I don't remember	10	1.2%	258	14.2%
	Total	**833**	**100.0%**	**1813**	**100.0%**

A key finding is that there is a wide variety of practice around version control and about maintaining copies (digital or hard copy) of interim versions. Academics are more likely to produce interim versions than members of the digital public.



- I printed out some versions and saved others
- Every time I worked on it I saved it under a different name
- I saved some versions under a different name but not all
- I just saved under one filename so I only ever had the final version
- I only worked on it once
- I do not remember

Table 31: A precious created file, file naming conventions

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
What kind of file name did you give it?	I didn't give it a file name	41	4.9%	192	10.6%
	The file name was descriptive of the subject	383	45.9%	1,019	56.1%
	The file name included a date-or-version number	147	17.6%	211	11.6%
	The file name included a date-or-version number and subject information	255	30.6%	252	13.9%
	I don't remember	8	1.0%	144	7.9%
	Total	**834**	**100.0%**	**1818**	**100.0%**

File naming conventions are similarly heterogeneous, although a large majority of both academics and the digital public incorporate at least some indication of the subject matter in the file name. It is intriguing that about 10% of the digital public respondents did not give their files names at all, presumably settling for the default name given by the system.

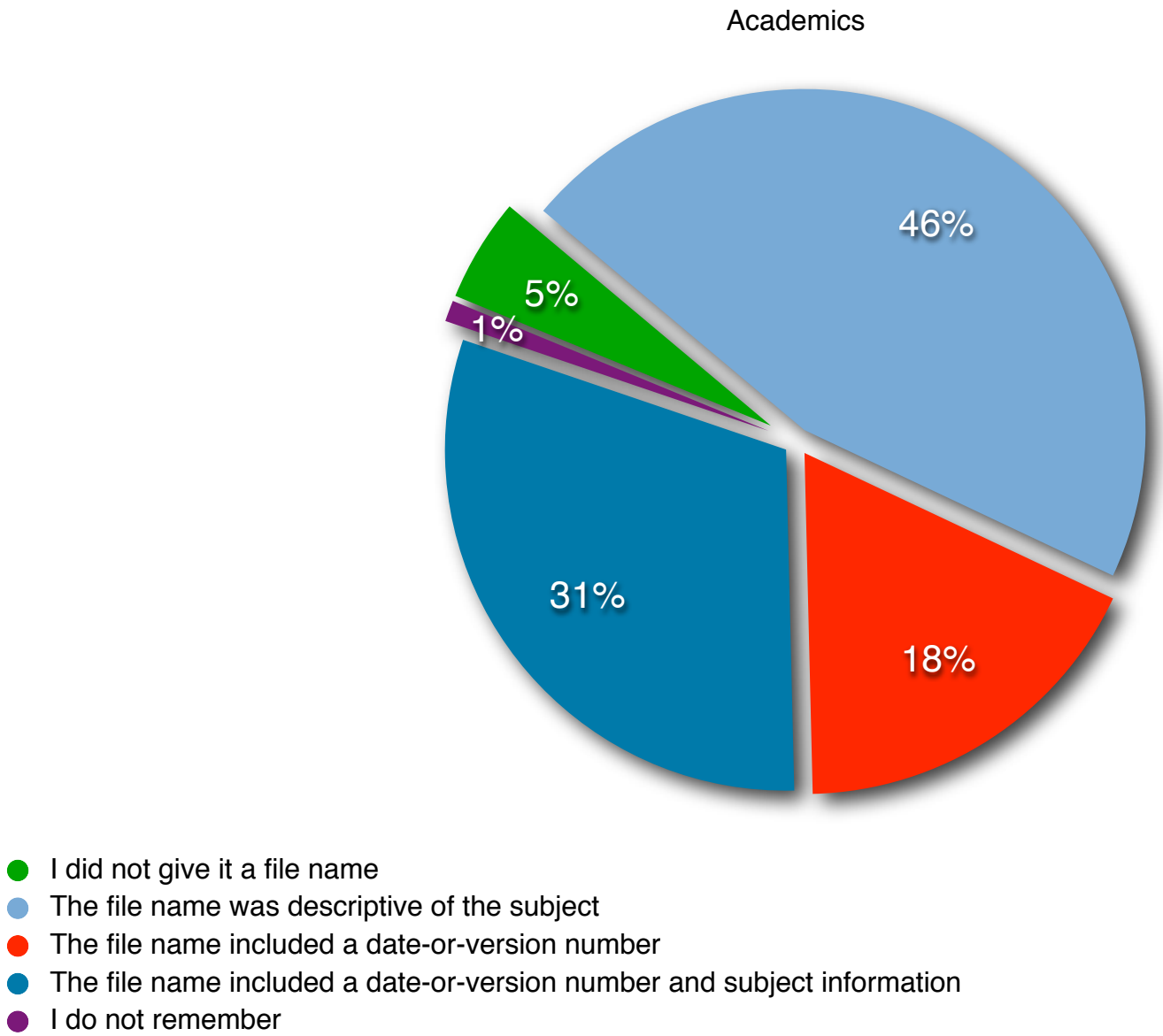
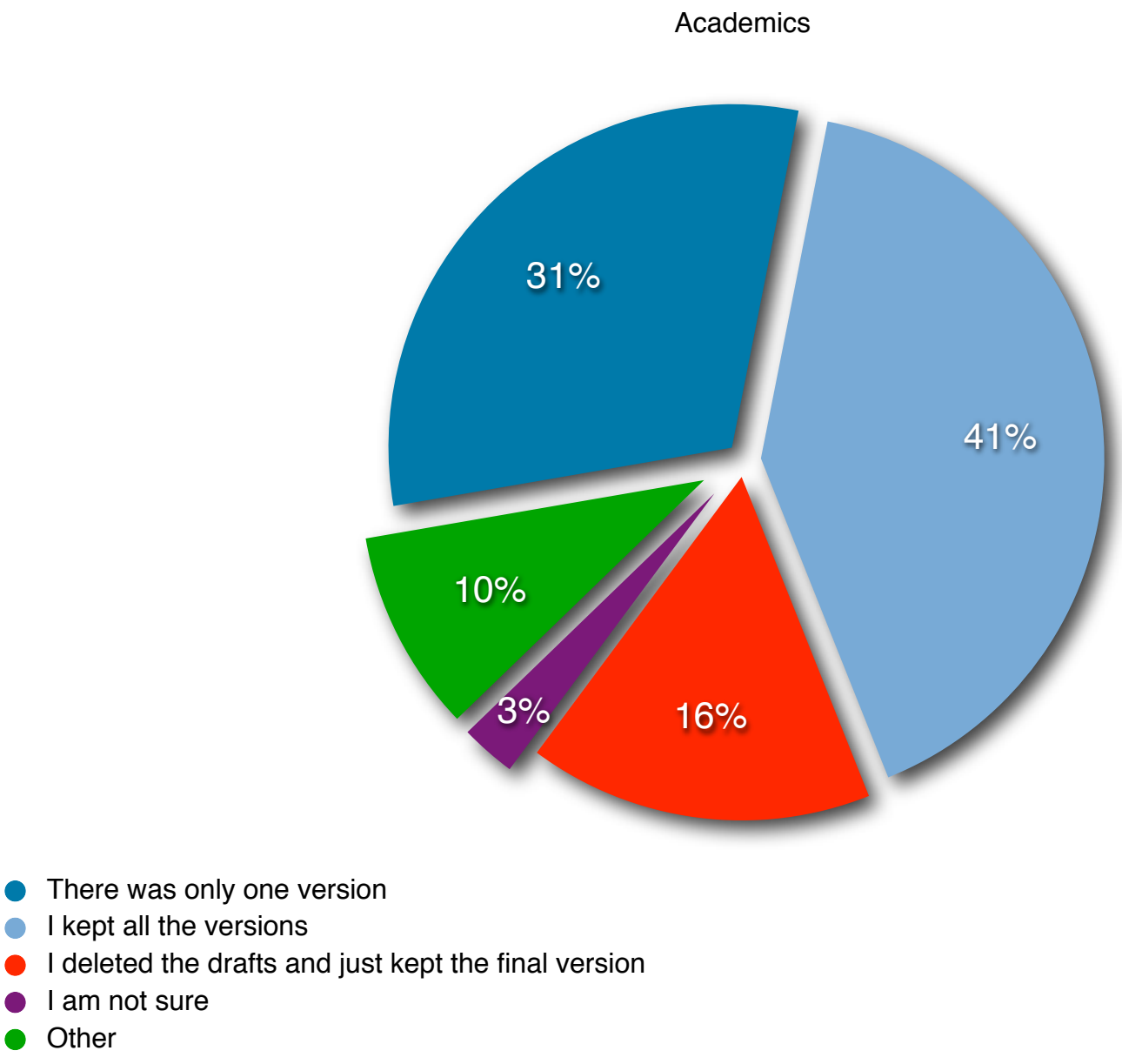


Table 32: A precious created file, fate of drafts?*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	%	n	%
What did you do with that computer file when you had finished with it?	There was only one version	285	30.8	771	42.3
	I kept all the versions	377	40.8	471	25.9
	I deleted the drafts and just kept the final version	150	16.2	355	19.5
	I'm not sure	24	2.6	135	7.4
	Other	88	9.5	90	4.9
	Total	**924**	**100.0%**	**1,822**	**100.0%**



The foregoing result suggests that with most people interim versions, drafts, of 'precious files' are not being retained, either because there is only one version (specifically, interim versions are not being created) or interim versions are actively deleted; approximately 15% and 20% of people actively deleted drafts. Table 30 indicates that 9% of academics report working on it only once. Academics are more likely (40.8%) than members of the digital public (25.9%) to keep all the versions of that precious file.

A precious computer file *that you acquired* (critical incident)

The next three tables relate to precious files that have been acquired from elsewhere.

Table 33: A precious acquired file, how sourced

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	%	n	%
How did you acquire that computer file?	It was sent to me as an email attachment	33	35.1	88	22.5
	It was given to me on a disk or some other transfer medium	21	22.3	125	32.0
	I was alerted to it, eg by email, RSS feed, personal recommendation	12	12.8	31	7.9
	I searched or browsed the Internet	19	20.2	99	25.3
	Other	9	9.6	48	12.3
	Total	**94**	**100.00%**	**391**	**100.00%**

This analysis highlights email communication as an important mechanism for file transfer for academics. It is also notable that between 20% and 30% of people obtained the precious file *via* the internet through browsing or searching on the internet. It would be interesting to explore the social nature of acquisition further: to establish to what extent people acquire precious files through other people directly compared with acquiring such files independently of anyone else, eg through browsing websites.

Table 34: A precious acquired file, renamed or not

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Did you rename the file when you acquired it?	Yes	38	40.4%	126	29.4%
	No	49	52.1%	233	54.3%
	I can't remember	7	7.4%	70	16.3%
	Total	**94**	**100.0%**	**429**	**100.0%**

More than half of individuals report that they did not rename these files. Academics appear to be more likely to rename acquired files than members of the digital public, which may be suggestive of better personal information management practice motivated by professional necessity. Renaming is likely to improve the ability to find the file later, but might conceivably also reflect a 'repurposing' of the file in some way - a finding purpose seems likely because many files acquired in this way lack meaningful and identifiable names and can get lost among numerous downloaded files.

Table 35: A precious acquired file, how renamed

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
What kind of file name did you give it?	The file name included a date-or-version number and subject information	9	23.7	23	17.7
	The file name included a date-or-version number	3	7.9	12	9.2
	The file name was descriptive of the subject	24	63.2	84	64.6
	I don't remember	0	0.0	5	3.8
	Other	2	5.3	6	4.6
	Total	**38**	**100.0%**	**130**	**100.0%**

Again, there is no statistically meaningful difference between the two populations and much variation in practice. Most acquired files that have been renamed include some level of subject description in the file name.

How you organise your computer files (all respondents)

At this point in the survey, the questionnaire moves away from analysing a critical incident (that 'precious file') to questions more directly concerning personal information management practices generally.

Table 36: Strategies for organising computer files so that they can easily be found

Overlapping categories, more than one response possible

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Which of the following strategies do you use to make sure that you can subsequently find files on your computer?	I give descriptive and meaningful filenames	789	93.8	1,151	61.9
	I leave important files on the desktop	251	29.8	317	17.0
	I leave files in the default folder (eg 'My Documents')	166	19.7	497	26.7
	I put my files in named folders organised by subject only (eg `Summer holidays')	357	42.4	433	23.3
	I put my files in named folders organised by type only (eg `memorable photographs')	173	20.6	462	24.8
	I put my files in named folders organised by type and subject (e.g. `Summer holiday snaps')	463	55.1	519	27.9
	I put my files into date folders (eg `this week')	118	14.0	127	6.8
	I make several copies of my files at different locations	262	31.2	199	10.7
	I send files to myself as email attachments	367	43.6	214	11.5
	Nothing	3	0.4	82	4.4
	Other	82	9.8	133	7.2
	Number of respondents	**846**		**1731**	

This table again reinforces emergent themes from the previous section (§3.4): (i) there is much variation in personal information management practice, and (ii) most people provide some level of subject description in their file naming conventions and/or their use of thematically organised folders to help them subsequently find material. A few further findings stand out: 43.6% of academics send themselves files as email attachments; moreover many

creators are using default options for organising files with 29.8% of academics leaving files on the ‘desktop’, and 26.7% of the digital public using a default folder.

On the whole there is evidence of a widespread effort and desire to manage information.

There is no significant discrepancy between academics and the digital public overall. It is extraordinary, however, that 31.2% of academics make several copies of files in different locations in order to increase the likelihood of finding files. This result together with the use of email as a place to store files for future finding suggests that the design of better systems for organising files would be welcomed.

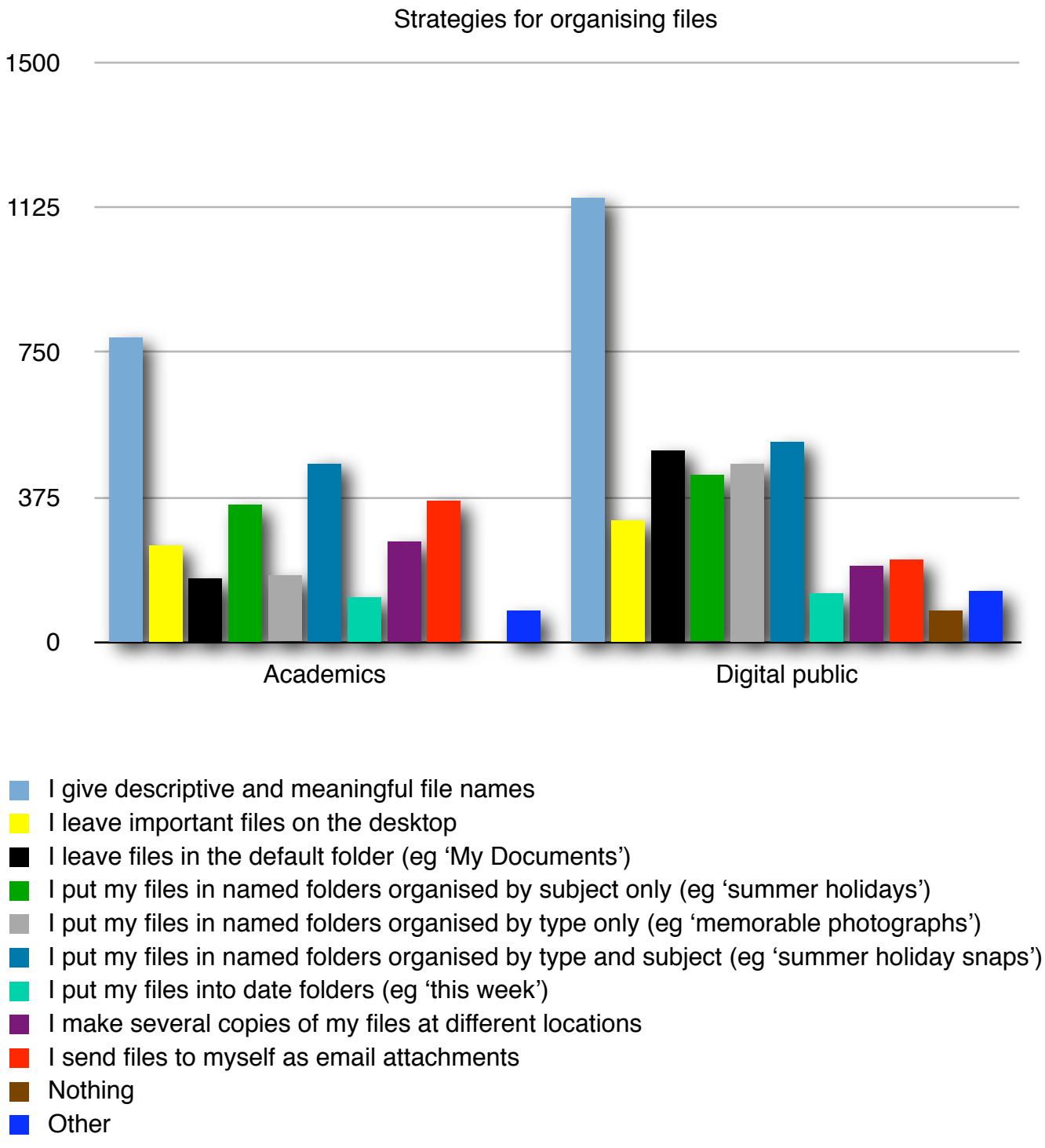
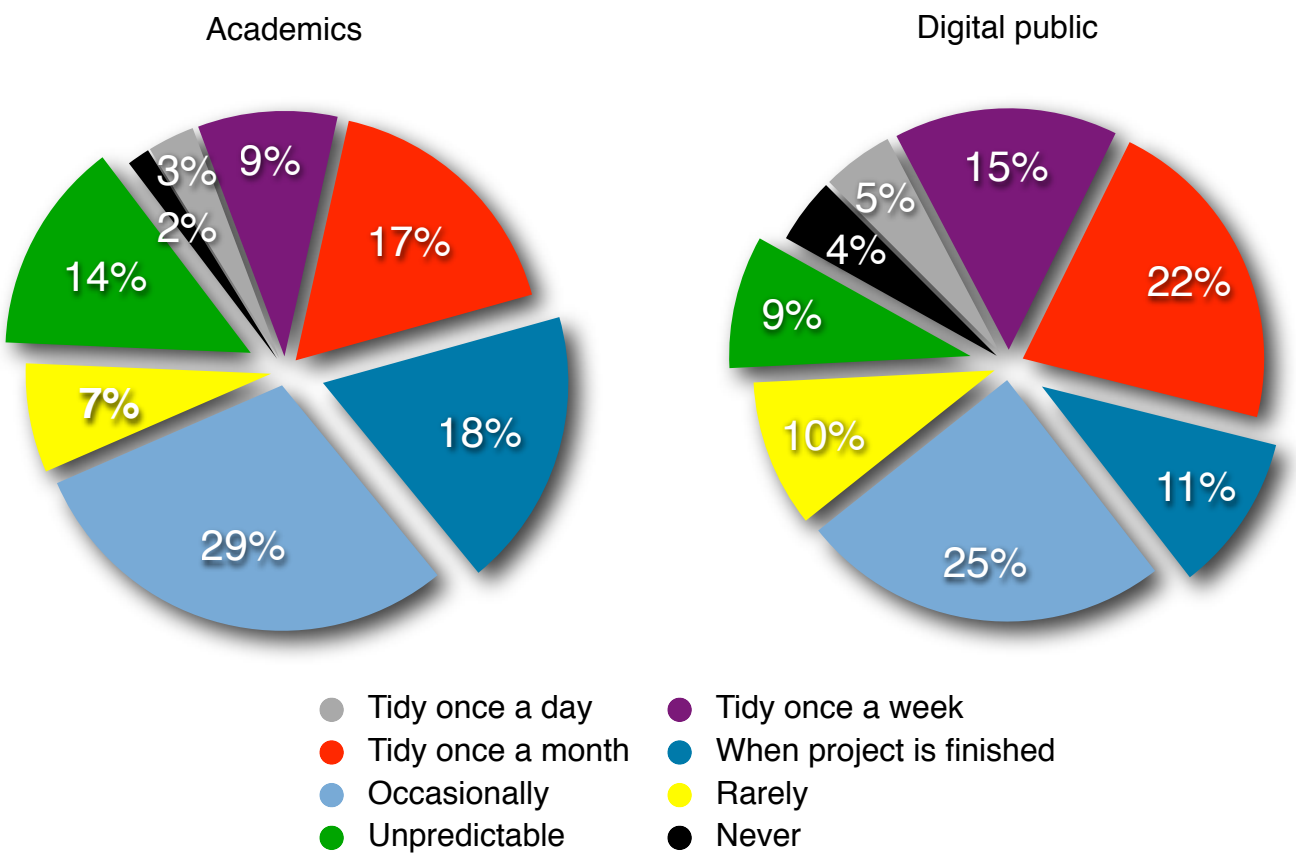


Table 37: Frequency of tidying computer files and folders*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	*n*	%
Which of the following statements *best represents* your approach to organising your computer files?	Tidy at least once a day	29	3.1%	89	4.8%
	Tidy once a week	85	9.2%	278	15.0%
	Tidy once a month	159	17.2%	401	21.6%
	When project is finished	170	18.4%	197	10.6%
	Occasionally	271	29.4%	460	24.8%
	Rarely	66	7.2%	184	9.9%
	Unpredictable	129	14.0%	166	8.9%
	Never	14	1.5%	82	4.4%
	Total	**923**	**100.0%**	**1857**	**100.0%**



The most popular approach is ‘occasionally’. Academics seem to be quite good ‘finishers’: 18.4% say that they organise their files at the end of a project (compared with 10.6% in the case of the digital public). At the same time 41.4% of the digital public tidy up regularly, once a day, week or month (compared with academics with a total for regular tidying of 29.5%).

It provides a clear difference in approach, and yet 47.9% and 52.0% of academics and digital public, respectively, either tidy regularly or do so at the end of projects.

Only 8.7% of academics and 14.3% of the digital public admit that they ‘rarely’ or ‘never’ systematically organise their computer files.

How you find your computer files (all respondents)

The next five tables explore respondents' perceptions of how reliant they are on different strategies for finding computer files when they need them.

Table 38: Dependence on browsing for finding files*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Finding files by 'browsing'	Not at all dependent	140	16.5%	497	28.8%
	Slightly dependent	367	43.3%	751	43.5%
	Quite dependent	251	29.6%	503	21.9%
	Highly dependent	90	10.6%	100	5.8%
	Total	**848**	**100.0%**	**1727**	**100.0%**

Those who are 'quite' or 'highly dependent' on browsing: academics 40.2%; digital public 27.7%

Table 39: Dependence on memory of file name for finding files*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Finding files by 'memory of the file name'	Not at all dependent	42	5.0%	262	15.2%
	Slightly dependent	200	23.6%	584	33.9%
	Quite dependent	350	41.3%	589	34.2%
	Highly dependent	255	30.1%	288	16.7%
	Total	**847**	**100.0%**	**1723**	**100.0%**

Those who are 'quite' or 'highly dependent' on memory of file name: academics 71.3%; digital public 50.9%

Table 40: Dependence on recent files option for finding files*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Finding files by 'using the "recent files" option within a package'	Not at all dependent	314	35.4%	767	41.8%
	Slightly dependent	348	39.2%	657	35.8%
	Quite dependent	175	19.7%	341	18.6%
	Highly dependent	51	5.8%	70	3.8%
	Total	**887**	**100.0%**	**1834**	**100.0%**

Those who are 'quite' or 'highly dependent' on recent files option: academics 25.5%; digital public 22.4%

Table 41: Dependence on keyword search tools for finding files*

Column percentages

		Academic or digital public			
		Academic		Digital public	
		n	%	n	%
Finding files by 'a keyword search tool (eg Google Desktop)'	Not at all dependent	359	39.7%	842	47.2%
	Slightly dependent	289	32.0%	508	28.5%
	Quite dependent	164	18.1%	321	18.0%
	Highly dependent	92	10.2%	112	6.3%
	Total	**904**	**100.0%**	**1784**	**100.0%**

Those who are ‘quite’ or ‘highly dependent’ on keyword search tools: academics 28.3%; digital public 24.3%. Worth noting, perhaps, are the small differences between academics and public in being ‘not at all dependent’ and correspondingly in being ‘highly dependent’. Is there a difference in this context between Apple Mac users (with the Spotlight search tool) and others?

Table 42: Dependence on memory of folder or location for finding files*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
Finding files by ‘memory of the folder or location’	Not at all dependent	27	3.1%	250	14.1%
	Slightly dependent	104	12.2%	537	30.8%
	Quite dependent	342	40.0%	642	36.8%
	Highly dependent	382	44.7%	321	18.4%
	Total	**855**	**100.0%**	**1744**	**100.0%**

Those who are ‘quite’ or ‘highly dependent’ on memory of folder: academics 84.7%; digital public 55.2%.

Considering Tables 37 through to 42: both academics and the digital public favour most especially memory of folder or filename, with browsing as a clear runner up. Keyword search is the least favoured with 39.7% and 47.2% of academic and public respondents, respectively, indicating that they are ‘not at all dependent’ on it as a strategy. It will be interesting to see if attitudes change in the coming years.

An obvious consideration for further study is the fact that the *quantity* of files held will vary among individuals as will the nature and variety of the files; and quantitative variables might be expected to determine the need for personal information management to a significant extent.

Table 43: Ease of finding computer files

Column percentage

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
How easy or difficult do you find it, generally, to find your computer files when you need them?	Very easy	380	41.3%	832	44.4%
	Quite easy	510	55.4%	973	51.9%
	Quite difficult	29	3.2%	60	3.2%
	Very difficult	1	0.1%	9	0.5%
	Total	**920**	**100.0%**	**1874**	**100.0%**

A large majority of both samples (96.7% of academics and 96.3% public) perceive that it is ‘quite easy’ or ‘very easy’ to find files when needed.

If we accept this at face value and reflect on the large variation in file naming (or renaming) and folder naming conventions, the differences in version control and in the frequency of tidying up their computer files, it seems very likely that different styles of personal information management work for different people. Perhaps the concept is better understood as an ecology of preferences and processes rather than as a ‘right way’ to organise digital lives. This, of course, has substantial implications for training programmes and goals.

The apparent ease of locating files may seem counterintuitive bearing in mind the observation that serious data loss was commonly reflected in an inability to find the file (Table 14), which alludes to poor personal information systems. However, there is possibly a temporal factor at play. If most attempts at finding are directed at files that have been **recently** created, acquired, amended or simply viewed, problems of finding may arise when more time has passed, with memories becoming less reliable: bearing in mind the preference for relying on memory of filename and of folder or location. This interpretation suggests that necessary improvements in personal information management lie exactly in the archival realm of making fuller use and reuse of files that were last handled some time ago[21].

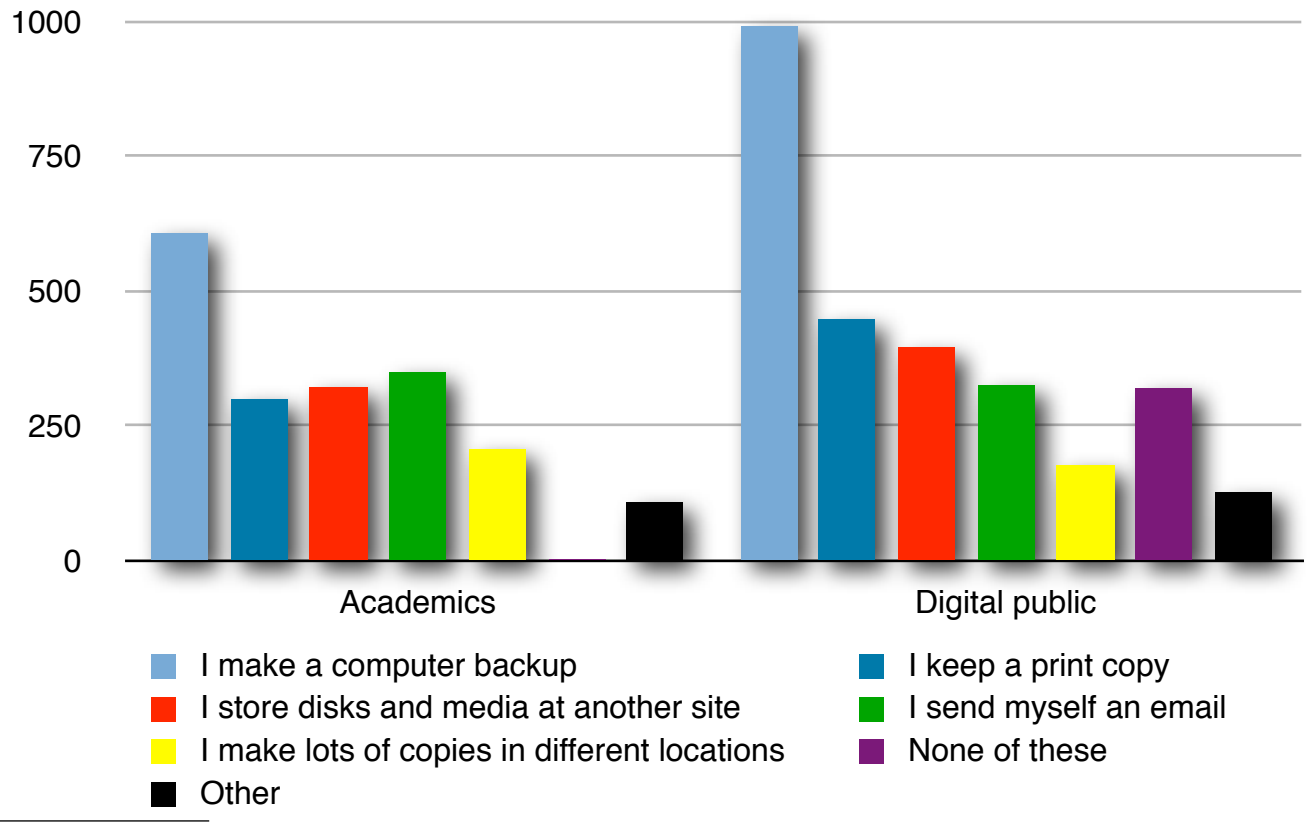
Your personal archive of computer files (all respondents)

The remaining four tables are directed at the participants' attitudes towards the notion of a sustainable personal archive.

Table 44: Strategies for preservation for the future*

Overlapping categories, more than one response possible

		Academic or digital public			
		Academic		*Digital public*	
		n	%	n	%
Which of the following strategies do you use to ensure the preservation of important computer files for the future?	I make a computer backup	609	72.1	994	54.4
	I keep a print copy	301	35.6	449	24.6
	I store disks and media at another site	323	38.2	396	21.7
	I send myself an email	351	41.5	327	17.9
	I make lots of copies in different locations	208	24.6	178	9.7
	None of these	2	0.2	321	17.6
	Other	110	13.0	127	7.0
	Number of respondents	**845**		**1,826**	



[21] The concept of recency is well known to psychologists: further research in the present context is merited

On the face of it these results point to a dire situation regarding the future preservation of personal digital archives even in the short to medium term, with nearly half of the digital public population not even backing up their files. Academics are much more likely than members of the public to ensure that their computer files are preserved at least for the immediate future but in the long run the situation is little better.

An option was available for respondents to indicate any other approach adopted. It represented an opportunity for respondents to mention best practice digital preservation solutions (such as the explicit use of, or early conversion to, interoperable file formats) if they were aware of them. In fact obvious awareness was not revealed: notwithstanding the making of several copies and storing of information at separate sites by numbers of respondents.

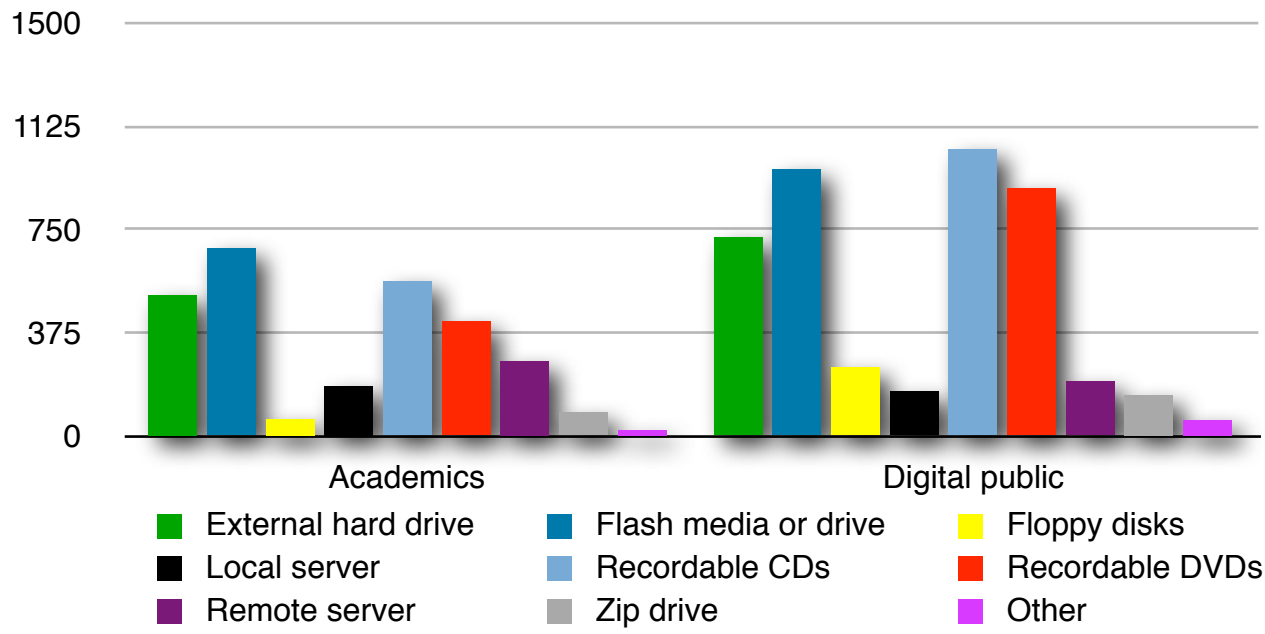
The use of printing by 35.6% and 24.6% of academic and digital public respondents, respectively, may point to the continuing lack of confidence in the sustainability of digital information as well as the convenience of paper. It supports the widespread view that curators can expect personal archives to be hybrid for the foreseeable future.

The findings also underline the continuing prominence of email systems in academic information management practices (as highlighted in the literature review too).

Table 45: Storage media used at home*

Overlapping categories, more than one response possible

		Academic or digital public			
		Academic		*Digital public*	
		n	%	n	%
Which of these storage media do you use?	External hard drive	513	60.0	722	38.5
	Flash media or drive eg USB stick	683	79.9	968	51.7
	Floppy disks	64	7.5	251	13.4
	Local server	184	21.5	163	8.7
	Recordable CDs	563	65.8	1045	55.8
	Recordable DVDs	417	48.8	899	48.0
	Remote server	274	32.0	199	10.6
	Zip drive	87	10.2	150	8.0
	Other	24	2.8	60	3.2
	Number of respondents	**855**		**1873**	



For academics the media that feature most commonly were flash media (such as USB sticks), recordable CDs, external hard drives and recordable DVDs; the same media feature in the top four for the digital public too although the percentages are lower (see Table 45) and yield a different order of likelihood.

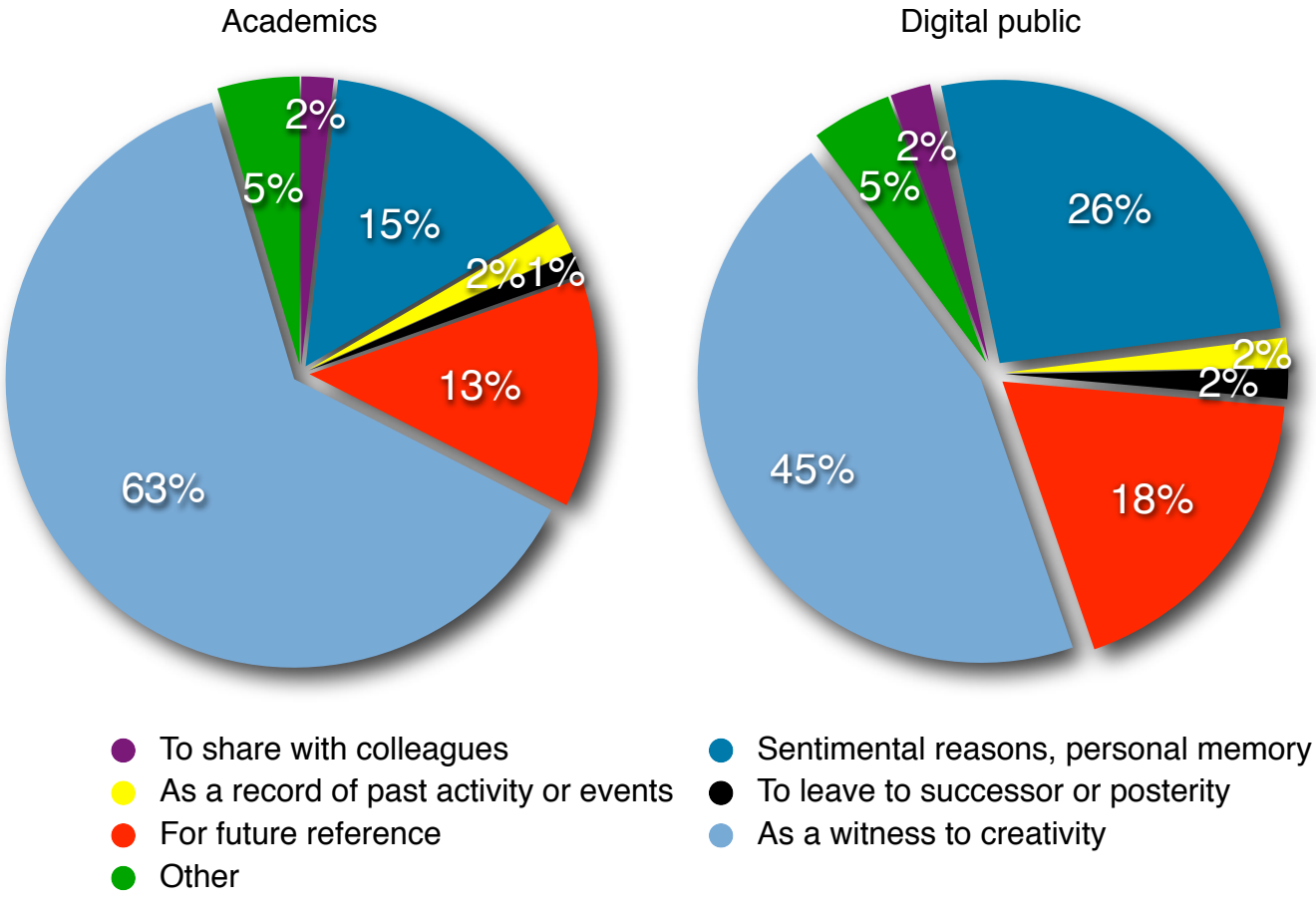
Of these four the one that involves magnetic media, namely the external hard drive, is used by 60% of academics and much more than by people generally with 38.5%. Possibly the academic has more files as well as a more prevalent professional imperative. Future questionnaires should seek to quantify the personal digital holdings.

Academics make greater use of server technologies, which is not surprising. The continuing use of floppy disks by 13% of the digital public is indicative of persistence in the use of obsolete technologies even if used primarily for information transfer.

Table 46: Reasons for archiving computer files*

Column percentages

		Academic or digital public			
		Academic		*Digital public*	
		n	%	n	%
What is your *main* reason for archiving computer files?	To share with colleagues	17	1.8	43	2.3
	Sentimental reasons, personal memory	137	14.8	482	26.3
	As a record of past activity or events	16	1.7	31	1.7
	To leave to successor or posterity	12	1.3	32	1.7
	For future reference	119	12.9	334	18.3
	As a witness to creativity	582	62.9	823	45.0
	Other	43	4.6	85	4.6
	Total	**926**	**100.0%**	**1,830**	**100.0%**



This Table shows some of the most interesting findings of the surveys. For both the digital public and academics, the three most prominent explanations offered for archiving files are (in descending frequency): (i) for witnessing creativity; (ii) for sentimental and personal memory; and (iii) for future reference.

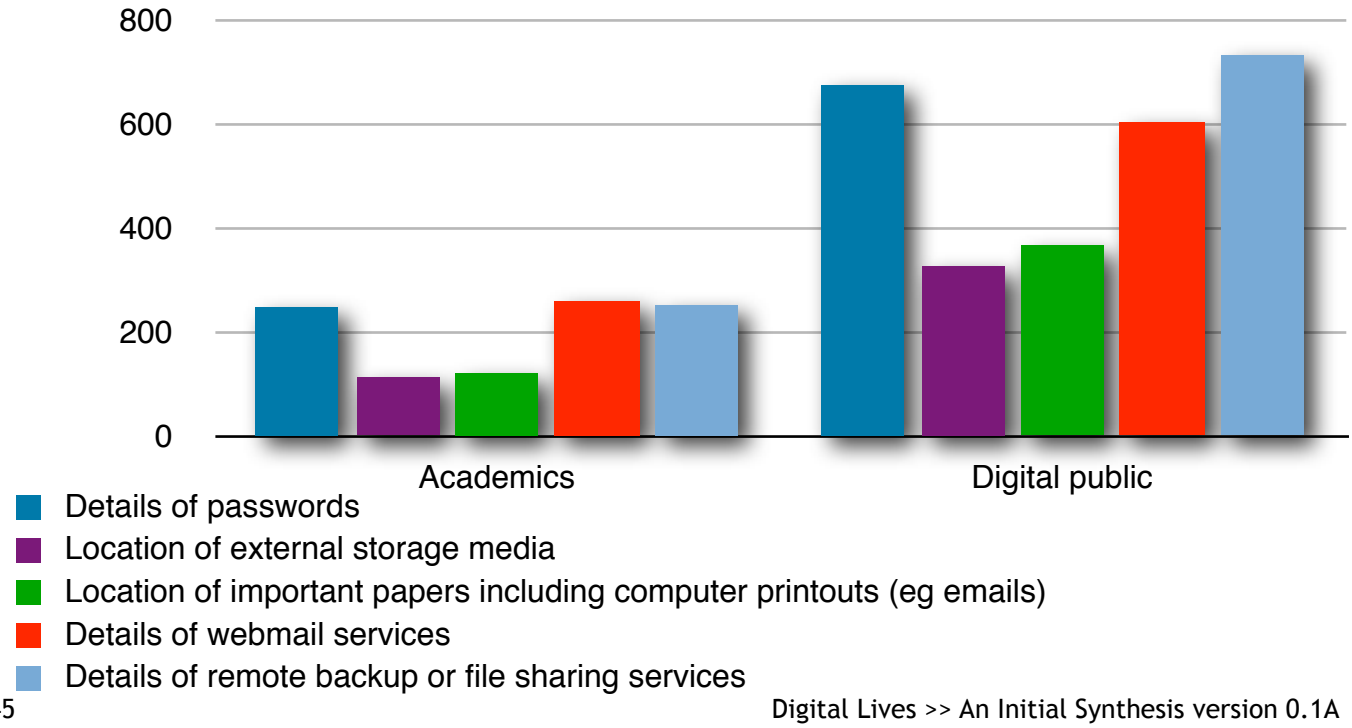
That sentiment and personal memory feature in the list is expected but nonetheless important. That academics tend to value the witnessing of creativity above all else might very well be anticipated. Nonetheless it is interesting that this explanation is so conspicuously shared as the most important reason by members of the digital public as well as academics. It might speak for creativity as a core human value and need or as self-validation. This is an informal, exploratory survey; interpretation requires care. It would be good to follow up this finding.

By contrast, less than 2% of people indicated a desire to leave files to a successor or posterity as the principal reason for archiving computer files. This can be compared with Table 26 where 'Interest to future historians' was prominent. In examining the findings in both Tables 26 and 46 it is important to remember that these data are *relative*, focussing simply on the *primary* value. Table 46 supports Table 26 in corroborating the considerable diversity of reasons indicated by respondents for attaching value.

Table 47: Arrangements for sudden death or incapacity*

Overlapping strategies, more than one response possible

		Academic or digital public			
		Academic		*Digital public*	
		n	*%*	*n*	*%*
In the event of sudden death or incapacity (you fall under a bus), do you have any of the following arrangements in place so that family or archivists could continue to have access to your files? (those responding 'yes')	Details of passwords	249	23.6%	676	35.4%
	Location of external storage media	114	10.8%	328	17.2%
	Location of important papers, including computer (eg email) printouts	122	11.6%	367	19.2%
	Details of webmail services	261	24.8%	604	31.6%
	Details of remote backup or file sharing services	253	24.0%	733	38.4%



This interesting set of results requires further investigation and follow up: in each category, far fewer academics admit to having made arrangements in the case of sudden death or incapacity than members of the digital public. This finding should be of some concern to archivists and curators.

Verbatims

The questionnaire closed with an open-ended question in three parts. The verbatim answers will be analysed for a later paper. Here the answers are presented simply as Wordle visualisations[22].

If you could have a wish list come true, what would make your digital life easier?

[22] See http://www.wordle.net

(1) Organising your digital life more effectively...

(2) Finding your computer files more easily...

(3) Preserving your computer files for the longterm future...

The survey, in fourteen points

(1) Serious loss of data affects academic and nonacademics in much the same way.

(2) Inability to find files is the main cause of serious data loss, not hard drive failure.

(3) Members of the digital public exhibit greater levels of anxiety about computer security than academics.

(4) Academics are more likely to take steps to safeguard data following the purchase of a home computer.

(5) Nonetheless overall the situation is dire from the perspective of longterm archiving of the files; and there was very little obvious awareness of digital preservation practices, even though some people were conscious that the production of multiple copies of files stored at different sites is a sensible precaution.

(6) There is a wide diversity of practices in version control, and in the retention of interim versions, drafts (digital or hard copy). Academics are more likely to keep interim versions than members of the digital public.

(7) Most people use some level of uncontrolled subject description when they name their files.

(8) Practices vary widely with regard to organising files. Few people confess to ‘rarely’ or ‘never’ systematically organising their computer files.

(9) People rely heavily on personal memory for finding their computer files.

(10) There is wide variation in personal information management practices: different things work for different people.

(11) Email is a prominent element in academic personal information management practice.

(12) Very few people, academics or members of the digital public, have arrangements in place to ensure continuing reliable access to their data in the case of sudden death or incapacity.

(13) The most frequent type of precious file for academics was the ‘word processed document’ followed by ‘photographs or digital art’, whereas it was ‘photographs or digital art’ followed by the ‘text-based documents other than word processed documents (eg PDFs)’ for the digital public.

(14) Flash media have become very popular very quickly while floppy disks are correspondingly less popular, and yet these retain a foothold with over 1 in 10 members of the public still reporting the use of them.

Chapter 4: Legal and Ethical Issues

4.1 Aims[23]

This component of the project sought to provide a description and analysis that would help enable repositories to engage with, and encourage an expansion of, personal digital archives while adhering to the public values that laws and ethical principles attempt to protect.

In particular it aimed: (i) to look at the practical application of existing law in the United Kingdom to personal archives; and (ii) to consider the influence of prevailing perceptions of the benefit and value of online sharing and content creation.

4.2 Method

The research entailed a review of the academic literature along with an online examination of past practices and policies of repositories combined with feedback from key user groups on the legal and ethical issues.

A legal and ethical workshop took place at the British Library Conference Centre in June 2008: a morning session was attended predominantly by archivists and curators; and an afternoon session was attended by a more mixed group including researchers and historians.

Morning

- Guy Baxter, Victoria & Albert Museum
- Frances Harris, British Library
- Arwel Jones, National Library of Wales
- Jack Latimer, Community Sites
- Hannah Little, HATII, University of Glasgow
- Kathleen O'Riordan, Media & Film, University of Sussex
- Susan Thomas, Bodleian Library, University of Oxford
- Dave Thompson, Wellcome Library
- Lynn Young, British Library

Afternoon

- Else Churchill, Society of Genealogists
- Maxine Clarke, Nature Publishing Group
- Luke McKernan, British Library
- Helene Snee, University of Manchester
- Tilli Tansey, Wellcome Trust Centre for the History of Medicine
- Lynn Young, British Library

4.3 Findings and Analysis

The current scenario

(1) The increasing ability to create, acquire, retain and share digital data has legal and ethical implications for both individuals creating personal digital archives and for organisations that seek to host such collections.

---

[23] This chapter is based almost entirely but not entirely on: A. Charlesworth (2009) Digital lives >> legal & ethical issues, Digital Lives Research Paper, 14 October 2009, http://www.bl.uk/digital-lives/index.html. The chapter was compiled by Jeremy Leighton John who is responsible for any misapprehensions

(2) Digital technologies allow individuals to create immeasurably more personal information content – user-created content - than ever before, and through web 2.0 services such as Flickr (photos), Facebook (social networking) and YouTube (videos) to share it, on a scale hitherto feasible only for publishing houses and governments.

(3) There is a potential for discordancy between the law and public expectations in areas of intellectual property, privacy and confidentiality. Archives and repositories need to develop strategies to avoid legal actions, and to maintain a balance between the research imperative and the rights of individuals.

(4) With emerging digital technologies and their social and economic consequences, legal systems are presented with new, often unanticipated, scenarios for which they were not designed.

(5) The availability of 'amateur' user-created content such as photos has implications for professionally produced content and for associated services and repositories. The phenomenon has also resulted in individuals bypassing – not entirely effectively – the process of peer-review in publications.

(6) Individuals who place content on sites such as Facebook may choose not to restrict access to family and friends or be unaware of the possibility or the implications or simply not care enough to learn how to switch on the restrictions, or to do so properly.

(7) The combination of almost limitless quantities of user-created content and a lack of clarity about the users' real intentions significantly complicates the process of negotiating satisfactory legal and ethical outcomes.

(8) One approach that is aimed at simplifying the processes of use and reuse by others is the Creative Commons movement directed at reducing copyright barriers to sharing through widely applied and recognisable licenses. The extent to which this system will be understood and taken up and used properly by members of the digital public remains uncertain. There are existing cases of misunderstanding.

Commercial online service providers

(1) The phenomenon of online user-created digital content is driven by four factors: (i) digital content is more easily and inexpensively created in large quantities than analogue; (ii) digital content can be more easily accessed, backed up and shared by the creator through web 2.0 services; (iii) the digital content accessible through web 2.0 services is more readily used effectively by others than is the case with analogue content; and (iv) the commercial online service providers are able to generate revenue (eg through advertising) from the traffic to the site.

(2) Commercial online service providers have adopted alternative strategies for dealing with legal and ethical issues such as copyright and privacy, and have successfully encouraged new attitudes towards them.

(3) An especially profound change has been the essentially reactive approach to legal issues adopted by commercial online service providers such as YouTube and Flickr. The onus is placed on those individuals who upload content to these services to ensure that doing so is legal. If complaints are received then the material is taken down either permanently or pending

investigation. The approach has two advantages for the online service provider: (i) most material will not provoke legal claims; and (ii) most claims will be readily addressed.

(4) A potentially significant risk with the reactive 'notice and takedown' approach stems from matters of privacy. A specific complication is that individuals who are careful about their privacy in one context are much less careful when it comes to blogs, photos and videos uploaded onto social networking sites and the like.

(5) Another hazard for 'notice and takedown' is that some holders of commercial rights are questioning its legality and in particular its effectiveness in protecting their intellectual property.

(6) The legal situation regarding privacy and data protection is even less clear than is the case with copyright because: (i) legal examination of copyright has been strongly motivated by rights holders, while (ii) there appears to be little appetite shown by the courts or parliament in the United Kingdom to examine the impalpable boundaries of privacy.

(7) Commercial online service providers seek to maximise revenue generation, and this means it is typically possible for them to tolerate: (i) the loss of content through its voluntary or forced withdrawal; (ii) uncertainty about authenticity; (iii) lack of context and metadata.

(8) In comparison with public repositories, commercial entities can absorb occasional legal actions and loss of reputation, and can focus on just one specialised service.

(9) Under existing circumstances there is no legal obligation on the commercial online service providers to retain and preserve material for the longterm and to maintain the highest standards and practices for storage and access.

Public repositories

(1) The full implications for repositories and their ability to collect, hold and make available the elements of personal digital archives are difficult to determine at this time.

(2) There are clearly aspects of repository functionality and service that might be provided by public repositories. Namely: authenticity, integrity of content, metadata and context, longterm security and preservation.

(3) Two potential roles for public repositories are: (i) promoting the capture and preservation of personal digital archives directly from a diverse selection of originators; and (ii) gaining access to user-created content held by commercial archives and services. In both cases a key facilitator will be the provision of advice and systems for addressing legal issues.

Legal requirements and risk

(1) Broadly, the legal requirements are: (i) copyright and intellectual property; (ii) data protection and privacy; (iii) freedom of information and openness; and (iv) liability, including defamation, contempt of court, obscenity and indecency.

(2) In the face of uncertainty and complexity it may be tempting for repositories to choose to be more cautious with digital content than with analogue materials of a similar type of content. The novelty of applying legal and ethical principles to digital content may mean that they end up being treated more conservatively.

(3) Regarding the role of public repositories in collecting a large and wide range of eMSS, the experience of web archiving is of some relevance. The process of obtaining permission from rights holders proved to be a considerable burden that has severely curtailed the archiving of the United Kingdom's web space - yet the pragmatic approach adopted by commercial online service providers indicates that the legal risks are not necessarily high, and to date have in fact been relatively low.

Key legislation

>> British Library Act 1972

>> Legal Deposit Act 2003

>> Council of Europe, Draft Recommendation: A European Policy on access to archives

>> EU Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market

>> Copyright, Designs and Patents Act 1988 (unofficial consolidated version)

>> Data Protection Act 1998; The Data Protection (Processing of Sensitive Personal Data) Order 2000

>> Electronic Commerce (EC Directive) Regulations 2002

>> Freedom of Information Act 2000 (England & Wales)

>> Contempt of Court Act 1981

>> Criminal Justice Act 1988; Criminal Justice and Immigration Act 2008; Criminal Justice and Public Order Act 1994

>> Defamation Act 1996

>> Obscene Publications Act 1959; Obscene Publications Act 1964; Protection of Children Act 1978

(4) Risks need to be characterised based on the likelihood of legal action taking place, the consequences, and the degree to which it can be tolerated. The managing of risks and the securing of licenses need to be matched by efforts to encourage individuals with personal digital archives to engage with a repository.

(5) The strategy for repositories should be to keep licensing and risk amelioration procedures simple and transparent for those wishing to deposit or to access, use and reuse eMSS. Users should not be expected to spend inordinate amounts of time establishing what is or is not permitted, and potential depositors should not be deterred unduly from engaging with repositories.

(6) A balance has to be found in accepting and rejecting risky materials. Significant losses to the historical record can be later regretted with changes in attitudes to matters such as obscenity. At least the tolerated level of risk for digital archives should not be less overall than for their analogue counterparts.

(7) There is a need to anticipate or least plan for future changes in laws and ethical attitudes. It is necessary for there to be a flexible framework for legal policy that allows coherent

change management and longterm sustainability. It needs to be grounded in risk assessment, as well as clear and evident processes for deposit and access management, and for audit control and risk amelioration.

(8) Risk assessment needs to take place early in the lifecycle.

(9) Remedial measures and policies need to be developed.

(10) In some cases repositories will already have in their holdings digital material without licenses covering the interests and rights of third parties for example. It might not be known whether a particular letter or a particular photo was created during employment hours for example. A conservative way to deal with copyright is to make the digital object available on a single computer system in a reading room (in other words there is effectively only one presentation of the object being made available at given time). In the cases where there is clearly no privacy and data protection concerns, the possibility and risk of making available a digital object more widely, simultaneously in more than one reading room for instance will need to be conducted on a case-by-case basis. Could it be justified with copyright 'orphans', for example?

Legal reform

(1) With the anticipated emergence of legal deposit for a wide range of digital materials, a reappraisal of existing policy towards personal digital objects would be timely. An impact assessment of existing and proposed legislation and future research might be helpful for legislators.

(2) Considering personal collections generally, and in particular the widely available and public components (in contrast to the personal archive elements that are restricted to the individual, family and friends), a legal deposit scheme that permits collection with significantly reduced risk to public repositories, for instance, should be pursued.

(3) Attempting to access a creator's social networking account following his or her death might conceivably subject the curatorial institution to legal risk even if the permission of the family has been obtained. The process of legal reform might address this issue too.

Ethics

(1) Organisations may choose to hold higher ethical standards than the law strictly requires.

(2) Established notions of privacy and intellectual property have been weakened by online social networking, 'reality' television programmes, and the collection and use (and loss) of personal data by governmental organisations. Repositories and other institutions are consequently under pressure due to changing public expectations of access, use and time to availability.

(3) On the other hand, occupying a higher moral ground might be one way for a repository to be more attractive to depositors, engendering greater trust.

(4) From an ethical perspective, the interests of third parties represent a critical responsibility. In addition to considering the intellectual property rights of third parties who have been included in appended email messages or the privacy rights of those who are discussed in communications, there is the need to consider from the ethical standpoint the

possibility of real psychological harm and social damage to reputation through access to archival materials.

(5) A repository may need to consider the possibility of imposing ethical constraints on its users through access agreements or even ethical committees; and to put the onus on users in ensuring responsible research and reporting.

(6) Collaborative agreements among public repositories in maintaining agreed ethical standards would make interchange of materials much more straightforward, and provide consistent and widely understood principles.

Practical procedures and suggestions

(1) Repositories wishing to accession personal digital archives need to: (i) assess the risk; (ii) establish internal policies; and (iii) prepare guidelines. Most important is the need for clarity in the emerging solutions and practices.

(2) Metadata creation for 'high profile' personal digital archives might be feasible but bespoke processes are less feasible for handling large numbers of personal digital archives. One possibility is for depositors themselves to provide basic metadata regarding the legal status of their objects, motivated by using: (i) a straightforward system of annotation, (ii) guidelines and advice, and (iii) the existence of direct incentives (eg metadata being a requirement of acceptance).

(3) A metadata scheme might be devised that is similar or allied to the Creative Commons project with its standardised icons and layered meanings (designed to be read by people of varying expertise and by machines).

(4) 'Layering' is a technique increasingly used in data protection circles for privacy policies, and also by the Creative Commons for its licence scheme. The concept of 'layering' is to prepare explanations of policy, deposit agreements, licenses and end-user agreements at several levels of detail, technicality and length. The layers are directed at different audiences ranging respectively from plain language for the lay reader, more detail for the interested person, and full detail for the expert; there may also be a layer of machine-readable digital code directed at search engines and other applications. It allows the communication of a suitable form of information to a specific audience.

(5) Repositories could target the creators and collectors of personal digital archives with the promise of services such as advice about the legal framework and its implications, longterm preservation, archival access and so on.

(6) The development of software that automates or facilitates the process of rights compliance through the creation of metadata and the delivery of appropriate security measures, needs to be actively encouraged: mapping tools that help in the organising of personal digital archives (see §3 and §6), and in the compiling of components of these archives that reside in a fragmented way across diverse services.

(7) An operational relationship between users, repositories and commercial services could provide benefits for all parties. This would enable the harvesting and preservation of digital content for creators, and its archiving and future access for researchers, while providing commercial services with a potentially competitive endorsement by official or public repositories.

4.4 Summary Observations and Recommendations

(1) The laws relating to personal digital archives are in many respects little different from paper personal archives. The difference lies in how the spirit of the law is to be interpreted by archivists, researchers, the public and the courts at a time of radical change in digital society, reflected in online social networking and the widespread sharing of user-generated content.

(2) It is often implicitly assumed that archives and repositories must be directly responsible for legal requirements and adopt a gatekeeper role. It may be more apt in many situations for repositories to relinquish some control and serve as a facilitating guide or broker.

(3) New technologies are readily embraced by repositories. This openness needs to be extended to: (i) adopting new ways of operating cooperatively, forming new alliances and interactions with depositors, users and other organisations; and (ii) embracing new mechanisms for legal compliance.

(4) Digital and paper should not be automatically treated differently; rather the benefits and risks associated with retaining and making available personal papers and personal digital archives need to be assessed. It will be interesting to see what effect the social networking ethos has on the way paper archives are treated.

(5) In managing risk, it is likely to be helpful to coordinate with other repositories, bearing in mind the need to balance a desire for access with the concerns of depositors, and not to deter depositors unduly.

(6) The law is slow to change but it does change and it will need to do so repeatedly as more and more technologies and social innovations emerge. Repositories in turn need to be responsive.

(7) Archivists should adopt a pragmatic approach to the legal risks inherent in the collection and preservation of personal digital archives, an approach based on effective risk management rather than any impractical attempt to avoid risk entirely. The risk (reputational and financial) can be managed through: (i) risk assessement, (ii) measures for reducing risk, (iii) clear guidelines that are understood by all, and (iv) grievance procedures. Of particular importance is the need for convincing and actual efforts to respond to individuals quickly and appropriately. These will only work if these processes are clear, efficient and timely.

(8) The framework for legal risk policy needs to be flexible and be able to put in place coherent change management as the legal and attitudinal environment changes. In short, a highly dynamic risk management system that is responsive to change is required.

(9) The development of suitable tools and standards for legal processing and metadata should be encouraged. In part this requires agreement between archives and umbrella organisations in agreeing metadata standards along with collection and deposit policies

(10) Archival repositories should explore possible strategic relationships with online service providers in coming to mutually beneficial legal arrangements to safeguard digital content for future research.

(11) Future work should explore and determine suitable channels of communication and advice for legal interpretations and details.

(12) It will be helpful to consider lessons from novel approaches such as Creative Commons, with layered licenses and use of icons.

(13) Archivists need to establish procedures that produce little legal risk to members of the digital public, and be proactive in demonstrating that this is so while showing how personal digital archives can enable socially beneficial research.

(14) Archival organisations should consider how legal reform (including legal deposit) can help to enable the sustainability and use of personal digital archives, and to arrange for this understanding to be passed to legislators. (i) In particular, there is a need to make it possible for public repositories to collect and care for personal digital archives without undue legal risk, and, for example, to act according to the wishes of individuals (living or deceased) to obtain personal content from online service providers without inappropriate risk. (ii) It should also be possible for the larger organisations to delegate phases of the archival lifecycle to smaller ones subject to the maintenance of appropriate standards - in the case of the publicly available websites for instances.

Chapter 5: Users

5.1 Aims[24]

The aim of this component of the research into digital archives was to engage with users with expertise and experience in research activities involving personal archives.

Two specific objectives were: (i) to explore the needs and views of potential users of these collections such as historians and biographers; and (ii) to obtain opinions of users regarding how curators should meet user requirements.

5.2 Methods: User Forum Organisation

The principal research technique adopted was the holding of a one day User Forum in January 2008 at the British Library Conference Centre.

Invited members of the user community included representatives of professional and learned associations, specialist intermediaries, and selected individual scholars with relevant research expertise.

Participants (in addition to Digital Lives team members):

- Martin Campbell-Kelly, history of computing, Professor in Computer Science, University of Warwick
- Maggie Ferguson, biography, Royal Society of Literature
- Lorna Hughes, digital humanities and eScholarship, AHRC research methods network
- Anna Mayer, history of science and personal papers, National Cataloguing Unit for the Archives of Contemporary Scientists at the University of Bath
- Anne Sebba, biography and journalism, authors' group PEN
- David Shaw, printing history and drafts, Secretary of the Consortium of European Research Libraries (CERL)
- Boni Sones, contemporary political history and journalism, Women's Parliamentary Radio
- Lynn Young, archives, Corporate Information Management Unit, British Library

In the morning there were presentations from researchers and curators, outlining how modern personal digital archives of scholarly interest are being created, managed and disseminated; also elaborated were some of the significant differences from the past that have become evident in terms of format, content, intellectual property rights, and volume, and the possible implications for curators and users. The presentations were followed by questions and a general discussion.

In the afternoon, the participants formed two breakout groups and were asked to discuss and identify:

- generic user needs for 'personal collections' collected by research repositories, and the underlying requirements that remain unchanged;

[24] This chapter is the first public report on the Digital Lives research involving the users of personal archives. It is prepared by Jeremy Leighton John based almost entirely on a draft report on the User Forum compiled by Katrina Dean

- emerging key changes from the users' perspective for digital collections, and how requirements can be factored into new approaches, tools and services;

- any differences that apply between distinct collection areas such as oral history, history of science, and literary manuscripts and correspondence; and

- ways in which curatorial selection, preservation and access for personal digital archives should be conducted.

The day concluded with a session of reporting back and further discussion.

5.3 Outcome: Emerging Questions and Observations

The various points made during the day are grouped under the following eight headings.

Creators

(1) How should potential creators of important digital collections be identified and assisted bearing in mind the benefits of early intervention regarding digital capture and preservation?

(2) How can writers generally (and not just the most recognised ones) learn to look after their own archives? There is a need for institutions to provide advice to wide communities of creators.

(3) Writers would welcome guidance and general advice in deciding what to keep. On the other hand, traditionally there has been little involvement by curators regarding the nature of what a creator chooses to keep.

Personal digital objects and authenticity

(1) The convergence of digital formats across collecting areas has implications not only for collecting policies but also resource discovery.

(2) With email, people in general are not communicating in the same way, and are not writing emails with literary care and thoughtfulness; on the other hand, emails replace to some extent the unrecorded telephone conversations.

(3) To put the digital era in perspective and to address a fear that content is being lost, it was noted that there has always been a dearth of material. One participant suggested that most people are not interested in keeping a personal archive.

(4) The authenticity of digital objects and content is of special concern. How can the authenticating information itself be authenticated and provided with a secure provenance?

(5) Accounts based on electronic sources may be an improvement on unsubstantiated accounts based on unnamed, oral sources. Emails, for example, automatically record contextual information. On the other hand, emails can be 'doctored'.

Users

(1) The very large volumes of digital material point to the inevitable reliance on machine reading or trawling in the first instance to help researchers identify information of primary interest.

(2) Users are concerned with having: (i) speedy and effective access to both the archival objects and to metadata (indexes, catalogues, and cross-references); (ii) finding tools that can be used with large archives; (iii) information that is demonstrably authentic; and (iv) the design elements of digital resources (especially websites) captured for future scholars.

(3) One participant anticipated access to an archive as a 'digital dump' combined with advanced searching functionality and finding aids, and would prefer this option to prior selection by curators. There is, however, a corresponding issue of privacy and the detection of instances that require curatorial decisions. Copyright may also affect the feasibility of retaining a 'digital dump'.

(4) One user noted that there is still a perception that collecting institutions and repositories present a closed culture, and that the digital environment will precipitate a culture of more open access.

(5) How much technical support will repositories be able to provide, in enabling access to obsolete digital objects and media?

(6) The characteristics of repository users are changing. There are now more young people using the British Library for research purposes, and so curators and archivists need to be aware of this change and tailor their services accordingly.

Institutions and processing

(1) A significant potential cost in dealing with personal digital archives lies in the examination and decision making necessary in order to address privacy concerns.

(2) Will curators simply be dealing with the originators of an archive or will they need to interact with third parties too – people such as those who write to the originator?

(3) Considering parallels with paper personal collections, the relationship between writer and publisher has long been very important in shaping what was published and how it was edited. What are publishers keeping today? Traditionally, publishers were involved in the end of the research process; now some of them are trying to integrate themselves into the workflow of scientific researchers – by providing scientists with information services to help them obtain research grants, for example. Correspondingly, archival institutions could be encouraged to provide research grants to enable study of digital archives or to improve finding tools or to help in process of making the archival content available.

(4) The continuing role of universities in the future needs to be contemplated. In the past universities have been responsible for the storage of research information, which has not always been well preserved and has subsequently required the use of data recovery techniques.

(5) Participants emphasised that the organising of institutions according to media format does not make sense when dealing with hybrid collections. It was stated that splitting collections along lines of media might weaken provenance to little benefit. Instead curators and archivists and scholars and researchers need to operate with a multiformat approach that cross-references and indexes a holistic system.

(6) There is a possible role for repositories in providing thought leadership across a range of new issues affecting personal digital archives.

Rights and markets

(1) There is a need for greater clarity and understanding regarding ownership of digital objects and digital rights: from email messages to the content of social networking sites.

(2) With regard to the ownership of literary works, the focus may be moving away from the objects to the rights.

(3) What is the monetary value of digital archives? Will the market move from the sale of objects to the sale of rights? What will be the role of manuscript dealers? Who will establish the market?

New skills, new sources

(1) New skills will be required of researchers. Examples include the interpretation: (i) of body language in moving image, and (ii) of new forms of text analysis in authorship attribution. Correspondingly, an understanding of computer tools will be needed for effectively using and analysing multimedia; eg documenting artistic performances.

(2) Digital moving image and voice will become increasingly important resources along with blogs and websites generally. There are also informal oral history accounts that are not part of an archive and have not been conventionally transcribed, but will be accessible in more and more detail with increasingly sophisticated finding aids that use voice recognition and automated indexing.

Fakes

(1) Public relations representatives are being employed in some cases to 'clean-up' an individual's record on the world wide web, in order to protect or enhance the person's reputation.

(2) Deliberate attempts to create misleading evidence through the creation and propagation of digital content of various kinds have occurred. Lobbyists have been engaged to establish fake 'grassroots' campaigns. Similarly, there are fake blogs, and there are also cases of blogs being funded by organisations (eg political) that deny responsibility.

(3) It will be the role of curators and scholars to identify the authentic and the fake. There has been a longstanding historical and research interest in 'fake' personas and the like.

Online service providers

(1) The move to online services suggests that personal digital objects will be stored remotely in the future. Will less, little or none be stored locally at home? What are the implications for creators and curators? How would this change impact personal digital preservation? How will digital estates and executors deal with personal archives held by online service providers?

(2) Commercial service providers are often less risk averse - in the area of data protection, for example - than public bodies, which feel more restrained by regulations.

(3) Use of online services can be accompanied by the loss of some digital rights. This phenomenon may result in new players entering the market to purchase archives of commercial value in the future. It highlights the need to examine the terms of service and other legal information on the websites of online service providers, and to provide legal advice and awareness.

(4) Online service companies will often aggregate information on individuals for analysis and for identifying marketing profiles. This information is preserved and captured for marketing but is potentially useful for researchers of social trends and unofficial history, for example. It is possible in an electronic environment not only to capture the content of resources (eg health information online) but also the use made of them, as witnessed by computer transaction logs.

(5) At this time authors and creators cannot rely on others to secure their eMSS: in particular, many commercial services will not necessarily retain an individual's digital content or keep making it accessible. In these circumstances the onus is on creators to secure and preserve their digital output for themselves. It is a challenge for scholars as well as curators.

5.4 Concluding Remarks

(1) User representatives voiced a wide range of concerns. The examination of the needs and views of users revealed: (i) a deep sense of impatience among users who want access to personal digital objects to be made possible quickly – through automated searching and indexing, and if necessary, at the expense of detailed selection and cataloguing; and (ii) at the same time a strong concern for being able to ascertain in a convincing way the authenticity and provenance of digital objects.

(2) Selection by collecting institutions was considered less important by user representatives. Creators of collections and scholars could share the responsibility for evaluating and describing or annotating material in digital archives.

(3) While the desire for prompt access will not surprise professional curators, it represents evidence - derived from a focus group of diverse and prominent scholars - for the need for access to digital content to be expedited by means of streamlined and efficient workflows and processes; and for resources and effort to be directed at the entire lifecycle.

(4) The responsibility of a curator to ensure necessary privacy for third parties as well as originators and depositors – in the face of burgeoning quantities of digital objects – was acknowledged by user representatives when this issue was raised. These kinds of curatorial issues will require ongoing communication between researchers and repositories, alongside the provision of advice and information to creators.

(5) The requirement for authenticating procedures is to a significant extent met by forensic and related technologies, although these do need to be used properly and in an accountable way. The users' concern regarding this aspect of digital curation supports the attention that has been given by this project to authenticated capture. There is also a clear need to inform the scholarly and user community (as well as the creators and originators) about the available techniques in order that implications for authenticity are understood.

(6) It was anticipated by some user representatives that primary responsibility for managing personal digital archives would likely shift to creators and their representatives, with repositories providing advice, information and services that facilitate these activities. There is a significant risk in relying on commercial entities alone – most especially regarding the loss of rights and access. The impact of new technologies and services on the market for personal digital archives remains to be seen and understood.

(7) It was agreed that there is an urgent need for repositories to engage with three groups of originators and intermediaries in a timely way in order to establish ownership, immediate requirements and ultimate destiny of both the digital object and associated digital rights: (i) individual creators early in their careers, (ii) universities, and (iii) commercial entities, especially online service providers.

(8) It was felt by some user representatives that with personal archives embracing an extremely wide range of media types, some restructuring of the administration of repositories is inevitable in order for personal collections to maintain to fullest advantage their context and provenance as well as enabling - through curatorial oversight - integrated and effective access to the whole.

(9) The user community is embracing new types of users, encompassing to an ever greater extent social and natural scientists alongside historians, biographers and literary scholars.

(10) Curators and institutions could play a leading role in advocacy and thereby seek to facilitate the development and understanding of new research techniques and opportunities.

Chapter 6: Technologies

6.1 Aims[25]

The aim of this component of the project was (i) to identify promising and potentially transferable technologies and tools for supporting the curation of personal digital archives, and (ii) to consider the role of online service providers.

Tools of interest would: (i) assist professional archivists in the acquisition of eMSS, the maintenance of extant digital replicates, the securing of provenance and authenticity, the extraction of metadata, and the conversion of files to interoperable formats; and (ii) enable people generally to safeguard their personal archives.

6.2 Methods: Transferable Technologies

A combination of a literature review and online research was conducted; this was supplemented by direct experience at the British Library as part of the internal Digital Manuscripts project including work with computer forensics, and by consulting colleagues who specialise in information technologies, digital curation and digital preservation including at workshops and conferences, notably the Digital Lives research conference[26].

Special attention was directed at the PLANETS project[27], which is funded by the European Union in order to research and develop digital preservation technologies and services, and which - like Digital Lives itself - is also led by the British Library, offering possible synergies.

With three projects being funded by the European Union (CASPAR[28] and DigitalPreservationEurope[29] along with Planets) as well as the Digital Curation Centre[30], all collaborating under the banner of WePreserve, the situation in the United Kingdom and Europe has been transformed in recent years[31].

---

[25] This chapter is written by Jeremy Leighton John. It is derived from, and supplements substantially, the conference paper: J. L. John (2008) Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives and tools, iPRES 2008 Conference, the Fifth International Conference on Preservation of Digital Objects, the British Library, London, http://www.bl.uk/ipres2008

[26] The First Digital Lives Research Conference at the British Library, Personal Digital Archives for the 21st Century, 9-11 February 2009

[27] Preservation and Long-term Access through Networked Services

[28] Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval, http://www.casparpreserve.eu

[29] DPE, http://www.digitalpreservationeurope.eu

[30] Sometimes known by its acronym: DCC

[31] A. Farquhar and H. Hockx-Yu (2009) Planets: integrated services for digital preservation, International Journal of Digital Curation 2(2): 88-99. The Digital Preservation Coalition is another important, influential and longstanding group representing a number of institutions in the United Kingdom

6.3 Findings: Transferable Technologies

The perspective of manuscripts and personal archives

(1) In order to better understand the curation of eMSS and the challenges presented, it is helpful to first examine the perspectives of the creators, curators and consumers or users or reusers of analogue manuscripts, and to contemplate personal archives in the widest sense.

(2) Manuscript scholars and historians study: (i) information content, its textual, pictorial and symbolic content, subjecting it to textual analysis and iconography; (ii) the original creation of the manuscript and the manner in which it was written (loosely palaeography) as well as the overall organisation of the manuscript including its layout and style; (iii) the way it was protected, bound, contained and stored (loosely codicology); and (iv) associated connexions, chronology and context. (Who was communicating with who and when? Who knew who? Who wrote to who? Who joined what organisation? Who read what, viewed what, listened to what? Who collected what?)

(3) The literary scholar in particular often wants to understand (i) the development and history of an originator's amendments of the manuscript through various drafts until the final version is reached, along with (ii) the subsequent history of the manuscript and the way it is copied and distributed and modified by others.

(4) Some scholars will be interested in the physicality of the object and the way it is handled or manipulated and interpreted by people. Some will be interested in even the smallest fragments of unique information.

(5) Curators are concerned with: (i) scholarly, pedagogic and heritage value and, in short, the usefulness for researchers of the objects in a personal archive; (ii) provenance and authenticity along with ownership and its history; (iii) copyright and intellectual property; and (iv) the original position and contextual associations of an object. The curator has long served as an effective mediator, directly or indirectly, between the originator and the scholar.

(6) There is a long history of scholarship oriented towards forgeries and fakes. Establishing the authenticity of objects that emerge independently of an archive is a particular challenge.

(7) The need for confidentiality and the identification of privacy and data protection issues are of special concern.

(8) Collection care entails the protection of the integrity and stability of the object as well as its conservation (any necessary repair and recovery) and longterm preservation.

(9)The provision of access is facilitated by cataloguing, description and metadata, and the production of finding aids, along with security measures.

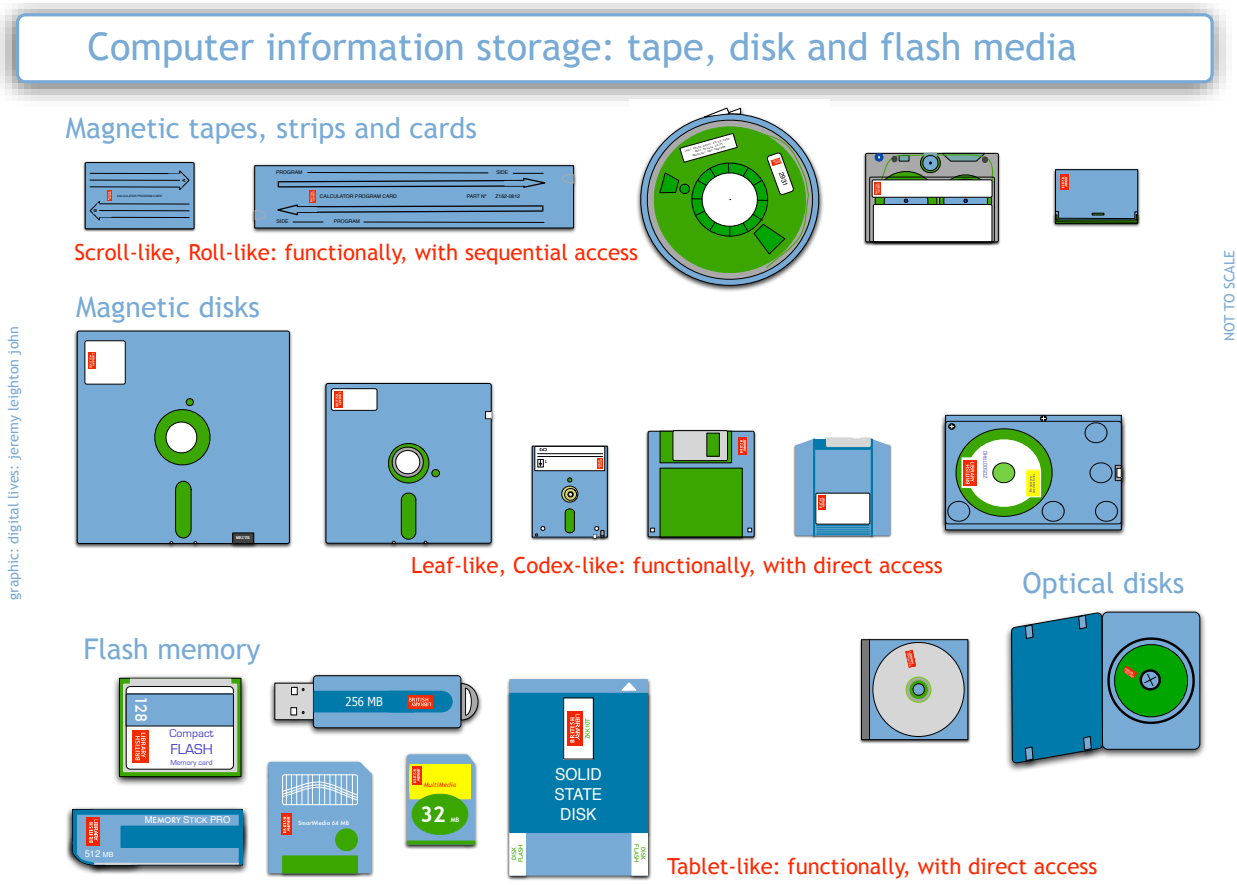
Personal archives in the digital era

(1) There are four key observations that impact the technical aspects of personal archives and their curation: (i) technological obsolescence, (ii) media degradation, (iii) information distributed over networks, most particularly cloud computing, and (iv) encryption.

(2) These are not entirely new phenomena. Even in medieval times scribes sometimes encountered difficulties in interpreting the writing and language of earlier generations; and even in the absence of obsolescence scribes made mistakes or amended texts in the process

of copying them; in addition, an archive or indeed elements of a single manuscript would exist in a distributed state (eg leaves or quires extracted and loaned to another individual or institution). Encryption techniques have been designed and employed for centuries. The difference lies in the pace of technological change, and the reach and power of technology in the 20th and 21st centuries.

(2) Correspondingly, the elements of a personal digital archive reside on media in essentially four critical situations: (i) media and digital object organisation that are obsolete (eg an 8" floppy disk) and at risk of degradation (if not already degraded); (ii) media and files that are not yet obsolete but are nonetheless at risk of obsolescence and degradation (eg a contemporary hard drive recently extracted from an active and working computer); (iii) files that reside on the remote servers of online service providers (eg photos on Flickr restricted to family and friends); and (iv) files that are securely or insecurely encrypted.



(3) A further consideration is the fact that for the foreseeable future the personal archive is likely to comprise analogue and digital objects, and their study by researchers will require an integrated solution. A printout of an essay as a word document may be unique and a digital version of the same essay may likewise be unique representing earlier and later drafts of the same ultimate essay.

(4) There is also the matter of the behaviour and psychology of creators. People have always differed in the extent to which they retain variations and hoard and collect their personal objects. Nonetheless, as the work with personal information management (this project, see §3) has indicated the volumes and nature of eMSS is such that information can be lost in the complexity and course of everyday computer use.

(5) The curation of personal archives essentially consists of three phases: (i) the bringing of the archive to the repository, the acquisition; (ii) the curatorial processing of the archive; and (iii) the making available of the archive to researchers, providing suitable and effective access.

(6) The pace of obsolescence and media degradation, however, puts the initial emphasis of effort on the capturing of the information (the digital capture imperative). In essence the approach is: (i) to copy as exactly as possible the information and transfer it from the original media to fresh, modern media; (ii) to monitor these primary files and pass them to fresh media at regular intervals; and (iii) to retain the original media.

(7) There are two basic approaches to digital preservation and access over the longterm: (i) emulation and (ii) conversion (somewhat confusingly termed 'migration' by the preservation community). (A third is sometimes elaborated, namely encapsulation[32].) Both emulation and migration are advisable and necessary at this time. In the case of emulation, exact replicates of the original files are presented to the researcher using hardware or software emulators. In the case of migration, the files are converted to an interoperable form suitable for use with modern and future (hopefully) computers[33]. Both approaches have advantages and disadvantages.

(8) The advantage of emulation is that the file requires (ideally) no modification and with a high fidelity emulator the look and feel of the original experience can be made available. The disadvantage is that there might be no such emulator especially for the most esoteric systems and objects.

(9) The advantage of conversion is that the files are converted into a form that is interoperable and will hopefully be acceptable to future computer systems. The disadvantages are: (i) that some of the original qualities (sometimes referred to as 'significant properties') of the files are lost and so the look and feel may not be retained exactly; and (ii) that considerable effort is required to undertake the conversion properly.

(10) In both cases there will be some uncertainty about the nature of future technologies, which is a reason for adopting both approaches.

(11) There are three groups of people who might want to engage in digital capture and preservation on some level: (i) professional digital specialists, (ii) professional archivists and curators, and (iii) creators of personal digital archives (from academics to members of the digital public). Of course, researchers and users of personal archives also need to understand

---

[32] Encapsulation involves the retention of the original object that is encapsulated with instructions for its interpretation (eg details of the file format), with this encapsulating information being written in XML for example. Digital objects that are dependent on complex software would require very sophisticated description even involving an executable program. It is an approach which in its most comprehensive form begins to merge into emulation and virtualisation. Digital Preservation Testbed Project (2002) XML and digital preservation, Digital Preservation Testbed White Paper, Den Haag, September 2002, 34 pp.

[33] Migration may take place in several ways: for example it might involve conversion to a standard specified file format (in which case the migration may be referred to as 'normalisation' or, in an even more special sense, 'canonicalisation'); it might take place on demand, when access to the digital object is requested or it might take place as and when new generations of software and hardware arise; see P. Wheatley (2001) Migration - a CAMiLEON discussion paper, Ariadne 29, October 2001; S. Thomas and J. Martin (2006) Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives, Journal of the Society of Archivists, 27(1): 29-56; C. Lynch (1999) Canonicalization: a fundamental tool to facilitate preservation and management of digital information, D-Lib Magazine 5(9)

the processes in order to be able to appraise the personal digital objects encountered (including the steps taken to protect authenticity), but these individuals will often belong to the category of creators too.

Box: File formats[34]

The organisation of a computer file is a fundamental issue for longterm preservation and sustainable use. If the file system is understood - in other words if the way files are organised on the disk or other media is recognised - it is a generally simple matter to copy the file; but the binary information of the file itself still needs to be interpreted. The file format is a model that gives meaning to this binary information: it determines the way elements of the content information are arranged, it provides locations for information about the specific file, and it provides practical information that allows the file to be processed by software in its technical environment so that the information content is expressed in the appropriate way. Unfortunately, in many cases the file format is proprietary or simply not open to the community at large, in which case preservation specialists recommend conversion - 'migration' - of the file into a form that adopts an open file format that is widely or freely understood. This remains a key approach of digital preservation. It can only be a good thing for the preservation and curation community to understand the nature of the file formats and how to characterise and extract information, for this serves capture and emulation as well as migration and, of course, information science and computer history.

An important point for the uninitiated is that the preservation quality of a file such as a PDF or a JPEG often varies depending on its precise nature (eg whether lossy compression has been applied or not). In short one cannot make a recommendation based purely on broad file type category: some forms of PDF are better than others.

Even with a fuller understanding of file formats there are challenges. Firstly, migration frequently will result in a loss of information, in style or functionality, and there is the concern that repeated migration may be necessary over the years and decades (as software and hardware continue to evolve), and consequently there will be a tendency to erode the original information. Secondly, there may be a mismatch between the formal specification of a file format and the demands of the software; if a new version of the software is less forgiving and adheres more strictly to the specification, the digital object may become unreadable by the new version of the software; thus even using apparently suitable software may be no guarantee. Thirdly, the fact that many digital objects simply do not comply with the specification for the file format or may be mildly corrupted, raises the question of what to do about valuable but technically invalid files. Fourthly, remarkably little is known about the robustness of file formats: their susceptibility in the face of quantifiable corruption or damage; this is one reason why there is no universally agreed set of preservation file formats. These issues are being addressed by the Planets project.

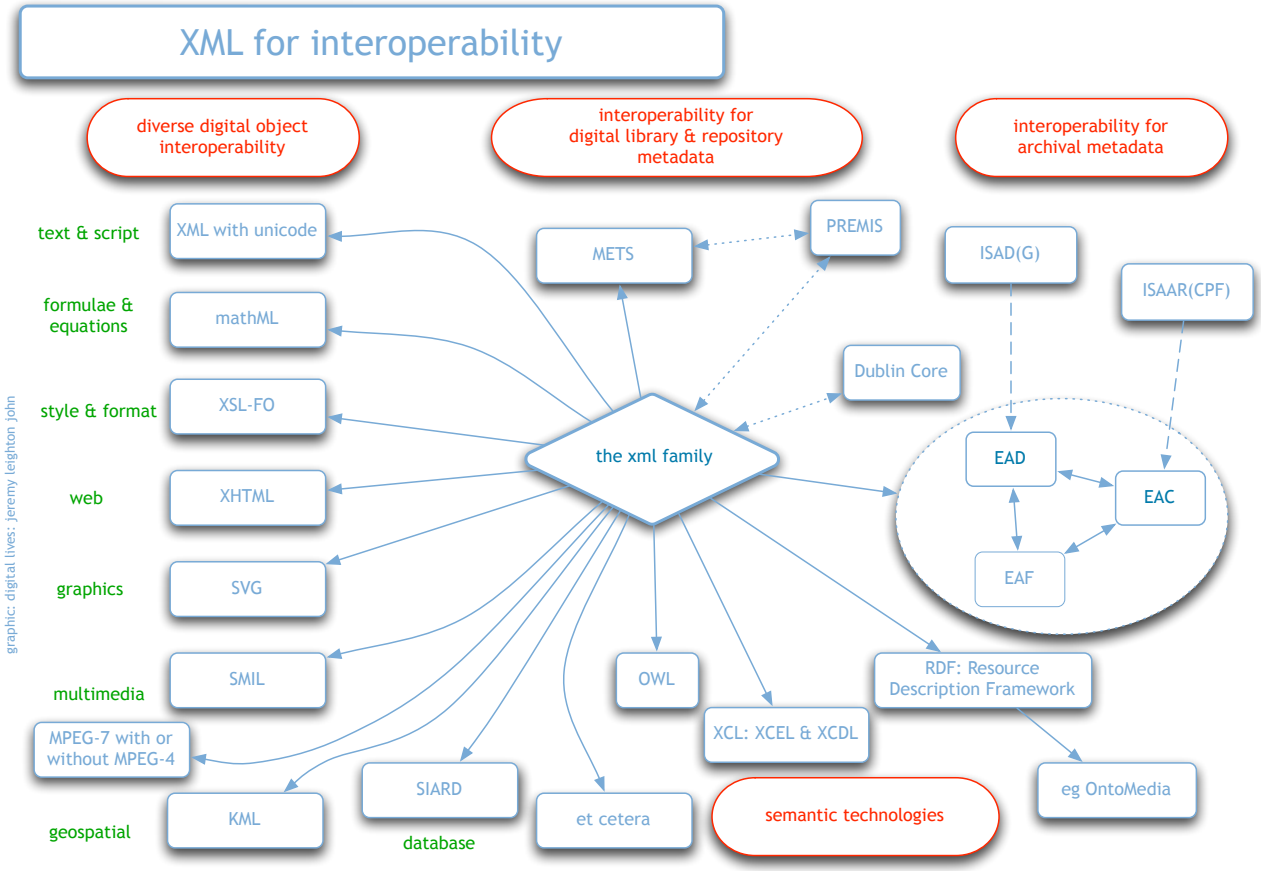
(12) There has been over the years much attention directed at individual files and their conversion to some standard of interoperability and sustainability acceptable for archival digital preservation. Indeed for a time this was the favoured approach. While it is an

[34] V. Heydegger (2009) Just one bit in a million: on the effects of data corruption in files, ECDL 2009, LNCS 5714, pp 315-326. S. Abrams (2006) Knowing what you've got. Format identification, validation, and characterization. DCC/LUCAS Joint Workshop, University of Liverpool, 30 November - 1 December 2006, http://www.dcc.ac.uk. S. Abrams (2007) File formats. DCC Digital Curation Manual, October 2007, Version 1.0, 53 pp. S. Barve (2007) File formats in digital preservation, ICSD 2007, pp 239-248.

important approach, the emerging files are effectively 'digital facsimiles'[35]. It is essential therefore to capture as far as possible: (i) exact replicates of the original files, acknowledging their historical and informational value; and (ii) as much as possible of the contextual space within which these files exist (the entire hard drive or set of hard drives, the entire folder directory structure, the map of remote locations of the personal information (eg in the cloud, as represented by online service providers).

(13) The capture of context may be made even more comprehensive through enhanced curation activities such as panoramic photography of the work and home environment.

(14) A pragmatic approach is to provide for immediate access at the same time as seeking to capture and retain exact replicates of files (in the absence of, say, a high fidelity emulator). This might entail the extraction of raw text or an imprecise rendering of an image sufficient for some but not all scholarly purposes. With new and more rigorous tools emerging from the digital preservation and other communities, this compromise hopefully will become less pronounced and less commonplace – it being possible to undertake a precise conversion in more and more instances, and to employ more exact emulators.



Sources of tools

(1) Five potential sources of tools were considered: (i) computer forensic resources, (ii) ancestral computer enthusiasts, (iii) the digital preservation community, (iv) consumer

[35] The term 'digital surrogate' is used by the Digital Lives project for the digital objects that result from the digitisation of analogue objects

computer software and hardware producers, and (v) evolutionary computing technology developers. In this document most attention is directed at (i), (ii) and (iii).

(2) In contemplating the relevance of tools, all three phases of the lifecycle – acquisition and capture, processing and examination, and access – have been borne in mind although the priority has inevitably been given to the first and second phases of the lifecycle.

(3) Attention was directed at open source software as well as proprietary tools; the latter can be acceptable provided the output is nonproprietary.

Forensics

Historic document analysis

(1) There have long been skills and tools shared by the forensic and manuscript communities due to their mutual interest in establishing authenticity, in document analysis and in determining past events and the histories of objects.

(2) The initial motivation to explore forensic computing techniques arose from the simple fact that even turning on a computer risks altering dates and times of files. Computer forensic scientists wish to be able to capture digital files with demonstrable rigour and to explore them without altering their contents or associated metadata.

(3) Moreover, computer forensic scientists are routinely expected to meet certifiable standards in order to satisfy legal authorities. Across the land, and from week to week, many of these technologies and their uses are being accounted for and defended in a court somewhere. This provides some confidence although it can never be complete, and techniques still have to be applied properly[36].

The fundamental approach

(1) Instead of switching computers on, internal hard drives are extracted from the computer and attached to a write blocker that prevents the curator's workstation from writing or modifying the original collection hard drive; similarly, floppy disks, external hard drives, flash media and other removable media are all carefully write-protected beforehand. Next a forensically sound bitstream 'imaging' of the original disk is conducted yielding a file that encapsulates the entire disk[37].

(2) Exact digital replicates of the original files can be exported from the bitstream 'image' file. Digital facsimiles, interoperable and hopefully future-proof, can be created from these

---

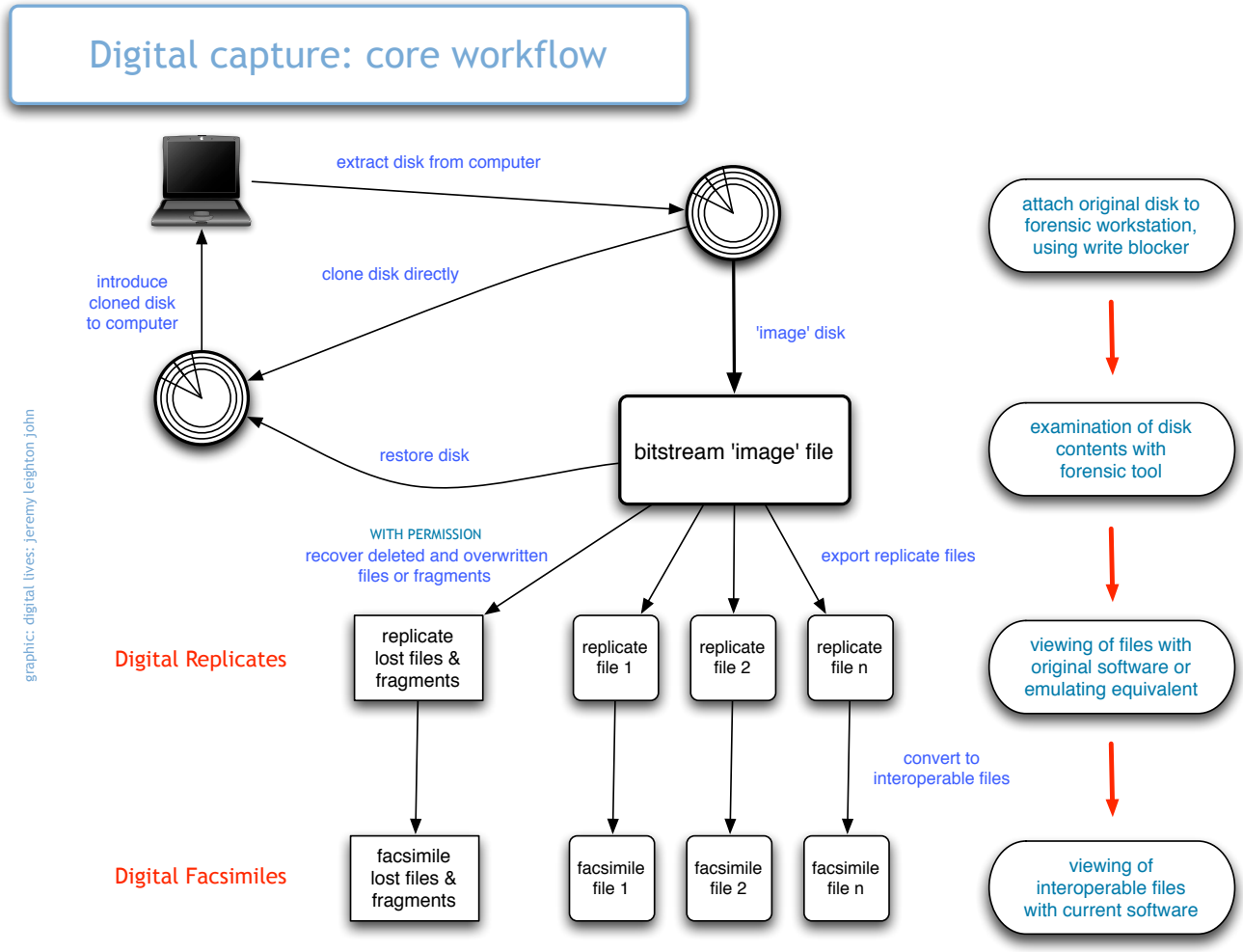
[36] A further source of confidence is that forensic tools (both hardware and software) are subject to independent testing. For example, the Computer Forensics Tool Testing (CFTT) program is undertaken by the National Institute of Justice (which is the research and development organisation of the US Department of Justice) and the Office of Law Enforcement Standards and Information Technology (which is part of the US National Institute of Standards and Technology, NIST). The program publishes its test results. For example the data acquistion software DCCIdd (a forensic version of the dd disk imaging tool) has been evaluated: M. B. Mukasey, J. L. Sedgwick and D. W. Hagy (2008) Test results for digital data acquisition tool: DCCIdd (Version 2.0), NIJ Special Report, January 2008, 122 pp

[37] The 'imaging of hard drives and other disk media' has recently been strongly supported by prominent members of the digital archive community: eg see M. Kirschenbaum, E. L. Farr, Kari M. Kraus, N. Nelson, C Stollar Peters, G. Redwine and D. Reside (2009) Digital materiality: preserving access to computers as complete environments, iPres 2009 Conference, Stanford University Libraries, preprint. See also: J. L. John (2006) Digital manuscripts: capture & context, Digital Curation Centre Conference, Email curation: practical approaches for long-term preservation and access, 24-25 April 2005, Newcastle, http://www.dcc.ac.uk/events/ec-2006/EC_Digital_Manuscripts_Jeremy_John.pdf; S. Thomas and J. Martin (2006)

digital replicates by means of conversion to suitable file formats such as members of the XML and PDF families.

(3) In the process of 'imaging' a collection disk, hash values are created for every file as well as for the entire disk. These unique hash values derived from MD5, SHA1 or SHA256 one-way algorithms are effectively 'digital fingerprints'. If one hundred years later a file is subjected to the same algorithm, and the same hash value is obtained one can be confident that the file has not changed over the century.

(4) This core functionality can be achieved with equipment purchased at modest cost: essentially a series of write blockers (eg from Tableau) combined with forensic 'imaging' software (AccessData provides a well known imager that is free). There are also open source software tools. Of special note is the Advanced Forensic Format system that enables the storage of extensive metadata integrated within the 'image' bitstream file, and also incorporates the use of digital signatures that facilitate the process of authenticating the hash values themselves[38].



Features

(1) Two of the most well established forensic software packages are Encase of Guidance Software and Forensic Tool Kit of AccessData. Both seek to be comprehensive, providing in

[38] For further information about AFF see §9.2

one package a wide range of functionality, and are correspondingly more expensive than the more isolated tools.

(2) Additional functionality includes the bookmarking and annotating of files, timeline viewers for investigating times and dates of file creation, modification and access, while taking into account different time zones, provisional identification of file types based on file signatures and extensions, maintenance of an examination audit trail, refined searching, file viewing, and reading of emails with attachments, and viewing of hexadecimal code and disk organisation.

(3) Some examples of what can be done illustrate the immediate applicability of forensic tools.

- All of the computer media from one archive (the hard drives and removable media) can be integrated by the forensic software as a single case, and searches can be conducted across the entire collection.

- Libraries of known hash values are maintained by institutions such as the National Institute of Standards and Technology (NIST) in the United States of America[39]. The hash values of the files in the personal archive can be compared with a known hash library for application and operating system software files, allowing these to be filtered out from searches if desired.

- The identification of the software may also be facilitated and in turn this could be used to establish likely candidates when seeking to identify the file formats existing in a collection with digital preservation tools.

- One of the most time consuming aspects in the curation of personal digital archives is the tagging of files where digital rights and ethical concerns may apply. Existing and customisable scripts are available that allow searches for specific content such as, for instance, files containing credit card numbers, telephone numbers, post codes, and email addresses. Similarly searches for the most commonly used URLs may point to favoured online service providers such as social networking sites and webmail accounts[40].

- The bitstream ‘image’ can be mounted in read-only mode and it can be scanned for viruses prior to the exporting of digital replicates.

- Metadata found in association with files - eg file extension, file type, file signature, dates and times, permissions, hash values, logical size, physical location, file extents (fragmentation) and others - can be exported and subsequently imported into a spreadsheet or cataloguing system.

---

[39] The National Software Reference Library (NSRL) based at NIST, profiles the full range of software and maintains the NSRL Reference Data Set (RDS) that allows software files to be readily identified, http://www.nist.gov/srd/nistsd28.htm

[40] S. Garfinkel and D. Cox (2009) Finding and archiving the internet footprint, paper presented at the Digital Lives Research Conference, British Library, 9-11 February 2009, http://simson.net/clips/academic/2009.BL.InternetFootprint.pdf

- The entire corpus of captured information can be automatically indexed (including text within files). At the same time it is possible to filter out certain classes of files (eg those subject to data protection) in the indexing.

- Passwords are sometimes forgotten or records are accidentally lost, and with the permission of family and originators, decryption and password recovery tools can be used with varying levels of success.

- On occasion, some deleted information (eg whole or fragments of earlier draft versions of a novel) might be recovered with the permission of the originator.

(4) It is essential that potential depositors are fully aware of the latent power of digital technologies in searching, data fusion and mashing, and in forensic analysis. This aspect of the relationship between depositors and curators, and approaches towards establishing informed consent and understanding available options, warrants further study.

(5) There are clear archival benefits to capturing computer media in their entirety not least for aiding authenticity and providing context - for example, a file that is manifestly derived from an entire disk is much easier to authenticate than an isolated file.

(6) Nonetheless an originator may wish to simply donate some specific folders or files rather than an entire carrier such as a disk. This requirement can be met with a forensically sound 'logical' acquisition of files akin to a 'physical' acquisition of a disk. This option might be preferred when the repository is dealing with a 'living archive' where many files are still being used, created and amended.

(7) An initial examination of a digital archive can be conducted at a creator's home using a forensic laptop and a preview facility that allows curators and creators decide whether an archive should be transferred to a repository. In some cases such as with obsolete media, it may well be necessary to take the material to the repository for examination. The digital objects residing on obsolete media such as 5.25" floppy disks may not have been seen for many years, and so again it will be necessary to involve the originator or family, and perhaps cater for the possibility of a private viewing.

(8) The significant involvement of depositors and originators, of course, has many potential benefits in, for example: (i) aiding the identification of personal digital objects that are likely to prompt data protection and confidentiality issues, and (ii) supplying contextual and corroborating information that increases the scholarly and historical value of the entire digital archive.

(9) The use of forensic technologies needs to be put in perspective, bearing in mind a number of observations: (i) manuscript curators, conservation experts and document analysts have long employed forensic techniques; (ii) most of the functionality is widely available in tools produced outside the forensic community; (iii) much of the analytical power of digital forensics ultimately stems from the nature of digital media and technology, and everyone - creator, curator and researcher - needs to have some understanding of the emerging possibilities (without necessarily knowing the details); and (iv) the principal benefits of using forensic technologies are their certified or peer tested status, and their integration and relative convenience of use.

(10) Some of the forensic systems offer a variety of ways of: (i) allowing colleagues to have access to the digital objects, without risk of changing the original data, the hash values and other metadata, or the records of the history of the objects since joining the forensic system; and (ii) controlling and recording access to some or all of the information.

(11) The available systems offer a source of ideas if not initial models of controlled access for users. Along with providing practical solutions that can be adopted in initial stages of the lifecycle in capture and examination, the way the information is presented on the computer screen is demonstrated with several viewing panes for: (i) exploring the folder directory structure, (ii) modifying the listing of objects and viewing of metadata in different ways, (iii) experiencing the actual contents of objects, text, image, audio and video, along with (iv) panels and toolbars as finding aids for access, with search, filter, keyword and index functionalities[41].

Ancestral computer enthusiasts

(1) In the archival context, there will be circumstances when it is essential for the personal digital object and the computer environment to be made available to a researcher, as it was perceived by the creator - with styles, layout and behaviour precisely represented.

(2) There is a need to understand the way users interacted with their computers, how these operated, the nature of available applications, and the character of the files produced — just as curators of conventional manuscripts are required to know about the ways in which writing media (wax, parchment, vellum, paper) and associated technologies (pen, ink, pencil, stylus) were designed and used.

(3) An important source of useful equipment in this regard is the community of ancestral computer enthusiasts, many of whom are highly expert in this field[42].

(4) Technologies are being produced specifically to serve this community. Two types of product can be briefly outlined, both associated with a small company Individual Computers: (i) a modern universal floppy disk controller (Catweasel) that can be inserted into modern computers to allow them to interpret the disk formats of early format floppy disks; and (ii) a modern computer with a hardware system (C-One) that can be configured (and reconfigured) to behave like one of a number of ancestral computers. There is also software that emulates hardware, operating systems and applications; an example is Amiga Forever; others include MESS and SIMH[43].

---

[41] The Forensics Wiki provides an invaluable source of information and ideas that bear on the capture, authentication and proper management of digital objects, http://www.forensicswiki.org/wiki/Main_Page. It covers a wide range of topics and tools

[42] Another and in some sense emerging source of expertise is the current generation of researchers and developers working in the field of personal information management and the manifest research output in this area

[43] http://simh.trailing-edge.com and http://mess.redump.net

"Without the original bits, authentic recreation of the digital object is impossible. Capturing the bits from the original data carrier is therefore crucial and requires sophisticated and reliable tools" *quoted from KEEP website*[44]

Virtual machines and emulators

(1) Digital preservation specialists have produced a plethora of policies, tools and guides. One that is highly relevant to personal archives is the Dioscuri emulator, open source software that emulates a computer environment based on the x86 lineage of computer processors (responsible for many generations of personal computer). Significantly it has been designed with longterm preservation in mind, and is both durable and flexible[45].

(2) The Dioscuri emulator is organised as a series of software modules, each of which mimics a component of hardware. The modules can be rearranged according to the structure of the hardware that is to be emulated. Architecturally, Dioscuri lies on top of a virtual machine that potentially would be able to operate on numerous underlying hardware systems, which in turn means that Dioscuri is able to run on diverse computers.

(3) The Dioscuri work is now being incorporated within another exciting project KEEP, Keeping Emulation Environments Portable. The aims are: (i) to understand an extensive range of media carriers such as 8" floppy disks and tapes, and (ii) to develop new tools if necessary for copying information from these media; and (iii) to create an Emulation Access Platform that will allow (a) static and dynamic digital objects to be rendered accurately, and (b) the transfer of data from the emulated environment to the host environment partly for researching the emulation.

(4) Virtual machine technologies are widespread feature of modern computing (from Vmware to open source Xen), but two are special in being directed at digital preservation: Olonys and UVC.

(5) The Olonys virtual machine by Vincent Joguin has been embraced by the KEEP project in order to take advantage of the support offered by the Olonys system for external communication with peripheral devices.

(6) The Universal Virtual Computer promoted the concept of a universality that allows the virtual machine to be readily ported to almost any hardware computer system. It is founded on a highly condensed minimal set of operations that future computers can implement. The Planets project has supported the UVC and is ensuring that it has models for engaging with peripheral devices.

Box: Universal virtual computer

The UVC stemmed from a concept of R. A. Lorie at IBM Almaden USA that has been realised to a significant extent by IBM Netherlands in collaboration with the National Library of Netherlands, and research and development continues in association with the Planets programme. There are two key ways in which the UVC is being applied.

---

[44] Keeping Emulation Environments Portable (KEEP), http://www.keep-project.eu/, a project that is being led by the Bibliothèque nationale de France (BnF), http://www.bnf.fr

[45] http://dioscuri.sourceforge.net

In the first approach, a digital object is converted to another object (a Logical Data View) by a UVC program which ensures that the converted object will be readable in the future by means of another UVC program acting in conjunction with a Logical Data Schema. In this first approach the original digital object such as a Macwrite document would be converted to an interoperable form: in effect this is a migration but one that is compatible with the future-proof UVC (an advantage of using the UVC approach is that this migration should only need to be done once)[46].

In the second approach, the UVC is envisaged as a means of retaining the function - to some extent - of computer programs including application and operating system software[47]. This approach essentially entails the writing of a UVC program that emulates the operating system (partially or wholly) on which the archived program depends. If the original program does not require input or output interactions then the archived operating system will suffice but often this will not be so, and it is necessary to write additional UVC programs that mimic the diverse input and output devices and processes. In this second approach the original digital object such as a Wordstar document can be preserved and viewed in its original form by means of emulation.

The initial implementation of the modular emulator Dioscuri (for example) runs on the open source and widely used Java Virtual Machine but it is envisaged that in due course a fully fledged UVC will be adopted and be made available.

Planets: overview

(1) Preservation and Long-term Access through Networked Services, PLANETS, is a four year research and technology development project that commenced in June 2006 and is funded by the European Union[48]. It is directed at libraries, archives and any organisation that is responsible for safeguarding digital information; and in particular at providing solutions (i) for integrating and (ii) testing tools, (iii) for handling collections and not just isolated digital objects, and (iv) for dealing with complex and dynamic objects.

(2) From the perspective of digital archives, it is reassuring that the Planets project is founded on the concept that there are two kinds of fundamental actions that can be performed to ensure that digital objects continue to be accessible in the future: (i) transform the object itself (preservation migration that converts an object into a new format so it can still be read in the future), and (ii) produce a tool that mimics the environment in which the digital object operated so that its users can still interact with the digital content of the object.

(3) Planets is deemed to be a practical implementation of the OAIS Reference Model which was originally devised by the Consultative Committee for Space Data Systems[49] and which is now an ISO standard for digital preservation tools and repositories. There is a difference in

---

[46] This aspect of the UVC is outlined by N. J. C. Kol, R. J. van Diessen and K. van der Meer (2006) An improved Universal Virtual Computer approach for long-term preservation of digital objects. Information Services & Use 26: 283-291

[47] R. A. Lorie and R. J. van Diessen (2005) Long-term preservation of complex processes. Archiving 2: 14-19

[48] http://www.planets-project.eu/

[49] Consultative Committee for Space Data Systems (2002) Reference model for an open archival information system (OAIS), CCSDS 650.0-B-1, http://public.ccsds.org/publications/archive/650x0b1.pdf

emphasis (notably where the OAIS model emphasised migration more than emulation at the time of its inception), although precise alignment with OAIS continues to evolve and it is acknowledged that future refinement and extension of the OAIS model can be expected in time[50].

(4) It aims to provide in the form of software that can be downloaded and installed with a few clicks a set of tools and services that allow: "the administration, configuration, and deployment of preservation services and workflows"[51]. For example, it is anticipated that an end user will be able to choose from several options: (i) look at the digital object with the original viewer, running under emulation; (ii) launch a viewing tool that renders the digital object on the screen without any further functionality; (iii) examine a migrated (converted) version of the digital object in current software.

(4) The Planets project consists of five research and development activities: (i) preservation planning, (ii) characterisation of digital objects, (iii) preservation actions, (iv) a test bed, and (v) an interoperability framework.

(5) The essential approach is: (i) to incorporate and enhance existing and well established tools and services, (ii) to build new complementary tools and services, (iii) to integrate them all in a consistent and accessible way, and (iv) to ensure the potential for ongoing extension, incorporation and integration over the network.

Planets: planning

(1) Preservation planning is conducted with the Plato tool[52]. It enables archivist's to make an informed, accountable and comprehensively documented selection of the most appropriate preservation plans in the context of the archivist's specific purposes and requirements.

(2) The user is taken through four steps: (i) the definition of context and requirements, (ii) the selection and evaluation of potential actions based on sample content, (iii) the analysis of outcomes with recommendations, and (iv) the identification and elucidation of a preservation plan based on this empirical evidence. At present each plan typically would be directed at a specific or broad category of files, eg jpg files or image files.

(3) An enhancement is planned for the Plato tool version 3 (to be released Spring 2010) in the form of a recommender system that will reduce time and effort required in planning and will simplify aspects of the workflow[53].

---

[50] A. Farquhar and H. Hockx-Yu (2008) Planets: integrated services for digital preservation. Serials 21(2): 140-145. See also a different version of the article in International Journal of Digital Curation 2(2): 88-99, (2007)

[51] Farquhar and Hockx-Yu (2008), ibid

[52] Planets Preservation Planning Tool: Plato 2.0, User Manual, vo.8, 7 November 2008. See http://www.ifs.tuwien.ac.at/dp/plato. The planning process emerged from a variety of case studies including: (i) C. Becker, S. Strodl, R. Neumayer, A. Rauber, E. Nicchiarelli Bettelli and M. Kaiser (2007) Long-term preservation of electronic theses and dissertations: a case study in preservation planning, Proceedings of the 9th Russian Conference on Digital Libraries, RCDL 2007, Pereslavl, Russia; (ii) C. Becker, G. Kolar, J. Küng and A. Rauber (2007) Preserving interactive multimedia art: a case study in preservation planning, ICADL 2007, LNCS 4822, pp 257-266. Among other things the former supported the preservation value of PDF/A as a format of choice for some text documents, while the latter considered the special requirements of interactive multimedia

[53] Planets (2009) News round up: work starts on recommender support in Plato. Planetarium: the news bulletiin of the Planets programme, issue 8, December 2009, pp 9-10

(4) Digital preservation audit and risk management is also supported by the DRAMBORA[54] toolkit, while cost analysis is being investigated and developed by a sister project of Digital Lives at the British Library, LIFE3 (following the successes of LIFE2 and LIFE)[55].

Planets: characterisation of digital objects

(1) It is essential to be able to identify the features – often subtle - of digital objects that characterise them. For example if file conversion is being contemplated it is necessary to understand the implications and the possible loss of details of style and behaviours including interactive potential. The aim is to make it possible for such characterisation to take place automatically. Planets builds on the well known tools: (i) DROID (designed for identifying file types, file formats)[56], and (ii) JHOVE (designed to extract file properties and validate the file format of digital objects)[57].

(2) The Characterisation Registry is based on PRONOM[58] and contains technical information about the significant properties of digital object types such as file formats, compression methods and character encoding schemes. (PRONOM and GDFR[59] have recently joined forces to create the Unified Digital Formats Registry, UDFR.)

(3) Three other things are required: (i) a language with which to express the characteristics, (ii) an understanding of how to extract the characteristics from the files, and (iii) a means of comparing them. Two eXtensible Characerisation Languages (XCL) have been developed by Planets to meet these requirements[60].

(4) The Extensible Characteristics Description Language (XCDL) allows the description of digital objects: characteristics such as colour depth of an image, number of images within a textual document, metadata embedded within a file or the font type of text.

(5) The Extensible Characteristics Extraction Language (XCEL) makes it possible to define how characteristics can be extracted from a digital object.

(6) Using a software application called an Interpreter, Planets can automatically execute the XCEL definitions to extract XCDL characteristics from a file. The concept has been tested with images and text documents.

---

[54] DRAMBORA, Digital Repository Audit Method Based on Risk Assessment, http:www.repositoryaudit.eu, developed jointly by the Digital Curation Centre and DigitalPreservationEurope. It is aimed at providing a repository with a methodology for self-assessment. See DRAMBORA Interactive, User Guide, M. Donnelly, P. Innocenti, A. McHugh and R. Ruusalepp (2009)

[55] http://www.life.ac.uk

[56] DROID, Digital Record Object Identification, http://droid.sourceforge.net

[57] JHOVE, JSTOR / Harvard Object Validation Environment. A draft version of the Architectural Overview of JHOVE2 has been prepared; see also http://confluence.ucop.edu/display/JHOVE2Info/Home

[58] PRONOM, a technical registry based at the National Archives of the United Kingdom, http://www.nationalarchives.gov.uk/pronom. See in particular A. Brown (2006) The PRONOM Unique Identifier Scheme, DPTP-02, issue 2, pp 1-9, http://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf

[59] GDFR, Global Digital Format Registry at Harvard University

[60] C. Becker, A. Rauber, V. Heydegger, J. Schnasse and M. Thaller (2008) Systematic characterisation of objects in digital preservation: the eXtensible Characterisation Languages. Journal of Universal Computer Science 14(18): 2936-2952

(7) A third tool that has been developed is the Comparator, which makes it possible for the XCDL characteristics of two files to be compared automatically. It allows the automatic checking that conversion (from say a word document to a corresponding PDF) has taken place successfully.

(8) The ability to convert the output of JHOVE to XCDL is being incorporated so that Planets can enable functionality for any file formats that JHOVE has addressed.

Planets: actions and testbed

(1) The series of activities directed at preservation action are: (i) the design of a tool for describing preservation tools, (ii) an initial tool registry, and (iii) a range of current action tools such as the Dioscuri emulator.

(2) Planets has identified the file formats most commonly encountered by digital archives and has created an inventory. A survey of the tools that exist to characterise and preserve each file format has been conducted in order to identify where tools do not exist or are insufficient.

(3) The gap analysis in tool provision identified 137 different file formats, submitted by 76 respondents. The Planets Core Registry (PCR) is an ongoing registry for information about file formats and preservation action tools (specifically, migration tools and emulation tools) along with other complementary information such as pathways for action. In October 2009, the list contained 57 migration tools[61], which together are able to migrate the most common file formats (with the exception of XML which is explained by its acknowledgement as a preservation output for many migrations tools)[62].

(3) The project, along with its partners, has already embraced a series of tools for a variety of tasks relating to characterisation, conversion, reading and writing of files, including: Abiword, Ghostscript, InkScape, JJ2000, JTidy, MsgText, Pdf2Ps, Pdf2PdfAMayComputer, SanselanMigrate and Xena[63].

(4) Three example activities for text, image and database include:

- producing tools for extracting XCDL descriptions from binary Office formats, bearing in mind that Microsoft is developing tools for transforming earlier Office formats into OOXML;

- enabling the characteristics extracted by tools such as ImageMagick to XCDL, potentially valuable since ImageMagick can read, write and characterise bitmap images in more than one hundred formats;

---

[61] S. van Bussel and F. Houtman (2009) Gap analysis: a survey of PA tool provision. Planets Programme, Preservation Action, Deliverable PA/2 - D3, pp 1-26, 12 October 2009

[62] A careful survey of tools has also been conducted by the Cairo project, a collaboration between Oxford University Library Services, the John Rylands University Library and the Wellcome Library: S. Thomas, F. Baker, R. Gittens and D. Thompson (2007) Cairo tools survey: a survey of tools applicable to the preparation of digital archives for ingest into a preservation repository, 21 May 2007, pp 1-50

[63] A. A. Blekinge (2009) Preservation action toolset update. Planetarium: the news bulletin of the Planets programme, issue 8, p 7, December 2009

- adopting SIARD for the conversion of relational databases such as MS Access, noting that the SIARD suite is able to convert SQL and Oracle to SIARD format.

(5) The aim of the Planets Testbed is to provide a controlled software environment for scientific experiments and testing of practices in digital preservation[64]. It will enable the empirical evaluation of strategies and tools for particular content and digital objects. It followed conclusions by several studies that "digital preservation has been characterised by practices and processes that could best be described as more art and craft than science, and that "frameworks for *experimentation* must be central to the design and practice of digital preservation research".

(6) By providing the testbed as a web service, Planets enables many users to perform a wide range of experiments ranging from the testing of migration and characteristation through to the loading of new workflows. A common Testbed experiment for a workflow might involve: (i) requesting a characterisation service to determine the significant properties of some input data and to identify appropriate migration tools; (ii) subsequently employing a migration service to migrate the data; and (iii) using a second characterisation service to assess the results of the migration automatically[65]. Other experiments might test the process for some very large files or for very many files.

Planets: interoperability over the network

(1) The Planets Interoperability Framework makes it possible for all the tools, registries and services to be incorporated within a manageable preservation system that allows exploratory requests and discovery.

(2) A central perspective is the integration of planning and actions, with the Planets system making flexible and adaptable use of existing services such as CRiB (Conversion and Recommendation of Digital Object Formats), a publicly available web service (Service Oriented Architecture) that offers various kinds of migration possibilities and assessments[66].

(3) Remote access to emulation tools such as Dioscuri and UVC is being enabled by a new service known as Grate developed by the University of Freiburg as part of Planets.

(4) A vital requirement is the monitoring of the quality of the web services: service quality and tool performance measurement that is required for both the emulation services and the migration services[67].

(5) The Framework is designed to be extensible with third party tools and services being plugged in as they arise. Guidelines have been prepared for 'wrapping' preservation tools for integration with the Planets system. This 'wrapping' is the provision of a web service interface that allows the tool to be incorporated into the preservation workflows. By this

---

[64] B. Aitken, P. Helwig, A. Jackson, A. Lindley, E. Nicchiarelli and S. Ross (2008) The Planets testbed: science for digital preservation. Code{4}Lib Journal 3, 23 June 2008

[65] B. Aitken, P. Helwig, A. Jackson, A. Lindley, E. Nicchiarelli and S. Ross (2008) The Planets testbed: science for digital preservation. Code{4}Lib Journal 3, 23 June 2008

[66] C. Becker, M. Ferreira, M. Kraxner, A. Rauber, A. A. Baptista and J. C. Ramalho (2008) Distributed preservation services: integrated planning and actions, ECDL 2008, LNCS 5173, 25-36; for CRiB see http://crib.dsi.uminho.pt

[67] C. Becker, H. Kulovits, M. Kraxner, R. Gottardi, A. Rauber and R. Welte (2009) Adding quality-awareness to evaluate migration web-services and remote emulation for digital preservation, ECDL 2009, LNCS 5714, pp 39-50

means remote access to new file format identification, file conversion and emulation and media imaging tools can be made possible, for instance.

(6) The Interoperability Framework has been ported to the website GForge for initial dissemination. The date for final project release is May 2010[68], and the software will be available as an open source project at Sourceforge[69].

Guides to obsolescence and the presevation of software

(1) The Automated Obsolescence Notification System 2 (AONS II) based in Australia provides up to date information about file formats and technologies that are at risk of becoming obsolete.

(2) A useful resource for archivists and individuals and also being developed in Australia is Mediapedia: it provides images and information about obsolete and contemporary media, 5.25" floppy disks, tape data cartridges and the like[70].

(3) For a long time there has been a presumption that the preservation of software is untenable and even largely undesirable from the preservation perspective. The need for a systematic approach directed at preserving software is receiving attention by members of the preservation community as well as by historians of computer science and use.

Box: Software preservation

Software is complex, having many components and an inherent fragility due to a sensitive dependency on each component and the entire environment; furthermore, the software systems are extremely diverse. Software preservation has received low priority among professionals due to this complexity and diversity and to an uncertain future requirement for software. In fact there are good reasons for researching and preserving software. Foremost among these reasons is that sustaining the usability of software supports the preservation of digital objects generally[71].

---

[68] Planets Programme (2009) Interoperability framework. The glue that holds Planets together. Planetarium: the news bulletin of the Planets programme, issue 8, p 2, December 2009

[69] http://sourceforge.net

[70] N. del Pozo, D. Elford and D. Pearson (2009) Mediapedia: managing the identification of media carriers. In: DigCCurr 2009. Digital Curation. Practice, Promise & Prospects. University of North Carolina at Chapel Hill, North Carolina, USA: School of Information and Library Science, University of North Carolina, Chapel Hill, pp 76-78

[71] Members of Rutherford Appleton Laboratory of Science and Technology Facilities Council (STFC), one of the institutions representing the Consortium of the Digital Curation Centre, recently completed a report for JISC recommending greater awareness of and research into the issue of software preservation for its own sake and for its role in digital preservation and curation processes: The significant properties of software: a study, March 2008, B. Matthews, B. Mcllwrath, D. Giaretta and E. Conway. Further support for the archiving of software and documentation: T. Reichherzer and G. Brown (2006) Quantifying software requirements for supporting archived office documents using emulation, ICDL 2006; J. L. John (2008) Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives and tools, iPRES 2008 Conference, the Fifth International Conference on Preservation of Digital Objects, the British Library, London; [P. Wheatley] (2004 ca) Survey and assessment of sources of information on file formats and software documentation, Final report, Representation and Rendering Project, University of Leeds (http://www.jisc.ac.uk/whatwedo/programmes/fileformat.aspx); D. von Suchodoletz and J. van der Hoeven (2009) Emulation: from digital artefact to remotely rendered environments. The International Journal of Digital Curation 4(3): 146-155. See also: http://www.softwarepreservation.org/ and http://www.computerconservationsociety.org/index.htm

Several recommendations have emerged: (i) raise awareness; (ii) develop a careful consideration of the role of software in preservation generally and in relation to OAIS processes; (iv) involve experienced software engineers in embedding software sustainability in code development, maintenance testing and reusability; (v) consider the special issue of preserving the way users interact with a software package; (vi) explore the role of software preservation in emulation and migration; (vii) examine further the identification of significant properties in the context of software digital objects; and (viii) develop methodologies for testing software in the context of preservation. It has been recognised that there is also a need for an institutionally coordinated approach to the preservation of software for the longterm.

Specific challenges that have been mentioned include: (i) the fact that the complexity of software continues to increase, which in turn means that (ii) there are more and more ancillary components to address (hardware drivers, plug-ins, decoders, fonts); (iii) legal issues, the online nature of software activation and the sourcing of regular updates to keep the software operational; (iv) understanding how the software works and ensuring that essential information about it is retained (eg manuals and guides); and (v) capturing in some sense the diversity of software releases (eg various languages). To these recommendations could be added a need to catalogue and classify all software with a view to establishing historical and convergent relationships between them, a process which would support coordinating activities among other things.

With the growing complexity it seems likely that an approach along the lines of systems biology and evolutionary science will become necessary, not least with the emergence of autonomic computing which is itself borrowing concepts from biological phenomena. Most urgent is a need to ensure that institutions that are holding software collections are applying digital preservation standards: the US National Software Reference Library and the open source Sourceforge spring to mind. An important potential source of software and licenses for archival repositories that should not be overlooked are personal digital archives within which exist diverse and obsolete software.

Personal archiving with digital preservation

(1) Of particular relevance to the Digital Lives project has been the recent development of software aimed at archiving by individuals and by small office and home office enterprises: Hoppla (Home and Office Painless Persistent, Long-term Archiving). It is written in the open source language of JAVA and has been partly funded by Planets, and is derived from the Vienna University of Technology[72].

(2) Using preservation standards and tools, Hoppla seeks to provide in a user-friendly way digital preservation and backup jointly for the home user and small archive. It models its level of service on currently existing firewall and antivirus solutions. There is an emphasis on ease of use with updates and upgrades being provided through web services. Although it provides service functionality, the services are installed locally and do not require movement of a person's files over the internet to web services, thereby protecting privacy.

---

[72] S. Strodl, F. Motlik, K. Stadler and A. Rauber (2008) Personal & SOHO archiving, ACM/IEEE Computer Science Joint Conference on Digital Libraries (JCDL 2008), Pittsburgh, Pennsylvania, 16-20 June 2008 (SOHO is Small Office / Home Office); S. Strodl, F. Motlik and A. Rauber (2009) Hoppla - Digital preservation support for small institutions, Conference on Digital Curation: Practice, Promise & Prospects (DigCCurr 2009), University of North Carolina at Chapel Hill, 1-3 April 2009; http://www.ifs.tuwien.ac.at/dp/hoppla; also referred to as Home and Personal Persistent Long-term Archiving project

(3) Significantly, Hoppla embraces both migration and emulation by retaining the original files while at the same time producing versions converted into interoperable formats.

(4) The software allows for the possibility of acquiring information from diverse sources including the web - eg from an individual's website - or from email accounts. Versioning and time stamping is supported and can in principle embrace sophisticated synchronisation software such as Unison[73]. The proper behaviour and success of migration is checked and logged. All metadata created are stored in XML.

(5) Storage is based on the well established principle of holding multiple copies. It does not allow for encryption which is sensible at this stage, but may need to be reconsidered if future consumers want and expect some encryption capability.

(6) The initial and informal requirements specification appears to place less emphasis on authentication in the case of home archiving. This is understandable but from the perspective of repositories and others, the longterm authentication is paramount, and this need could be addressed as the software advances in its development.

(7) This is an exciting proposition and perhaps with some amendments could be employed by very small and local archives as well as individuals.

Consumer Computer Software and Hardware[74]

Unifying backup and preservation

(1) At present there is a wide lack of awareness of the distinction between backup and preservation, and the need to do more than backup to ensure longterm preservation. There are two distinct requirements: disaster recovery; and looking after your files for the longer term.

(2) Backup as carried out by individuals at home generally does not meet longterm preservation standards. Most backup solutions are directed at providing emergency protection in the event of a failure of storage media and, perhaps, in the event of inadvertent deletion of files by a user. In the event of a disaster such as a catastrophic failure of a hard drive, the computer system and its files (or just some files) can be restored. This involves the creation of a special file that encapsulates the contents of the disk.

(3) A specific and serious concern is the tendency for the backup files produced by commercial software to be proprietary. This means that future software might be unable to interpret and make use of the file. However, some companies implicitly acknowledge the issue and offer alternatives: eg Acronis Image Home 2009 and 2010 and East-Tec Backup 2009

---

[73] Unison is a well regarded open source file synchronisation program that runs on Windows, Mac OS X, Linux and Solaris, and can synchronise files across diverse platforms, even over the internet: ie ensures the most recent version of a file is retained at the two locations no matter at which location the modification takes place. It can be used to synchronise entire directory trees, and is relatively easy to use but is somewhat unforgiving: eg in the event of a synchronisation of a deletion. http://www.cis.upenn.edu/~bcpierce/unison/ and http://www-uxsup.csx.cam.ac.uk/pub/doc/suse/suse9.0/adminguide-9.0/node21.html. Beyond synchronisation, personal systems of reliable version control, open source, user-friendly and comprehensive would welcomed by many individuals and archivists. Apple's Time Machine is a step in this direction but is not free, of course

[74] Disclaimer: in mentioning consumer products, the Digital Lives research project is not recommending them but is seeking informally to portray the consumer market

make it possible to use the ZIP format, and Drive Image XML adopts an XML format. Nonetheless, for immediate purposes it remains advisable for individuals to additionally either directly copy files to suitable storage media, or use the backup software to restore specific files and folders, or even to clone the entire disk. Beyond the present, it is clear that there is a need for more suitable, user and preservation friendly software.

(4) Backing up and longterm preservation ought to be compatible and indeed unified. The act of storing should be future proof. Software could produce converted files (digital facsimiles) according to the latest standards as well as retain digital replicates. The software should recognise the media format and report or recommend in a timely way the migration and transfer to fresh media.

(5) The open source software Hoppla appears to be heading in the direction of a fully integrated and comprehensive solution for members of the public. At present it is not fully mature, and relatively few people are aware of it.

The persistent importance of consumer software

(1) For many people the perceived solution remains the use of commercial backup, cloning and copying software in conjunction with external hard drives or multiple hard drive systems (with or without RAID[75], and on a family network or not).

(2) Even popular computer magazines generally do not emphasise the difference between backing up and longer term preservation in their reviews. One option would be for representatives of archival and curatorial professions to highlight the distinction to reviewers of backup products. A possible route for alerting the public might be Which Online? It has in recent years been reviewing computer technologies and services, and could be encouraged to emphasise digital preservation in future reviews (and be made aware of emerging solutions such as Hoppla when it is more mature)[76].

(3) Software companies need to be directly encouraged and engaged. Although continuing to be oriented towards backup more than longterm preservation, products such as Apple's Time Machine (and Time Capsule), Microsoft's Home Server for Windows, and the Drobo system of Data Robotics all point to an understanding that there is a substantial market for personal storage, and that consumers are receptive to increasingly sophisticated functionality, most notably an ease of choosing and restoring earlier states demonstrated by the Time Machine.

(4) Cataloguing and Digital Asset Management tools are available in the form of consumer software such as Extensis Portfolio, Canto Cumulus and File Maker Pro. From the perspective of the professional archivist, these are severely handicapped by their proprietary nature but are relatively inexpensive and if used judiciously these software packages can serve a number of archival purposes. It is sometimes overlooked that it is not necessary to base a repository's catalogue entirely on such software for it to be useful: for the initial cataloguing, for example, with the data being exported as a simple file that can be imported into a primary

---

[75] RAID: Redundant Array of Inexpensive Disks; sometimes Redundant Array of Independent Disks; either way improved storage reliability is obtained through redundancy, ie replicates of information stored more or less independently

[76] For instance: Bath (2008) Protect your PC. The practical guide to PC and online security. Which? Complete Guide; in particular, see Keeping your data safe, pp 93-114. Like most guides of this kind, the aim is to protect against unexpected failure; nonetheless, such publications do represent a clear opportunity to get the message of longer term digital preservation across to a wider public

catalogue. Such functionality might be convenient for small archives or for offsite cataloguing: eg when recording the contents of a personal library or collection of artefacts. It is, of course, essential to ensure and routinely test the exporting of the data. It seems likely that even easier-to-use software will emerge as demand by individuals to manage their personal digital objects intensifies. Indeed it is already happening with photo managment tools being offered in conjunction with operating systems and as part of online services.

(5) In this context the rapidly advancing open source Archivists' Toolkit (AT)[77] is an important development for professional archivists and others. Although quite young it has existed longer than the two other key projects: Archon and ICA-AtoM. Recently and significantly, the planned integration of AT and Archon has been announced; but ICA-AtoM is important and exciting because of its collaborative capability, its adoption of international standards and clear aim to provide for multi-institutional description and interoperability[78].

Evolutionary technology, artificial intelligence, semantics and making connexions

Evolutionary perspectives and tools

(1) There are many examples of engineers adapting or copying technologies from nature. Various computer techniques in artificial intelligence, automation and creative design have borrowed concepts from nature. It can be expected that these technologies will increasingly influence curatorial and preservation practices.

(2) Another intimately related area of work is bioinformatics and genomics, fields of endeavour that face many curation challenges.

(3) Perhaps the advancements of most obvious relevance to the curation of personal archives lie in automated or supervised computer extraction and interpretation of pertinent information within digital objects, databases and also major compilations of scientific literature: either in the primary content of the object or in the embedded metadata.

(4) Computational text analysis and natural language processing of textual content could in time be helpful not only in characterising the primary content but also in helping to identify privacy and data protection issues, intellectual property and provenance.

(5) Evolutionary technologies have been used to evolve not only software but hardware: using a process of variation followed by selection and replication. These approaches have been used to improve designs of physical and digital artefacts: in qualities of reliability, optimality, robustness, versatility and even aesthetic appeal.

(6) Processes have also been improved using evolutionary computation: enhancing search functionality, recognition capability and usability. Thus these technologies and the evolutionary approach have relevance not only for cataloguing but for all areas of the life cycle from helping to ensure that the personal archive is retained and organised to the benefit of the creator in the first place through to giving scholars usability options that favour their own personalised way of conducting research.

---

[77] Archivists' Toolkit, http://www.archiviststoolkit.org and Archon, http://archon.org

[78] http://archiviststoolkit.org/node/161; see also .../node/136; and http://thesecretmirror.com/archives/the-state-of-open-source-archival-management-software, 21 December 2006

(7) Much research into knowledge discovery using text mining techniques has had to be done with structured databases. One way that the challenge of resource discovery with less structured texts has been addressed is through the use of a combination of natural processing with genetic algorithms; and it has proved possible to extract novel and apt information, even without reference to additional electronic resources or domain knowledge beyond the text itself[79].

Integrating the hybrid and engaging the creators

(1) Another interesting project is SALT at Stanford University: Self Archiving Legacy Toolkit. It is directed at the paper and digital archives that emanate from distinguished academics[80].

(2) There are four key phases: (i) the paper elements of an archive are digitised, yielding PDF files, which are subsequently integrated with the born-digital elements of the archive; (ii) the born-digital and digitised documents are automatically organised, tagged and mapped supported by editing and validation by the creator; (iii) further, the creator is asked to add insights and contextual knowledge through annotation; and (iv) the originator is invited to make the archive or some documents public on the web (with knowledge discovery facilitated by advanced web technologies), or to pass the documents to the University Archives and Digital Repository.

(3) The project is paying special attention to the design of semantic information processing aided by the creators.

(4) It is intended to develop the concept so that current researchers and faculty members can manage their personal information, their 'living archive'.

(5) It is understood that as more collections are added, links between them will yield a network of innovation and influence to be explored.

Future access and discovery

(1) A sign of the progress in software that is using advanced technologies and is directed at enriching personal information is provided by Photocopain[81]. It is based on semantic technologies: combining (i) contextual information obtained by a software instrument known as the Semantic Logger with (ii) image analysis and content-based annotation (using EXIF metadata embedded in digital photos for example), and presenting the information to the user through the AKTiveMedia image annotation tool that allows individuals to vet the automatically proposed annotations for images harvested from a website such as Flickr for example.

---

[79] J. Atkinson-Abutridy, C. Mellish, S. Aitken (2003) A semantically guided and domain-independent evolutionary model for knowledge discovery from texts, IEEE Transactions on Evolutionary Computation 7:546-560; R. Feldman and J. Sanger (2007) The text mining handbook. Advanced approaches in analyzing unstructured data, Cambridge University Press, Cambridge

[80] SALT: Self Archiving Legacy Toolkit, http://stanfordluminaryarchives.googlepages.com/salt, and for an initial demo see http://stanford.edu/group/salt_project/cgi-bin/feigenbaum/

[81] M. M. Tuffield, D. P. Dupplaw, K. O'Hara, N. R. Shadbolt, A. Chakravarthy, C. Brewster, F. Ciravegna and Y. Wilks (2006) Photocopain - annotating memories for life, 5 December 2006, AKT IRC, http://www.aktors.org/; M. M. Tuffield, S. Harris, D. P. Dupplaw, A. Chakravarthy, C. Brewster, N. Gibbins, K. O'Hara, F. Ciravegna, D. Sleeman, N. R. Shadbolt and Y. Wilks (undated) Image annotation with Photocopain, http://www.aktors.org; obtainable from eprints of the University of Southampton

(2) Another approach, SAPHARI, has provided semi-automatic annotation of batches of photos with person recognition based on clothing and event-based clustering of photos taken at around the same time or incident.

(3) Much work has been undertaken towards automated redaction of documents bearing personal information. Typically, it involves the use of look-up dictionaries, regular expressions (powerful fine tuned searching such as GREP) and simple heuristics. One successful trial was conducted in the automated de-identification of free text patient medical records; it was concluded that the algorithm was not of sufficient accuracy for processing publicly disseminated information, but nonetheless the open source software succeeded in outperforming a single human de-identifier[82].

(4) A recent, and rapidly notorious, instance of an inadvertent release of privacy information occurred when the web provider AOL made available supposedly-anonymised data pertaining to the search behaviour of 650,000 users. Within a very short time it became evident that it was possible to infer the identity of many of the people represented by the data; this was because queries often include so-called vanity searches that contain the individual's own name[83]. It clearly demonstrates the importance of recognising the issue, and developing effective and automated de-identification or controlled identity recognition procedures.

(5) Another area of increasing relevance to modern archives is visualisation as a means of portraying relationships between entities and the way these relationships evolve through time. A case in point is the Themail as a visualisation approach to email archives (see §9 for elaboration).

6.4 Methods: Online Service Providers

An informal survey of the websites of some online service providers (OSPs) was conducted with the terms of service, privacy and copyright statements being copied for careful examination.

List of some online service providers

30 Boxes >> ADrive >> Bebo >> Blist >> Box >> BT Digital Vault >> Carbonite >> Dandelife >> Delicious >> Dipity >> DropBox >> Dropio >> Facebook >> Flickr >> Gmail >> Humyo >> LegacyLocker >> LiveDrive >> LiveJournal >> MediaFire >> MeetWithApproval >> MobileMe >> Mozy >> Multiply >> MySpace >> Ning >> Photobucket >> Prosper >> RememberTheMilk >> Scribd >> SecondLife >> Sliderocket >> SlideShare >> StoryMash >> Storytlr >> StumbleUpon >> ThinkFree >> Twitter >> Vimeo >> Windows Live Sync >> YouSendit >> YouTube >> Zoho

Three overlapping classes of online service provider were considered: personal data storage; sharing and informal publishing of personal files and content; and online application services. The services arXiv, OurMedia and SourceForge do not belong in the same class of online service providers as the others, and for comparative purposes were also examined.

---

[82] I. Neamatullah, M. M. Douglass, L.-w. H. Lehman, A. Reisner, M. Villarroel, W. J. Long, P. Szolovits, G. B. Moody, R. G. Mark and G. D. Clifford (2008) Automated de-identification of free-text medical records, BMC Medical Informatics and Decision Making 8, issue 32

[83] D. Butler (2007) Data sharing threatens privacy, Nature 449: 644-645

The aim was to obtain a broad picture of the kinds of conditions under which OSPs expect their users to operate[84].

6.5 Findings: Online Service Providers

The terms

(1) It is striking how varied are the presentation and details of these legal statements.

(2) Some stipulate for example that after 30 days of any unpaid charges, the user's account may be terminated or suspended. Others indicate that if the license expires and is not renewed or is terminated or discontinued for any reason the company may deny access or delete content without notice. Others give a notice of a month before terminating a service. A number give much less notice: days rather than weeks.

(2) Even so most 'terms of service' have a clause somewhere that appears to give the company the right to terminate an individual's account without notice and for any reason. One example, from Blist: "Notwithstanding any of these Terms of Service, Company reserves the right, without notice and in its sole discretion, to terminate your right to use the Site… and to block or prevent future [sic] your access to and use of the Site…".

(3) In a number of cases the service providers are quite explicit that the company makes no claim for the intellectual property of the content placed on their service websites, acknowledging that it belongs to others, most especially the user of the service. Other service providers are much less explicit in this regard.

(4) Some service providers allow the user to download to his or her local computer (ie backup) much of the content of the user's space on the website but not necessarily all of it. Facebook is one OSP which does not allow a user to download all of his or her content, and press and media sources have reported an attempt by Facebook in February 2009 to modify the terms of service such that it reserved digital ownership of the content for itself even if an individual closed his or her account[85].

(5) Unsurprisingly, perhaps, liability and any guarantees for ensuring the preservation of the content is simply avoided. A typical example: "You are solely responsible for creating backup copies of and replacing any user content you post or store on the site at your sole cost and expense". Even in the case of online service providers that specialise in the backing up of data, it is expressly emphasised that only temporary storage is being provided, and that individuals should back up their information locally themselves.

- One company, namely Carbonite, is explicit that it "does not maintain a secondary copy of your data that you have Backed Up to our servers", and indicates that it will only make commercially reasonable efforts to create a replacement in the event of the company losing your data.

---

[84] Note that terms of service and other legal information are subject to change, June 2009

[85] F. Vogelstein (2009) The great wall of Facebook. The social network wants to dominate a new, friendlier internet - and keep Google out. There's going to be war, Wired, American edition, pp 96-101, 120

- Dropbox states: “You acknowledge and agree that you should not rely on the Site, Content, Files and Services for any reason. You further acknowledge and agree that you are solely responsible for maintaining and protecting all data and information that is stored, retrieved or otherwise processed by the Site, Content, Files or Services. Without limiting the foregoing, you will be responsible for all costs and expenses that you or others may incur with respect to backing up, and restoring and/or recreating any data and information that is lost or corrupted as a result of your use of the Site, Content, Files and/or Services”.

(6) While the importance of privacy is widely acknowledged by the online service providers, the ‘terms of service’ frequently mention not only the need to comply with requests from legal authorities, but other exceptions such as the need to respond to claims that the data violate the rights of third parties, and even the need to deal with technical problems. These may well be reasonable but do need to be understood.

(7) The privacy conditions, like the general ‘terms of service’, are typically subject to change at short or no notice.

(8) Many of the services allow for access to information uploaded by a user to be restricted to the individual or to family or friends or to be publicly available. Sometimes the public option is the default; in other cases the private option is the default. The requirement that private information of any other individual (third party) including addresses, phone numbers and email addresses is not uploaded may be understandable but it is not always clear what counts as private information.

(9) Furthermore restrictions are placed with regard to what an individual may hold and share with others (ostensibly even within a restricted group of friends and family, and even for the purposes of backing up). In theory this apparently includes – for example – (i) instances of content that the individual knows to be false or misleading, and (ii) a photograph of a person who has not given consent. What is less clear is the reality in practice: the extent to which services are terminated; the extent to which restrictions are enforced; and the extent to which data are actually lost; there are, however, anecdotal instances reported.

(10) The other key consideration is that an individual’s use of a service or site is subject to (ie governed by) the laws of a county and state in another country, typically the United States of America; and the user agrees to unconditionally submit to the jurisdiction of the courts in these places.

(11) LegacyLocker offers to pass to an individual’s friends and family information about his or her online accounts and profiles following his or her demise. But it is not certain to what extent other online service providers in their entirety will respect the wishes of the individual. Some service providers do explicitly indicate that they will make content available to the family where this reflects the wishes of the deceased. To what extent is an individual legally permitted to access another person’s social networking or webmail account even with the permission of the holder? This question should be researched in detail.

Summing up: online service providers

(1) Evidently these services in their present form cannot be relied upon solely for longterm storage and archival access. It is strongly advisable for content to be backed up locally by individuals perhaps ultimately using a more advanced form of Hoppla or some alternative.

(2) Individuals should be encouraged to maintain locally an account of their wishes and details of their online accounts. The hard drives and other media of the originator can – with permission of the family - be searched by curators in order to find URLs and email addresses (as is sometimes done by members of the forensic community); this would help serve as a check but it is always possible that some will be missed.

(3) Notwithstanding these concerns, there is a manifest and almost universal desire for the services being offered by web 2.0 companies, reflected in very considerable commercial success at least in numbers of users if not always immediately in advertising revenues.

(4) Moreover, there is evident interest by commercial entities in the advancing use of online services for personal storage and (hopefully) archiving. It is not irrelevant that the Digital Lives research conference and project have attracted the interest of representatives from Flickr (Yahoo), Google, Nature online, Microsoft Research, Amazon and Second Life[86].

(5) Many of these organisations are already using personal information in various ways beyond the direct furthering of advertising income, and in some circumstances are making it available to academic researchers for genuine scientific analysis (eg Facebook)[87].

(6) A key strategy must be to encourage the public to reward those services and products that support sustainable personal archives: favouring the capture by individuals of their own personal content, longterm preservation, and continuing reuse of their own personal digital objects. It is necessary therefore to make people aware of current limitations, and for consumer and professional computer magazines to be encouraged to promote digital preservation and archival purposes. One useful route for alerting the public might be Which Online?, as mentioned in an earlier section.

(7) Future research could examine the frequency at which the terms of service are changed and the extent to which individuals are supplied directly and actively with records of these terms and any changes.

(8) From the archival perspective there are two considerations: (i) there are opportunities to provide a *bona fide* quality service that meets basic archival standards and expectations; and (ii) if demand is there it is likely to be realised.

(9) There are a number of ways that public archival repositories might establish partnerships or agreements with commercial online service providers to their mutual benefit. For example a system for endorsement with vetting for individual rights and archival suitability could be offered, perhaps in collaboration with privacy and rights organisations

(10) In view of the diversity of presentation of the terms and conditions of use by Online Service Providers, it would be helpful to users and allow for greater transparency if OSPs

---

[86] The First Digital Lives Research Conference, Personal Digital Archives for the 21st Century, British Library, 9-11 February 2009. Further evidence of interest is shown by the recent convening of a meeting, led by Jeff Ubois, on Personal Archives, Saving Our Present for the Future: Personal Archiving 2010, 16 February 2010 attended by representatives of Microsoft Research, Yahoo! Research, PARC, Internet Archive, and others, http://www.archival.tv/personal-archives/

[87] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer and N. Christakis (2008) Tastes, ties, and time: a new social network dataset using Facebook.com, Social Networks 30:330-342

would be encouraged to adopt an approach akin to that recommended for repositories: namely the use of layered legal metadata with icons for users to readily understand.

Extracts from the terms of online service providers

Carbonite

"You are solely responsible for protecting the information on your computer such as by installing anti-virus software, updating your applications, password protecting your files, and not permitting third party access to your computer".

Dipity

"Dipity reserves the right to refuse service to anyone for any reason at any time".

"Dipity claims no intellectual property rights over the material you provide to the Site. Your Content remains yours. You can remove Content you submitted at any time."

"By submitting any Content to the Dipity Site, you agree to grant Dipity and its users permission to access and use your Content. You hereby grant to Dipity a worldwide, assignable, non-exclusive, sublicensable, royalty-free license to publish, use, reproduce, distribute, transmit, adapt, modify, create derivative of, provide user access to, publicly perform and publicly display, in every manner and medium now or hereafter known, any and all Content (in whole or in part) solely for the purpose for which such Content was submitted."

Dropbox

"Dropbox does not claim any ownership rights in Your Files."

Google

"Google does not claim any ownership in any of the content, including any text, data, information, images, photographs, music, sound, video, or other material, that you upload, transmit or store in your Gmail account. We will not use any of your content for any purpose except to provide you with the Service".

Humyo

"Humyo reserves the right, in its sole discretion, to reject, restrict, suspend, or terminate your access to all or any part of the Humyo Services at any time, with or without prior notice".

"Humyo does not claim any ownership rights in the text, files, images, photos, video, sounds, musical works, works of authorship or any other materials (collectively, 'Content') that you upload to Humyo Services."

"Humyo assumes no responsibility for any error, omission, interruption, loss, deletion, defect, theft, destruction or unauthorized access to, or alteration of any Content you upload to the Humyo Services".

"The HUMYO Sites are hosted in the United Kingdom and are intended for and directed to users in the European Union. If you are a User accessing the HUMYO Sites from outside the United Kingdom, through your continued use of the HUMYO Sites, which are governed by United Kingdom law, you are transferring your personal information to the United Kingdom and you consent to that transfer. In the event that HUMYO is acquired by or merged with a

third party entity, we reserve the right, in any of these circumstances, to transfer or assign the information we have collected as part of such merger, acquisition, sale, or other change of control."

Legacy Locker

"Legacylocker, Inc. shall have the right at any time to change or discontinue any aspect or feature of Legacy Locker, including, but not limited to, content, hours of availability, and equipment needed for access or use."

"Legacylocker, Inc. provides the User with the necessary resources, includng account log in information, to transfer the use and control of important websites, such as email accounts, Facebook, MySpace, YouTube, and eBay to a designated beneficiary upon the death or disability of the User (collectively 'Services').

- User enters account information, password, instructions, and designated beneficiary for each website.
- Legacy Locker contacts designated beneficiaries to inform them that they have been selected by User.
- Upon death or disability of User, email is sent to designated beneficiaries OR designated beneficiaries contact Legacy Locker with proof of death.
- Designated beneficiaries are granted information to access User websites."

"Either Legacylocker, Inc. or User may terminate this Agreement at any time".

Livedrive

"Livedrive does not claim any ownership rights in the text, files, images, photos, video, sounds, musical works, works of authorship, or any other materials (collectively, 'Content') that you upload to the Livedrive Services."

"Livedrive is not responsible for any error, omission, interruption, loss, deletion, defect, theft, destruction or unauthorized access to, or alteration of any Content you upload to the Livedrive Services."

MobileMe: Apple

"Apple may at any time, under certain circumstances and without prior notice, immediately terminate or suspend all or a portion of your account and/or access to the Service."

"Effects of Termination Upon termination of your account you lose all access to the Service... . In addition, Apple shall delete all information and data stored in or as a part of your account(s) including, but not limited to, data files, email, albums and preferences."

Mozy

"If this Agreement terminates, other than for your failure to comply, Decho will use commercially reasonable efforts to make your Data available for you to download for a period of three (3) days. Decho has no obligation to provide you with a copy of your Data and may remove and discard any Data."

"You are solely responsible for your conduct and your data related to the Service. You agree to indemnify, defend, and hold harmless Decho and its suppliers from any and all loss, cost, liability, and expense arising from or related to your data, your use of the Service, or your violation of these terms."

MySpace

"MySpace is not responsible for and makes no warranties, express or implied, as to the User Content or the accuracy and reliability of the User Content posted on or through the MySpace Services... ."

"You are solely responsible for ensuring that all Content that you post on or through any of the MySpace Services, and any material or information that you transmit to other Users, conforms to all applicable data protection and privacy laws, including ensuring that you have obtained the prior consent of any and all individuals whose personal information you use and/or disclose on or through MySpace."

"MySpace respects the intellectual property of others, and requires that our users do the same. You may not upload, embed, post, email, transmit or otherwise make available any material that infringes any copyright, patent, trademark, trade secret or other proprietary rights of any person or entity."

Windows Live including Live Sync: Microsoft

"Online Backup Feature. The Online Backup feature provides the capability to store and retrieve your digital photographs from our services *via* the Internet during the applicable Online Backup Subscription Period... . If you cancel your Online Backup subscription or your subscription lapses, the copies of your photos stored with Online Backup will be deleted immediately."

Zoho

"The Zoho Services are made available to you by AdventNet subject to the following Terms."

"We respect your right to ownership of content created or stored by you. Unless specifically permitted by you, your use of the Services does not grant AdventNet the license to use, reproduce, adapt, modify, publish or distribute the content created by you or stored in your Account for AdventNet's commercial, marketing or any similar purposes"

"You agree that AdventNet may terminate your Member Account and access to the Services for any reasons including, but not limited to, breaches or violations of the Terms or the Zoho Privacy Policy, a request by you to terminate your Account... ...unexpected technical issues or problems, extended periods of inactivity... ."

Postscript: related work

(1) A closely relevant study has been conducted by Jane Kaye and Heather Gowans of The Ethox Centre, Department of Health, University of Oxford[88]. It selected 10 websites and examined the terms and conditions. The websites selected were: Yahoo!, Google, YouTube, MySpace, Facebook, Blogger, Flickr, Twitter, NetMums and Delicious.

(2) Kaye and Gowans likewise noted that the thrust of the terms is to protect the service provider, and while some protection of personal data is offered, there are no guarantees regarding security. Should the controls on privacy be set high? Are users sufficiently aware? Is it reasonable to expect users to understand the implications of the privacy policies?

---

[88] J. Kaye and H. Gowans (2009) Sites for social networking and user generated content, IMDE Research, Innovative Media for a Digital Economy, University of Oxford, February 2009

(3) It is suggested that “there is no reference in the legal terms... to any external regulatory body operating as a ‘watchdog’ over service providers and the use of individual’s personal data”.

(4) The brief report also indicates that there are “advocates for so-called ‘data portability’ (allowing individuals’ to easily transfer information in and out of the web services they use), and others, such as Facebook, who argue that it needs to safeguard the information that it stores so that it is not misused”.

6.6 Brief Overview of Technology and Services

(1) A series of tools and perspectives useful to the curation of digital lives has been identified across the lifecycle. One of the key concerns to emerge from consultation with both users and curators is that of authenticity, provenance and integrity of the whole.

(2) In seeking suitable tools two requirements were: (i) to make it possible where desired to capture the whole space of a person’s computer media (and not just independent files); and (ii) to replicate and retain demonstrably exact copies of the original files, recognising their historical and informational value (and not rely solely on digital facsimiles, even if these match modern standards for interoperability).

(3) At the same time the practical utility of converting files to digital facsimiles for immediate access by creators in the wild and users in repositories is recognised.

(4) Crucial to the chosen approach are: (i) the creation and maintenance of hash values akin to unique ‘digital fingerprints’ for each file; and (ii) a system of provenance, custody and audit control. Of great interest is the Advanced Forensic Format (AFF) for forensic ‘imaging’ of disks.

(5) For this purpose forensic software offers solutions that are of immediate applicability in a number of ways, and that serve as quasi models for what is needed. The use of forensics progressed at the British Library and by the Digital Lives project is now being adopted across the world from Oxford and continental Europe to the archival institutions of University of Texas, Emory University and Stanford University for instance.

(6) The potential of features within existing forensic applications - for example automated but controlled searching and organising of files across a whole personal collection - could be explored for enabling user access to personal collections.

(5) Moreover, this approach has demonstrated the feasibility and practicality of adopting existing technologies and off-the-shelf commercial products – in some cases designed for the consumer and inexpensive, with others aimed at professional specialists but being open source.

(7) A pragmatic philosophy is to provide for immediate access to basic text, images and sounds (eg raw alphanumeric content which will suffice for many scholarly purposes); but to retain (by capturing and keeping exact digital replicates of disks and files) the potential to make available high fidelity versions that respect original styles, layout and behaviour. A further series of tools have been identified for capturing and making available eMSS derived from ancestral computers, most especially for emulation of hardware and software classic computing activities.

(8) In highlighting specific tools it was endeavoured to embrace as far as possible: (i) adequate and good quality but inexpensive commercial software; and (ii) open source software. Some exciting examples are coming out of the Planets project including Dioscuri for emulation, and Hoppla for archiving by small institutions and individuals.

(9) A review of a selection of online service providers included those that focus on: (i) backup and storage such as Carbonite, (ii) restricted sharing or full publishing of personal files such as Flickr; and (ii) online application services such as Zoho.

(10) A preliminary examination of 'terms of service' revealed a number of cases: (i) where the service provider reserves the right to suspend and terminate its service without warning and at its sole discretion; (ii) where a service provider will on termination make commercially reasonable efforts to make an individual's data available for download for a very limited period without being under any obligation to provide a copy and being free to discard any data; (iii) where immediate deletion is indicated if the subscription lapses; and, commonly, (iv) where liabilities and warranties are avoided.

(11) Nonetheless it is clear that OSPs offer a service that is desired by many people.

(12) It is recommended that repositories and archival institutions endeavour to make people generally as well as OSPs aware of the issues, especially regarding longterm sustainability, and of what is necessary, so that those commercial online service providers that do (hopefully) provide archivally-sound features are rewarded by more customers.

(13) The possibility of mutually beneficial partnerships and collaborations between online service providers and public repositories should be explored.

(14) Perhaps the most important single technology would be the ability for individuals to readily make safe their own personal digital archive. This should continue to be explored vigorously and combined with an archival perspective on personal information management.

(15) The close and explicit (eg in the Terms of Service) involvement of a regulatory body should be considered.

Chapter 7: Curation

7.1 Introduction with Objectives[89]

The primary aim of the workshop and associated activities was to seek the views and suggestions of archivists and curators who handle personal archives, and to establish some priorities and recommendations[90&91].

7.2 Procedures

An informal though detailed questionnaire was sent to prospective workshop participants and others. This was followed by an initiating workshop, held in September 2008. Subsequently, input from other archivists was sought internationally and professionally, in part through the Digital Lives Research Conference, especially Day 1.

Workshop for curators and archivists

The workshop consisted of two sessions, morning and afternoon, both attended by all invited participants.

Morning
During the morning session an overview of the Digital Lives project and initial findings was provided by Digital Lives team members, which was followed by an outline of experiences with personal digital archives given by Jeremy Leighton John, Susan Thomas, Natalie Walters, Ifor ap Dafydd and Gary Brannan.

Afternoon
In the afternoon, participants first worked together in pairs addressing a series of topics, ad then there was an open discussion, with some possible recommendations concluding the session.

List of participants attending on the day along with Digital Lives team members:

- Ifor ap Dafydd, National Library of Wales
- Fran Baker, John Rylands Library, University of Manchester
- Guy Baxter, Victoria and Albert Theatre Collections

---

[89] This chapter is the first report of the findings of the project's consultation with professional archivists and curators, and was prepared by Jeremy Leighton John based on notes during the workshop and informal questionnaires

[90] Emerging background references on practical aspects of the curation of personal digital archives include J. L. John (2005) Because topics often fade. Letters, essays, notes, digital manuscripts and other unpublished works, in M. Ridley, editor, Narrow roads of gene land. The collected papers of W. D. Hamilton. Volume 3. Last words, Oxford University Press, Oxford, pp 399-422; S. Kim, L. A. Dong and M. Durden (2006) Automated batch archival processing: preserving Arnold Wesker's digital manuscripts, Archival Issues 30(2); 91-106; Paradigm (2007) Workbook, http://www.paradigm.ac.uk; C. Stollar Peters (2006) When not all papers are papers: a case study in digital archivy, Provenance 24: 23-35; S. Thomas and J. Martin (2006) Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives, Journal of the Society of Archivists 27: 29-56

[91] Originating references include: A. Cunningham (1994) The archival management of personal records in electronic form: some suggestions, Archives and Manuscripts 22: 94-105; A. Cunningham (1999) Waiting for the ghost train: strategies for managing personal electronic records before it is too late, Archival Issues 24: 53-64; T. Hyry and R. Onuf (1997) The personality of electronic records: the impact of new information technology on personal papers, Archival Issues 22: 39; A. Summers and J. L. John (2001) The W. D. Hamilton Archive at the British Library, Ethology, Ecology & Evolution 13: 273-291

- Tilly Blyth, Science Museum
- Gary Brannan, West Yorkshire Archive
- Gareth Burfoot, British Library
- Maxine Clarke, Nature Publishing Group
- Sue Donnelly, London School of Economics
- Annette Faux, Laboratory of Molecular Biology, University of Cambridge
- Stella Halkyard, John Rylands Library, University of Manchester
- Oliver Urquhart Irvine, British Library
- Alysoun Sanders, Macmillan Group of Publishers
- Dorothy Sheridan, Mass Observation Archive, University of Sussex
- Bill Stockting, British Library
- Susan Thomas, Bodleian Library, University of Oxford
- Dave Thompson, Wellcome Library
- Natalie Walters, Wellcome Library
- John Wells, University of Cambridge Library
- Lynn Young, British Library

Topics

Nine principal topics were considered during the day:

- The nature of the personal archive
- The current and emerging situation: opportunities and challenges
- What needs to be done from the perspective of curation and archiving?
- What needs to be done with regard to networking, professional collaboration and the provision of advice?
- The possibility of participation by both creators and users in the archival process, and the impact of digital rights
- The skills and training necessary for dealing with personal digital archives
- Professional matters; the merits or otherwise of a group specialising in personal digital archives and possible means of communication
- What steps need to be taken in the near future?
- What recommendations should be made?

7.3 Findings

Workshop

Personal archives

(1) Personal archives are valued for: (i) providing unique views of events and speaking for the occasion; (ii) representing individuals and their richly individualistic activities, and - for instance - their relation to the community and state; (iii) offering evidence of the underlying processes behind achievements and historical events; (iv) enabling in particular the creative process to be documented, and (v) bringing the world of the past alive in evocative ways.

(2) Personal archives are: (i) derived from diverse media, (ii) less structured and unstructured, (iii) characteristically unedited, and (iv) idiosyncratic in organisation and content. It is through this primary source material, however, that scientific history, social history, and literary history and scholarship are enabled.

Opportunities

(1) More personal information is being created, and it will be possible in principle to store and preserve more; future collections will be more comprehensive and more inclusive.

(2) More and more people are documenting their own lives.

(3) Tagging is already happening, and participation by creators and others in metadata creation and other aspects of curation is fast materialising.

(4) Clear benefits to be gained are remote and wide access by multiple users, across collections with more flexible and powerful analysis of information becoming available.

(5) There will be a ready ability – where digital rights permit - to provide facsimiles (if not replicates).

(6) A particular theme that was highlighted was the opportunity for archivists and curators to establish new ways of archiving, to have an impact at this juncture.

Challenges

(1) It is difficult to assess the scholarly or monetary value of digital archives or digital elements of archives. The owners will no longer know what is on an obsolete disk or tape. There is also the impact of uncertain authenticity and provenance on value.

(2) The sheer quantities involved, with a proliferation of versions, is making appraisal specifically, and the overall curation generally, complex and demanding.

(3) In addition to the burgeoning quantities of contemporary digital media and files, there is also a growing legacy of neglected personal digital objects that require retrospective processing urgently.

(4) Hybrid collections remain and, in particular, the volumes of paper and other analogue components continue to grow in size.

(5) Small archival institutions are insufficiently resourced; recommendations need to be feasible for small institutions.

(6) Digital preservation policies are not currently suited to personal archives.

Attitudes and expectations

(1) There is a need to manage as well as meet the expectations of users along with those of creators and depositors.

(2) Managers often do not understand the implications for processes and resources; there is a sense that digital materials are demanding attention on top of everything else that needs to be done.

(3) The overall cost at least in establishing the capacity to handle personal digital archives is greater than the day-to-day requirements of analogue objects; in short there is a failure of resources to meet the initial setup cost.

(4) There are also the costs of developing new skills, the need for which will not disappear since technology is moving so fast and, moreover, legal and ethical requirements and interpretations correspondingly evolve.

(5) The attitudes of creators towards personal digital objects are not the same as those towards analogue objects such as diaries and notebooks.

(6) There are deep-seated issues to be addressed relating to the form and detail of cataloguing, and to the nature of access and resource discovery.

What needs to be done? What are the priorities?

(1) Curators and archivists expressed a universal desire to be provided with suitable training, in particular for digital capture and introductory forensics, and for the creation of interoperable files and relevant aspects of digital preservation.

(2) There is a corresponding wish for instruction manuals that can be adopted by all archivists, and which enable them to advise in turn the digital public.

(3) Scalable solutions that can be scaled down as well as up are required. This will require the development of robust, straightforward and pragmatic workflows

(4) Solutions that allow large volumes of information to be handled by small numbers of archivists are required, and mechanisms and procedures for enabling the appropriate participation of creators and users of personal digital archives.

(5) Establish a register or list of expertise, with centres of excellence; and enable collaboration between computer museum specialists and archivists and their respective institutions.

(6) Promote the provision of resources, and the value and requirements of personal digital archives.

(7) Consider ways of influencing software development.

(8) There was little enthusiasm for developing ontologies in the context of the curation of personal digital archives and eMANUSCRIPTS.

Questionnaire

Points Made

(1) There was vigorous support for hands on training so that even local archivists can undertake digital capture (eg through forensic techniques), creation of interoperable files and other processes.

(2) Support for facilities such as a wiki and specialist journal or newsletter for personal digital archiving was met with considerable enthusiasm but this was qualified by concerns about longterm viability and a desire to work with existing conduits. The use of a chat facility was also recommended. There was some support for a specialist forum dedicated to digital aspects of personal archives.

(3) Archivists were very receptive to the involvement of creators, family members, researchers and users in the provision of metadata, although the need for some kind of quality control or tagging was recognised (eg to distinguish between curatorial and other input).

(4) There was strong encouragement for the 'living archive', recognising the need to approach individuals regarding their archives (eg eminent writers and scientists) sooner rather than later, and to capture complementary information that provides context.

(5) It was noted that most small and local archives might not be able to afford a dedicated digital archivist. With most archives being hybrid in the foreseeable future, the need for multitasking curators who are able to work with both digital and paper archives was emphasised. Digital specialists can advise and support other curators (eg regionally in the case of local archives).

(6) On the whole curators felt generally well informed regarding Freedom of Information, Data Protection and Copyright (as these are pertinent to paper archives) but nonetheless would welcome clear legal guidance specific to eMANUSCRIPTS.

(7) It was widely agreed that - in frequently being hybrid – personal archives would be best served and make the most of digital potential for powerful access if the paper component would be digitised so that it can be accessed along with the eMANUSCRIPT component.

(8) There was emphatic encouragement for international and cross-disciplinary collaboration.

Miscellaneous extracts from the questionnaires

What is the primary value of personal archives? What is the primary value of personal digital archives?

"The primary value of personal archives is in 'telling the story' of the individual in the cultural context. The particular value of digital archives is that they have a much greater functionality."

"Personal archives give the 'insider's view', not hearsay, not someone else's opinion but the thoughts and workings of the individual. They may contradict or reinforce other material but they give a personal quality that is often missing from other material and give a richness to the overall picture"

"Personal digital archives are likely to provide even fuller records of their creators' lives, e.g. emails are more informal than letters and often capture the kind of exchanges previously made only in telephone calls (although conversely, we may lose some of the record unless creators have particularly good record-keeping practices – e.g. drafts of literary works composed on a word processor might be overwritten rather than saved as distinct versions)."

What does your institution do with personal digital archives or what would they probably do if presented with them (eg if a floppy disk was found among 'personal papers')?

"It is not unknown for there to be catalogue entries along the lines of '2 floppy disks'."

Would your institution deal with the digital materials itself?

"This is something that we need to decide during the coming year, and it all depends on what kind of resources the management is willing to commit to digital preservation. Personally, I think that… …we should be trying to deal with digital material in-house, although I realise that the relevant technical expertise is something we currently lack. I would be wary about outsourcing from a data protection and confidentiality point of view as much of the digital material we take in is likely to be highly sensitive."

"Commercial data recovery firms are not considered sufficiently 'archival' or sensitive to context of the material to provide a meaningful service."

How does or would your institution store a personal digital archive?

"In acid-free envelopes. (no kidding). At present we try to stop any ingress of digital material."

Is there an official policy statement for personal digital archives?

“At present our ‘policy’ is not to collect digital archives. However, some material has got through the net….”

How does or would your institution balance Freedom of Information with Data Protection…?

“In the case of digital archives, I think we would have to ask donors to impose closures – either until we can look at material in detail (which would usually be at cataloguing stage), or until such time as data protection is no longer an issue (although this would be an extremely lengthy period and far from ideal). In our deposit agreement… [the inclusion of a statement has been suggested] to the effect that ‘uncatalogued digital archive material will normally be closed to all except you and Library staff’.”

“We would be extremely reluctant to take material that donors or depositors would not want to make available for a long length of time, and would try to persuade them otherwise. This has caused problems in the past, with one organisation using the archive as a storage facility and then expecting Archivists to retrieve and send records to them. The material they were offering us would have to be extremely high value for us to be willing to store and work on material that could not be made available to readers immediately.”

How would or does your institution go about meeting copyright, intellectual property, requirements?

“In relation to access, I understand that unless we have obtained explicit permission from copyright holders, we would not be able to make even the non-sensitive content of digital archives accessible remotely. This is likely to apply to the vast majority of material in recent and contemporary archives….”

How do you feel about the situation regarding personal digital archives: regarding your institution and regarding the wider world?

“I don’t think there is enough thought to this and not enough policies/guidelines provided by institutions to either archives or to individuals….”

“I feel that we are currently failing our living donors/depositors to some extent, in that… we are asking them to print off key records which we should really be taking in digital form, and these records are therefore losing some of their most important characteristics, as well as their original context….”

“In the ‘wider world’ I feel that perhaps the issue of personal digital archives does not have a particularly high profile, and that there’s a failure to recognise that archives are at risk. Much of the advice/guidance and standards issued by the digital preservation community focuses on digital assets like image libraries, institutional repositories etc, not on personal or ‘collected’ archives…; similarly, [there is a] focus on electronic records management….”

“I do feel quite positive really – we do have the chance, now, to create a remarkably broad picture of the times we live in, and to ensure that as many voices, attitudes and opinions are heard in the years to come. Certainly though, the situation is not perfect by any stretch of the imagination. There is a definite reluctance on the part of archivists to engage with personal collections as they know them in a digital format, presumably stimulated by a lack of knowledge about how to curate such collections.”

“I feel that nobody has really got to grips with the reality of this – current projects are providing useful pointers and ways ahead but are often producing solutions way out of the reach of the majority of archives. Does this mean that only the favoured few will be able to deal with digital archives, what does that mean in an age where information and access are supposed to be democratised – is this just a temporary phase and will more flexible and immediate solutions be found. In terms of providing interesting research materials I fear the end of serendipity.”

“My institution sees this opportunity and the planning and implementation is methodical but exciting. My own opinion is that this is a revolutionary trend in the organization of people’s material for themselves and for sharing, and will greatly change the way we plan and live our lives”.

Do you feel that the archival and curatorial community has the situation in hand, under control?

“I do not think the majority of the archival and curatorial community have the situation in hand. There is a need for greater education about working with digital materials to reduce the fear that many Archivists and Curators have. This makes them ignore digital material, and means that they are creating more problems for the future.”

“The archival & curatorial community has raised a lot of very good questions regarding the issues over digital archiving, and many technological solutions and models have been identified. We need more practical examples of how these can be worked through for personal digital archive curation as well as for organisations.”

"I don't think we do have the situation under control – in fact I'm not sure it will ever be possible to have the situation in hand as technology morphs at the speed it does. Besides, I don't think we've ever had the situation under control. I think we might have thought we had BUT digital archives have forced us to address this misconception."

What are the main obstacles to further progress?

"Need more smaller institutions to get on board with this and give their experiences – currently too much of this work is focused on larger and relatively well funded organisations – this will need some strong funding and training."

"The cost of digitization and the careful navigation of copyright and privacy protection".

When should we approach the creators or originators of personal digital archives?

"The key is targeting the right people, and getting them to think about what is important to keep. Looking at digital archives should be no different to looking at personal letters. You need to convince people that it is important to have their story, their views, to enrich their story with the personal detail. However, to leave it too late can cause problems, for example they may have encrypted or password protected their data, the material may be more difficult to index/catalogue/arrange without input from the creator".

"So I think we get involved as soon as possible with the creators of archives. Also it is a misconception that creators of archives are 'unconscious' about their archive creation in an unproblematic way. Its way more complex that this and archivists are already in the equation."

"We need to raise awareness as early in the life of a digital object as possible, to help ensure the survival of that which should be preserved for posterity. We will need to combine a general advocacy role to all depositors, with a specific approach to prominent or potentially prominent individuals."

Would you be open to the notion of users (eg known or assessed experts, or the originators themselves) assisting with the creation of some of the description or contextual information…?

"Yes, I think this is an important aspect of getting the most out of the material (in fact I do this where possible for analogue collections as well, e.g. photographs, artifacts etc) but it can make the process more time-consuming and it is useful to set deadlines so that the process moves along and doesn't get stagnant (often material turns up when people are retiring and they have other priorities.)"

"Yes this is desirable and absolutely vital to understand this material in context in the future. Some websites already do this on a large scale (e.g. Flickr) but it is not practical on a large individual scale, where people have developed archives at home not supported by web technology. Individual archives are rarely perceived to be of significance when they are first created, but subsequently."

"Very open – a great idea for both groups. Very practical and desirable – we can't do it all ourselves anyway!"

What is the role of the curator, the archivist, of personal archives in the digital era?

"…I don't think that Archivists will need to be ICT professionals, but I do think that Archivists will need to posses some of the skills [that the] ICT professional currently possesses. The real theory driven, archival science, side of the career will still exist, but will, I imagine, become more desk based…. We will, in essence, become remote data providers. I see us spending our time working on ingest and management… with access being essentially on demand (or as near on demand as can be)."

Is it feasible, useful, for curators and archivists and their institutions to provide a digital advisory service to enable people to manage their personal digital archives?

"There is an increasing interest in personal and family history and individuals who see the need to maintain a personal archive and a personal digital advisory service would be useful particularly at a local level."

"[Y]es that is the way to go. A service that can demonstrate the usefulness of organizing material in an archival way, and demonstrating a path to preservation... will reduce the barriers to donation."

"Ideally, archivists, digital curators, and information institutions should be in the position to advise individuals on the preferred ways to manage and preserve their digital records. But many institutions currently have difficulty implementing these procedures for their own records and employees."

Is there a need for training of archival and curatorial specialists?

"Training of current Archivists and Curators is the most urgent task as these are the people being presented with digital personal papers and not knowing what to do with them."

"I think there is a need for clear policies and advice for all archivists and curators, which could take the form of training. Most archivists/curators already possess a wealth of skills that are just as useful to digital materials as they are to analogue materials. These offer a good foundation that can be built upon to tackle the special requirements of digital materials".

"…and if young archivists are 'future proof' from the start this would help immensely. However, this does not discharge older archivists, and indeed I feel there is a need for training targeted at older archivists or those firmly committed to the physical sphere. This cannot be detailed, in-depth digital preservation theory, but enough to allow the participant to function as an archivist in a digital world…."

What is your perception of how demanding are or will be personal digital archives in terms of resources?

"Staff will be one of the key expenses, as the material will need to be processed quite quickly after being received and material will need to be reviewed regularly to ensure it continues to be in a useable format".

"Possible ways of finding savings:

Working collaboratively with other institutions undertaking similar work….
Working collectively to influence software companies to develop commercial repository products suitable for the requirements of personal digital archives.
Alternatively, encourage developers working on open source repository software to develop functions, interfaces etc. suitable for working with digital archives."

What would help? What tools, guidelines, strategies, tactics, advice, actions would be most useful now?

"Compulsion."

"I would like to see some standards for online representation of data so the data may be 'mashed up' across repositories".

Would a wiki site be actively used by archivists and curators...?

"[M]oderated chat rooms work for dialogue, wiki works for posting documents like specifications. I personally would value a chat room dialogue."

"Yes, interactive technologies such as blogs and wikis would be helpful in keeping up with best practices... ."

Would a forum for bringing together archivists and curators... with the more technically minded... be useful, feasible?

"Yes this would be very valuable to match up what is possible with what is desirable. I would add historians and researchers. I could imagine the cross-presentation of latest projects and expression of needs, facilitated into a list of requirements".

"[N]etworking and collaboration are the way these sorts of problems should be tackled... ."

Do you think many, perhaps most, personal archives will be hybrid…?

"I imagine they will be hybrid for some time. It seems common for scientists to keep printouts/paper archives of things considered to be 'important' whereas digital seems, psychologically speaking, to be more ephemeral. We see this attitude regularly with regard to the print/online editions of journals – for example, scientists think that an online edition is more easily change[d] after formal publication, without seeming to see that it is part of the published record just as print. Also, we know that authors prefer to have their papers in print edition of journals rather than with online-only supplements. The relative importance people ascribe to the digital vs print media seems to depend on which 'hat' they are wearing when thinking about it. The common wisdom is that attitudes will change, but people do not seem to know when."

"I think that, like the paperless office, the entirely digital personal archive is a myth, at least for the foreseeable future. Although there are some people out there who scan paper and then destroy the originals, there are more who print out documents to read and amend."

What qualities and abilities should an archivist or curator of personal digital archives possess?

"Above all else curiosity. Everything else can be taught."

"The same qualities that all archivists should ideally possess plus, in order to have confidence with digital materials, a high level of IT competence. I can send the person spec for my post… if required. It would cover everything from qualifications and training in archival management, evidence of recent and ongoing staff development, excellent communications skills, ability to teach and lead, ability to act effectively and strategically within parent institution when competing for resources, design exhibitions, run groups including with children of all ages, host adult education and professional groups, deal with the media including TV & radio interviews, handle all legal aspects of deposit, access and intellectual property, able to catalogue to ISAD G level and knowledge of all relevant portals and gateways eg Archives Hub, TNA etc, have experience of conservation methods and the ability to take care of different materials in storage, ability to design and develop websites, desk top publishing and the design and production of publications, leaflets, promotional… materials[,] ability to manage and give leadership… and run volunteer schemes, able to communicate in all sectors of education including with senior academics, be very well organised and efficient and know how to fund-raise and apply for grants."

7.4 Key Suggestions and Outcomes

(1) Opportunities as well as some of the challenges presented by personal digital archives were identified by a group of archivists and curators representing local, university and national institutions.

(2) Two significant findings that emerged were: (i) the strong perception that digital preservation and curation research and development - too date - have been insufficiently oriented towards the needs of repositories of personal archives and that this is even more so for local and small repositories lacking the scale of resources of national institutions; and, nonetheless, (ii) a strong desire for all archivists of contemporary collections to be able through suitable training to engage in the hands on processes of digital capture and curation – including forensic techniques and enhanced curation.

(3) There was recognition of the concept of the 'living archive', and the need to approach individuals sooner rather than later, as well as the benefits of capturing and creating complementary contextual information.

(4) Archivists were very receptive to the involvement of creators, family members, researchers and users in the provision of metadata, although the need for some kind of quality control - perhaps through metadata tagging - was deemed necessary.

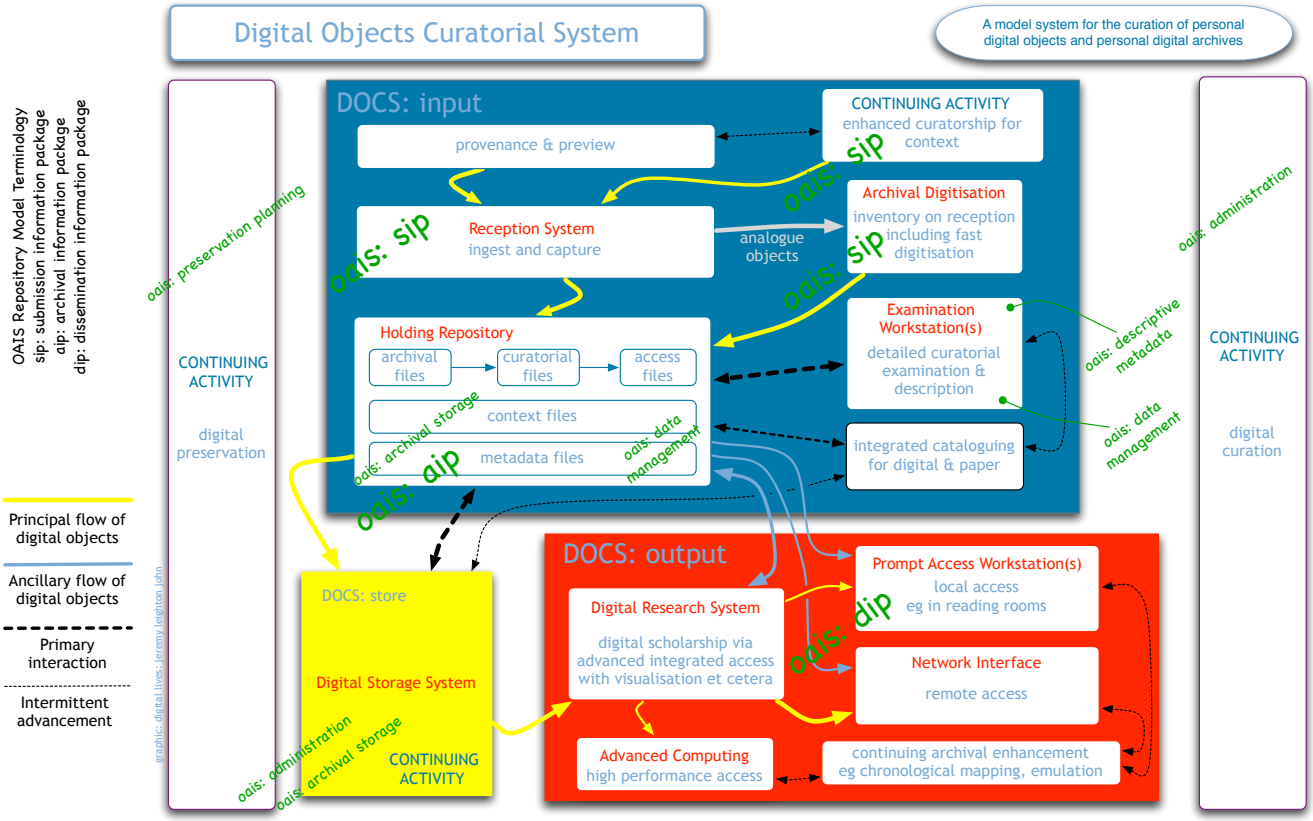
(5) Prompt digitisation of personal papers was adjudged to be a most effective way to integrate the digital with the paper elements of the hybrid archive.

(6) The notion of communities of practice was highlighted, and there was strong support for a forum, group communication, and a wiki or chat facility. A further step in this direction was taken at the Digital Lives research conference. The topic of digital lives and the bringing together of people with interest in personal digital archives was the theme of the first day of the project's conference at which international colleagues contributed[92].

[92] Further discussion took place in April 2009 at the University of North Carolina at Chapel Hill at a one day workshop organised by Cal Lee in association with the DigCCur 2009 conference. A discussion group has been established by Cal Lee, Personal Digital Archives Working Group, PDAWG at the DigCCurr website, http://ils.unc.edu/digccurr/institute.html, specifically the Digital Curation Exchange

(7) There is a need for more straightforward tools for: (i) individuals generally, (ii) local archivists, and (iii) and institutions in the developing world.

(8) Tools might be oriented towards the individual with the possibility of scaling up if resources permit it. In this way people generally would be served and the potential benefits of significant quantities of personal life information being captured across society would be enabled. This approach would ensure that the lives of writers, scholars, scientists and influential figures are captured early on - even before their pre-eminence and contributions are anticipated or understood.



Chapter 8: Information Media, Networks and Manipulation[93]

8.1 A Digital Revolution or Evolution: Personal Computing and the Internet

(1) This chapter seeks to look briefly at the rapid changes occurring in the digital era within the context of the past. There is change but also underlying continuity. The chapter concludes with an outline of possible future trends, highlighting in particular the potential value of personal digital archives to individuals and society, and most especially to research, supported by online curation.

(2) The digital revolution in its current manifestation springs from the interaction between personal computing, the graphical user interface with computer mouse or touch screen, the world wide web and others areas of the internet, and the desire of many individuals to create and share content, to socialise and to collaborate.

(3) As noted already, the International Data Corporation has predicted that by this year, 2010, nearly 70% of the digital universe will be created by individuals rather than organisations[94].

(4) Alongside this phenomenon has been the revitalised development in recent times of the information technology of interacting with handheld devices through gestures and subtle and dexterous finger control.

(5) These are revolutionary changes but it is possible to look back and see a trend, to observe the changing use of information media through history.

8.2 A Historical Perspective

(1) The making of tools, the use of the hand to investigate and interact with the environment, the creation of ornaments and the manipulation of material symbols can be traced far into the Palaeolithic[95].

(2) There are a number of defining phases in the long history of humanity but among the most momentous are the initial acts of storing symbolic information outside of the brain.

(3) Even the repeated making of sophisticated tools and ornaments – long considered the originating acts of creativity of humans – leads to devices and structures that bear

---

[93] This chapter was prepared by Jeremy Leighton John, and its research perspective is further elaborated in a paper entitled: iCuration: from I to i paper (see §15)

[94] J. F. Gantz (2007) The expanding digital universe. A forecast of worldwide information growth through 2010, An IDC White Paper sponsored by EMC, Framingham, Massachusetts; J. F. Gantz (2008) The diverse and exploding digital universe. An updated forecast of worldwide information growth through 2011, An IDC White Paper sponsored by EMC, Framingham, Massachusetts. Note that much of the individually created digital universe will reside in organisations according to the same pair of reports

[95] F. d'Errico, C. Henshilwood, G. Lawson, M. Vanhaeren, A.-M. Tillier, M. Soressi, F. Bresson, B. Maureille, A. Nowell, J. Lakarra, L. Backwell and M. Julien (2003) Archaeological evidence for the emergence of language, symbolism, and music - an alternative multidisciplinary perspective, Journal of World Prehistory 17:1-70; C. S. Henshilwood (2007) Fully symbolic sapiens behaviour: innovation in the Middle Stone Age at Blombos Cave, South Africa. In P. Mellars, K. Boyle, O. Bar-Yosef and C. Stringer, editors, Rethinking the human revolution: new behavioural and biological perspectives on the origin and dispersal of modern humans, McDonald Institute for Archaeological Research, Cambridge, pp 123-132

information, hints of how each was made, used and admired.

(4) Archaeologists are still deliberating the ways gesture, language, music, tool making and symbolic storage interacted and mutually influenced each other to yield the essentials of emerging modern humanity[96]. The matrix that unifies these phenomena is wide ranging sociality, manifested in the existence of diverse cultural mores, portable media and instruments, and networks of information exchange and learning.

(5) Over the centuries and millennia a wide variety of types of external information media and instruments of representation and symbolism have been adopted. Examples have included (in alphabetical order): antler, bamboo and clay, through papyrus, parchment and plaster, to stone, vellum and wooden tablet[97]. All have been used to code anthropic information: some monumental, some portable.

(6) With the passage of time the fundamentals of the expressive hand, information storage media and information networks have yielded profound social changes[98]. A major transition – merging the prehistoric into the truly historic - was the origin of written communication: the birth of written documents, of scripted recording of oral traditions after eons of oral inheritance and replication, and the building and transfer of accumulated information and knowledge from one generation to the next.

(7) But there is another deep effect much less often expressed: the gradual emancipation of individuals through history occurring with local and temporary setbacks and re-beginnings. It is an unending, inconsistently directed and unsteady process but seemingly inevitable – the striving of emancipation through the evolving power of writing, printing and network communication.

(8) Throughout much of history complex, beautiful and elaborate writing has been the prerogative of scribal specialists serving the elite and pre-eminent. Even the more informal everyday accounting of daily life - represented by the writing of notes of loans, debts, receipts, taxes and lists as well as occasional news and greetings and tokens of remembrance from distant places - would have been done more in the context of the lives of the well-to-do than the poorest of individuals.

(9) Yet, in any event, with regard to the more informal and personal writing, little has survived from ancient and early medieval times that speaks directly and comprehensively of everyday life; and even in the case of the most powerful people sources with convincing first hand details of the personal activities, thoughts and feelings of individuals - their private voices - are few and far between.

---

[96] N. J. Conard, M. Malina and S. C. Münzel (2009) New flutes document the earliest musical tradition in southwestern Germany. Nature 460: 737-740; M. C. Corballis (2009) Language as gesture, Human Movement Science 28: 556-565

[97] A.-M. Christin, editor (2002) A history of writing from hieroglyph to multimedia (originally published in French, 2001), Flammarion, Paris; M. P. Brown (1998) The British Library guide to writing and scripts. History and techniques, The British Library, London

[98] D. McNeill (1992) Hand and mind: What gestures reveal about thought, Chicago University Press, Chicago; R. T. Kellogg (1994) The psychology of writing, Oxford University Press, Oxford; F. Barbieri, A. Buonocore, R. Dalla Volta and M. Gentilucci (2009) How symbolic gestures and words interact with each other, Brain & Language 110: 1-11

(10) With exhaustive scholarship and a little bit of luck it has been possible to obtain glimpses of everyday life and language that are profoundly evocative. Examples range from: (i) letters from the frontier on wooden leaves at the Roman fort of Vindolanda near Hadrian's wall; (ii) domestic scenes in the margins of the Lutrell Psalter; (iii) naturalistic Flemish illuminations of Simon Bening showing local people engaged in daily tasks; (iv) surviving phrases of apparently genuine colloquial dialogue of Old English in the 11th century versions of Apollonius of Tyre; and (v) the incomparable collection of letters from several generations of the Paston family of Norfolk during the 15th century. Beyond Europe there are treasures of daily life from Mesopotamia to Dunhuang. But these are exceptional[99].

(12) Since the 16th and 17th centuries, paper has helped to open the way for widespread personal writing, printing has facilitated literacy and a more robust postal system has promoted distant communication. Even so unless an archive has gone to a repository or collector, personal and home archives have been kept in numbers by only the wealthiest families. It is during these centuries that history begins to witness lives in richer and extensive detail though it has taken intellectual skill and perseverance for historians to achieve it.

(13) An important process in capturing everyday lives of individuals, beginning in the 20th century, has arisen from the concept of mass observation, an "anthropology of ourselves". It is proactive, and entails the observing and recording of the lives of people in public by researchers, and the inducing of volunteers to keep diaries[100]. Another proactive and highly regarded and effective method has been oral history, where researchers conduct careful and detailed life story interviews, again originating in the 20th century, with the ability to record sound easily[101]. These activities in the predigital era have provided an important route to capturing everyday life though requiring considerable effort and organisation.

(14) With the emergence of personal computing and the TCP/IP internet in the 1970s, with the world wide web in 1989 and the proliferation of web 2.0 online participation, mass collaboration and crowd sourcing in the 21st century, an increasing emancipation of individuals is being recognised. Emancipation in the form of wide if not yet universal communication and collaboration, a flourishing of content creation and sharing and publishing.

(15) Equally far reaching, there is a significant sense of 'living online', of individuals carrying out more and more of their day-to-day activities through the computer, and currently through the wirelessly networked gesture-sensitive, touch screen, handheld device. The rising popularity of laptops, portable handhelds and 3G smartphones points to an increasing mobility with digital tools, devices and network access portals closer to the individual and more personal.

(16) Although there is some way to go in the wide adoption of the idea and practice of maintaining a personal digital archive, and doing so in a sustainable way, already many

---

[99] A. K. Bowman (2003) Life and letters on the Roman frontier. Vindolanda and its people, The British Museum Press, London; M. P. Brown (2006) The world of the Luttrell Psalter, The British Library, London; D. Crystal (2005) The stories of English (originally published 2005), Penguin Books, London; H. Castor (2004) Blood & Roses. The Paston family in the fifteenth century, Faber and Faber, London; S. McKendrick (2003) Flemish illuminated manuscripts 1400-1550, The British Library, London

[100] N. Hubble (2005) Mass observation and everyday life, culture, history, theory, Palgrave Macmillan, London

[101] R. Perks and A. Thompson, editors (2006) The oral history reader, Routledge, London; http://www.ohs.org.uk/

individuals and their families are holding their personal documents and memorabilia – digital and analogue - in quantities unheard of for past generations.

8.3 Benefits: iSCIENCE with Life Information

(1) Life information is potentially of enormous value to scientific advancement. All kinds of benefits can be proposed for medical research, social science and natural science: such as the appraisal of health in relation to lifestyle, social origins and environmental circumstances. For greatest research applicability it is necessary to embrace people generally. With moment-by-moment representations of relationships between people extended through long periods, personal archives will allow the dynamic and structure of interactions and communications among individuals and with their environment to be better understood[102].

(2) Longitudinal studies of cohorts of individuals are an important resource for social and political scientists. A case in point is the UK National Child Development Study that has sought the participation of 17,000 individuals born in one week in 1958. The typical instruments are postal communications with a few questions, more lengthy occasional questionnaires and personal face-to-face interviews. This carefully premeditated gathering of structured information is demanding on resources although it will be essential to continue and extend the approach as a baseline for quality data.

Box: Avon longitudinal study of parents and children, and the human epigenome

The Avon Longitudinal Study of Parents and Children (ALSPAC)[103] is based at the Department of Social Medicine at the University of Bristol. Sometimes referred to as "Children of the 90s" it is a longterm health research project founded on the participation of more than 14,000 mothers who agreed during pregnancy in 1991 and 1992 to contribute to the project. Ever since, both mothers and children have yielded an immense amount of genetic, medical, psychological and environmental information, producing a data resource that is supporting research by scientists across the world. Recent research papers have shown, for example, that women who are anxious, under stress, during pregnancy - especially late on - are more likely to have asthmatic children; another study has revealed that blood lead levels in early childhood are associated with effects on performance in school examinations and behaviour; other studies have looked at diet and aspects such as the timing of introduction of types of food to young children. A particular focus of the project is the interaction between genetic and environmental factors.

Recently, the project has contributed to pioneering studies of epigenetics involving an elegant combination of historical study, developmental biology and human genetics[104]. The story begins with some historical research by Dr Lars Olov Bygren of people living in the far north of Sweden in Norrbottens, a county that straddles the Arctic Circle and one that was subject in the 19th century to both bountiful years and famine years; using detailed agricultural records, he found indications that extreme environmental events (eg very poor harvest) have impacts that can be detected in subsequent generations (eg longer lifespan in grandchildren).

---

[102] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy and M. Van Alstyne (2009) Computational social science, Science 323: 721-723

[103] http://www.bristol.ac.uk/alspac/

[104] J. Cloud (2010) Why genes aren't destiny, Time magazine 175(2): 27-31, 18 January 2010

There is growing support for the notion of epigenetic inheritance[105]. In a collaboration involving the Institute of Child Health of University College London, the Karolinska Institute of Sweden and the ALSPAC research team, evidence of male-line transgenerational effects was found: for example, fathers who started smoking early in childhood were compared with those who started later; after various controls it was found that the male (but not female) offspring of the early smokers revealed greater body mass index[106].

With the human genome mapped, attention is now being directed at mapping the epigenetic markers, and the Human Epigenome Project[107] has already been launched; a pilot study investigated the Major Histocompatibility Complex - the most polymorphic region in the genome and one that is linked to many biomedical phenomena including the identification of self by the immune system[108]. The potential for using epigenetics to improve welfare is considerable but the Human Epigenome Project is an even greater undertaking than the Human Genome Project, requiring not only significant advances in computing but also an in depth understanding of the relationship between epigenetic markers and DNA, developmental and organismal consequences as well as environmental and social circumstances; and, of course, informed consideration of ethical implications. Access to life information of the kind found in personal archives could be crucial.

---

(3) Personal papers, historical manuscripts such as diaries of travellers, letters of diplomats, logbooks of ships' officers and local family archives have long yielded geological, meteorological, sociological and natural history data. In the digital era, much more 'freestyle' personal information can be beneficially garnered in the form of personal digital objects and archives for scientific analysis.

(4) A further indication of the research value of personal archives is provided by the Enron Email Dataset (which though derived from a corporate setting has provided a rare insight into the way email is actually used). It represents about 150 email users, mostly senior management of Enron, a major energy company that was subject to a very public financial investigation. To quote William W. Cohen who is making available the email corpus of roughly 500,000 messages: "This data is valuable; to my knowledge it is the only substantial collection of 'real' email that is public"[109].

---

[105] E. Jablonka and M. J. Lamb (2005) Evolution in four dimensions. Genetic, epigenetic, behavioral, and symbolic variation in the history of life, MIT Press, Cambridge, Massachusetts

[106] M. E. Pembrey, L. O. Bygren, G. Kaati, S. Edvinsson, K. Northstone, M. Sjostrom, J. Golding and the ALSPAC Study Team (2006) Sex-specific, male-line transgenerational responses in human, European Journal of Human Genetics 14: 159-166

[107] http://www.epigenome.org

[108] V. Rakyan, T. Hildmann, K. L. North, J. Lewin, J. Tost, A. V. Cox, T. D. Andrews, K. L. Howe, T. Otto, A. Olek, J. Fischer, I. G. Gut, K. Berlin and S. Beck (2004) DNA methylation profiling of the human Major Histocompatibility Complex: a pilot study for the Human Epigenome Project, PLoS Biology 2(12): e405, 2170-2182

[109] http://www-2.cs.cmu.edu/~enron/. The data were originally made public by the Federal Energy Regulatory Commission as part of its investigation. An example of the use of this corpus is provided by D. Zhong, D. Gatica-Perez, D. Roy and S. Bengio (2006) Modeling interactions from email communication, IEEE International Conference on Multimedia & Expo (ICME 2006), pp 2037-2040. There are smaller sets of emails available, of course: the recent 'climate change' emails might be one, although at this time their legitimacy is controversial

Box: Personal networks, communication and mobility: emails, phone calls

There has in the last decade or so been an explosion of research activity directed at social aspects of communication networks from the internet to the mobile phone[110]. A major focus of the empirical analysis has been the seeking of patterns and principles in the communications of networks and communities. Palla and colleagues analysed records of mobile phone calls and collaborations of coauthors in order to investigate complex structure of interconnected friends, families and professionals[111]. Clear differences were found in the dynamics of small and large groups and their composition and level of commitment of individuals. Other researchers were able to use information about mobile phone calls and the location of the individuals (inferred from locations of mobile phone towers) to gain insight into the relations between individuals, and even predict the existence of friendship with 95% accuracy; it was also revealed that self-reports of physical proximity by participants were subject to recency and saliency memory effects leading to deviations from mobile phone location data[112]. González and colleagues tracked the locations of 100,000 individuals again using mobile phone information for six months, and found that despite the diverse nature of the travel histories of individuals, the pattern characteristic for individuals was simple and highly reproducible, with most time spent at a few locations and with a high degree of regularity temporally and spatially[113].

Another study looked at email dynamics using the 'to', 'from' and 'time' fields in the log files: self-organised structures emerged akin to others found with nonlinear dynamic systems[114]. A key property of email is its ready capacity to involve more than two people in joint communication compared with the conventional use of the phone, and of course to document the communication. It was possible to distinguish between different kinds of groupings of the email correspondents: pertaining to temporary goal-oriented committees, and to static, organisational units such as institutional departments.

In an elegant analysis of the highly prolific correspondence of two scientists in the era of letter writing, it was reported that the communication dynamics of Charles Darwin and of Albert Einstein follow the same scaling law as modern email exchanges; on the other hand the scaling exponent distinguished letter communication from email[115]. More specifically, it was shown that the time taken to respond to a letter approximately follows the power law $P(\tau) = \tau^{-\alpha}$ where τ is the probability that a reply to a letter will take place in τ days, and $\alpha=3/2$ (and

---

[110] D. J. Watts (2007) A twenty-first century science. Nature 445: 489; B. Howard (2008) Analyzing online social networks, Communications of the ACM 51: 14-16; J. Kleinberg (2008) The convergence of social and technological networks. Communications of the ACM 51: 66-72; N. A. Christakis and J. A. Fowler (2009) Social network visualization in epidemiology, Norsk Epidemiologi 19: 5-16; D. Krackhardt (2009) A plunge into networks, Science 326: 47-48

[111] G. Palla, A.-L. Barabási and T. Vicsek (2007) Quantifying social group evolution, Nature 446: 664-667

[112] N. Eagle, A. S. Pentland and D. Lazer (2009) Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences USA, early edition, preprint

[113] M. C. González, C. A. Hidalgo and A.-L. Barabási (2008) Understanding individual human mobility patterns, Nature 453: 779-782

[114] J.-P. Eckmann, E. Moses and D. Sergi (2004) Entropy of dialogues creates coherent structures in e-mail traffic, Proceedings of the National Academy of Sciences USA 101: 14333-14337

[115] J. G. Oliveira and A.-L. Barabási (2005) Darwin and Einstein correspondence patterns, Nature 437: 1251; J. G. Oliveira and A.-L. Barabási (2006) Barabási & Oliveira reply, Nature 441: E5-E6

can be contrasted with studies with email where α=1). Both Darwin and Einstein replied to only a fraction of the letters that they received (just short of one third and one quarter respectively); but if replying mostly did so within ten days.

Kossinets and Watts analysed the emails of more than 40,000 students using the email headers, and emphasised the broad changes in social networks in time, manifested in the topology and structure of a network, with global properties appearing to approach a stable state even while properties concerning individuals are unstable[116].

As Kossinets and Watts note, some studies have found email in local social circles to be strongly correlated with face-to-face and telephone interactions[117]; others have found substantial differences between email and face-to-face interaction networks, eg a negative correlation between email and face-to-face communication[118]. This is one reason why a microscopic analysis of relations involving, for example, content of email as well as headers and a comprehensive set of communication pathways can be expected to be even more revealing; it is still early days, as a number of the contributing researchers have alluded[119].

Along with the potential to serve the human sciences, an understanding of personal networks can serve the individual too. Nardi and colleagues have directed research at enabling users to create their own visualisation maps for modeling and organising their individual contacts and the interrelations among them; initial tests were conducted in the workplace[120]. The software ContactMap extracts personal networks by analysing the individual's history of email interactions, proposes the network to the user who then amends it and classifies the contacts; a key aim was for the visualisation to support the memory of contacts, context and nature of relationships.

---

(5) The monetary value of information held by online service providers tends to lie in the area of advertising and individually targeted marketing. The research value of personal data held by social networking sites (SNSs) is already well evidenced, with a recent study looking at social networks based on this kind of information made available by Facebook[121]. Kevin Lewis and colleagues give an indication of what is feasible using data obtained from this SNS. Although there has been extensive research into social networks and longstanding interest in their manifestation through the internet, social networking sites such as Facebook offer another order of richness in research data. Information that resides on user accounts include background about school days, hometown, demographic information such as birthday and gender, and interests and cultural tastes from political affiliations to music, literature and

---

[116] G. Kossinets and D. J. Watts (2006) Empirical analysis of an evolving social network, Science 311: 88-90

[117] see also B. Wellman (2001) Computer networks as social networks, Science 293: 2031-2034

[118] F. Grippa, A. Zilli, R. Laubacher and P. A. Gloor (2006) E-mail may not reflect the social network, North American Association for Computational Social and Organizational Science Conference 2006 (NAACSOS 2006), Notre Dame, Indiana; see also D. Olguin Olguin, P. A. Gloor and A. S. Pentland (2009) Capturing individual and group behavior with wearable sensors, AAAI Spring Symposium on Human Behavior Modeling, Association for the Advancement of Artificial Intelligence, Stanford, California

[119] eg Oliveira and Barabási (2006), ibid

[120] B. A. Nardi, S. Whittaker, E. Isaacs, M. Creech, J. Johnson and J. Hainsworth (2002) Integrating communication and information through ContactMap, Communications of the ACM 45: 89-95

[121] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer and N. Christakis (2008) Tastes, ties, and time: a new social network dataset using Facebook.com, Social Networks 30: 330-342

movies; moreover users document their 'friendships' and share online photo albums as well as their own picture. Yet, Lewis and colleagues argue, "only a very small portion of the wealth of data available on Facebook" has been exploited; and so with permission from Facebook and the student institution, a new and versatile network dataset was sought; it was formed by downloading the profile and network data from a cohort of students.

(6) An even more detailed resource is emerging from activities known as 'life tracking', an ongoing capturing of personal metrics. People now record not only their activities (and 'inactivities' such as sleep) but track physiological conditions, respond to automated requests to indicate level of happiness, monitor coffee and alcohol consumption, and use of television and computer, diet and so on. Many life trackers like to share their data with other lifetrackers (eg mothers compare sleep patterns of their babies)[122]. The phenomenon of smartphone Apps has further promoted and strengthened this tendency[123].

Box: Personal science, life telemetry

The use of portable, mobile and handheld devices by individuals opens up a new universe of personal information that might be made available to responsible researchers for analysis of social, cultural, medical, epidemiological and environmental conditions. In the first instance, there are, of particular interest, the technologies that allow individual lives to be documented with wearable devices such as sociometric badges being used to understand social interrelations and conversational patterns. These technologies might be used, are being used, to identify personality types or to appraise human interactions: from the way people respond to novel products that seek to improve well being and safety to the way people use the space in their home or workplace.

Civil engineering and interior design can benefit from a greater understanding of how buildings and structures are actually used. A British construction company specialising in building houses tracked the members of a family electronically using bracelets with RFID tags, as part of Project: LIFE house. The family lived for weeks at a time in a 'concept home' fitted with 26 sensors throughout the house that regularly located family members. The aim was to help establish the optimum design for the layout of the building. David Burbeck, of Design for Homes, a nonprofit research organisation based in London concluded that the most interesting aspect of the study was its "attempt to clarify how people spend their time"[124]. A more modern system is reported by Libal and colleagues (2009). Netcarity is a project that seeks to improve the wellbeing, safety, independence and health of older people through new technologies[125].

---

[122] J. O'Reilly (2009) Your life as a number. All you do - from sexual activity to carb consumption - can be reduced to data. And by publishing your data online, other people can help you spot patterns. Welcome to the life-tracking trend, Wired, UK Edition, pp 144-149; G. Wolf (2009) Know thyself. The personal metrics movement goes way beyond diet and exercise. It's about tracking every fact of life, from sleep to mood to pain, 24/7/365, Wired, American Edition, pp 92-95. This enthusiastically quantitative phenomenon can perhaps be seen as one manifestation of a more general 'life logging' that incorporates a more informal qualitative digital capturing of personal lives, notably through image and sound recording; see G. Bell and J. Gemmell (2009) Total Recall. How the e-memory revolution will change everything, Dutton, New York

[123] R. Fisher (2009) Welcome to Appland. Is there nothing a smartphone can't help you do better? New Scientist, 22 August 209, pp 33-36

[124] B. Carney (2005) Home design's house of clues, BusinessWeek online, 24 August 2005

[125] http://www.netcarity.org/About.11.0.html. In addition to the technologies themselves, the project is exploring issues of privacy, ethics and data fusion

Correspondingly, Olguin Olguin and colleagues have used sociometric badges in corporate and organisational environments to enable 'sensible organisations' through automatic measurements of organisational behaviour. Olguin Olguin and Pentland have applied their sociometric badges in various environments to recognise daily human activities such as sitting, walking, standing and running ascertained *via* an acceleratometer; extract speech features to enable identification of nonlinguistic social signals such as interest and excitement; enable radio communications to be exchanged between badges of users and to transfer data to base stations; perform indoor localisation; communicate with bluetooth cell phones and other handheld devices; capture face-to-face interaction with an infrared sensor. In this way highly detailed and diverse information can be captured for hundreds of people in an unobtrusive way[126].

The use of electronic badges that automatically measure an individual's social activities can incorporate amount of face-to-face interaction, conversational time, prosodic style, physical proximity to other people, and physical activity levels. Social signals are derived from vocal features, body motion and relative location. Obvious applications exist in healthcare, intelligent hospitals, knowledge management and collaboration, and improvements to organisational and business operations. User experiences in virtual worlds and social networking sites could be enhanced by adding mobility and real world information gathered by the sensors[127].

It has also proved possible to identify individual personality traits (among five broad categories) as well as group performance based on data obtained from wearable sensors[128].

Beyond the urban and indoor environments, there is enormous potential too for research outdoors and not only for the social and human sciences but for the natural sciences.

Citizen science is a growing activity[129] where individuals participate in diverse science projects on a voluntary basis though not necessarily with the same intensity as an amateur scientist pursuing a scientific hobby or, of course, a professional scientist[130]. It has been anticipated for some time that mobile phone technologies would allow the environment to be monitored for health and local weather through people using wearable sensor units as manifested in the Nokia Eco Sensor concept for example[131]. The built-in geolocation of many mobile phones (iPhone and others) and dedicated geolocation devices existing today already brings the possibility of citizen scientists monitoring the environment of city, suburb and

---

[126] D. Olguin Olguin, B. Waber, T. Kim, A. Mohan, K. Ara and A. S. Pentland (2009) Sensible organizations: technology and methodology for automatically measuring organizational behavior, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39; D. Olguin Olguin and A. S. Pentland (2008) Social sensors for automatic data collection, 14th Americas Conference on Information Systems, Toronto

[127] D. Olguin Olguin and A. S. Pentland (2007) Sociometric badges: state of the art and future applications, IEEE 11th International Symposium on Wearable Computers, Boston, Massachusetts

[128] D. Olguin Olguin, P. A. Gloor and A. S. Pentland (2009) Capturing individual and group behavior with wearable sensors, AAAI Spring Symposium on Human Behavior Modeling, Association for the Advancement of Artificial Intelligence, Stanford, California

[129] For example: http://citizensci.com and https://app.secure.griggith.edu.au/03/toolbox

[130] L. Lyon (2009) Open science at web-scale: optimising participation and predictive potential, Consultative Report v1.0, JISC: UKOLN & Digital Curation Centre, University of Bath

[131] http://www.unwiredview.com/2007/12/10/nokia-eco-sensor-phone-concept

countryside if professional scientists produce suitable sensors. CyberTracker Conservation based in South Africa (and receiving some funding from the European Commission) has conducted a number of projects in monitoring wildlife and the environment and has a vision of establishing a worldwide environmental monitoring network. To support efficient data collection the project has made available CyberTracker software for installation on a personal computer and companion handheld device such as a smartphone or personal digital assistant[132].

A group at the University of Cambridge has set up a system whereby people, eg cyclists, collect information *via* wireless sensors as they engage in their daily activities: temperature and other environmental variables[133]. Steed and Milton established trials with GPS-tracked sensors of CO (carbon monoxide), motivated by the possibility of mapping pollution (and other properties of the real world at a fine scale) through the deployment of a multitude of sensors borne by "members of the general public who would carry them as they go about their normal everyday activities"[134].

Various projects at the Center for Embedded Networked Sensing (CENS) at the University of California, Los Angeles, have enlisted citizens in submitting observations from the field *via* mobile phone: algal blooms, invasive species, plant blooming, seismic events[135].

There are also people who use wearable devices to participate in policy discussion as citizens or to make a sociopolitical point, one of the very earliest pioneers being Steve Mann who practices sousveillance (watching from below rather than from above) by wearing a digital device that records 'more or less' what he sees[136].

Increasingly individuals are recording personal and environmental information for themselves, for their own benefit: life logging and life tracking. Thus even without citizen or amateur participation in science or sociopolitical campaigns the information is gathered. Kanjo and colleagues emphasise that individuals might want to install sensors in their private space, indoors and outdoors, to be used for observing home, possessions and themselves and thereby

---

[132] http://www.cybertracker.org/Welcome.html and http://www.cybertracker.org/Help/Introduction_v3.htm. It is a short step from environmenal issues to more social issues, and the usefulness of the software in the serving the need for rapid surveys as part of disaster relief is also mentioned

[133] E. Kanjo, S. Benford, M. Paxton, A. Chamberlain, D. S. Fraser, D. Woodgate, D. Crellin and A. Woolard (2008) MobGeoSen: facilitating personal geosensor data collection and visualization using mobile phones, Personal Ubiquitous Computing 12: 599-607; R. Kwok (2009) Phoning in data, Nature 458: 959-961

[134] A. Steed and R. Milton (2008) Using tracked mobile sensors to make maps of environmental effects, Personal Ubiquitous Computing 12: 331-342

[135] Kwok (2009), ibid; this paper also quotes Susan Teltscher of International Telecommunications Union: "About 85% of the world's population has access to a mobile signal". In this context it is noteworthy that the mobile phone is significantly less expensive than a laptop or desktop

[136] S. Mann (1997) Wearable computing: a first step towards personal imaging, Computer 30(2), February 1997, http://wearcam.org/ieeecomputer/r2025.htm; S. Mann, A. Sehgal and J. Fung (2004) Continuous lifelong capture of personal experience using EyeTap, Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004), 15 October 2004, New York; S. Mann (2004) 'Sousveillance'. Inverse surveillance in multimedia imaging, Proceedings of the ACM International Conference on Multimedia 2004 (MM 2004), 10-16 October 2004, New York; S. Mann, J. Fung, C. Aimone, A. Sehgal and D. Chen (2005) Designing EyeTap digital eyeglasses for continuous lifelong capture and sharing of personal experiences, Computer Human Interaction 2005 (CHI 2005), 2-7 April 2005, Portland, Oregon; S. Mann, J. Fung and R. Lo (2006) Cyborglogging with camera phones: steps towards equiveillance, Proceeddings of the ACM International Conference on Multimedia 2006 (MM 2006), 23-27 October 2006, Santa Barbara, California

monitoring health, physiological and environmental conditions[137]. The SenseCam is among the most celebrated of wearable devices that has been used to benefit individuals directly by assisting memory recovery[138]; increasingly the technologies are embedded and seek ultimately to be unobtrusive. As a means of optimising the effectiveness and efficiency of daily activities, it has been proposed that individuals should be able to monitor their energy consumption and activities using a Personal Energy Meter, possibly embedded within an individual's mobile phone[139]. These metrics combined with better understanding of the impact and constraints of technology use would allow for more optimal use of resources. Already smartphone software 'apps' are set to transform daily lives, 'always on, always on you', playing "a role in everything from social interactions to emotions"[140].

An important point is that although many people are content to share portions of their life log, others do not wish to do so. The life log is kept private due to personal inclination or for security reasons. Perhaps the most comprehensive and evangelical of practitioners of life logging is Gordon Bell who emphasises his desire to keep much of his personal life log private[141].

It seems reasonable to suppose that many if not most people will not want the entirety of their personal digital archive to be freely available on the internet. Some other route is necessary in order for researchers to make use of personal digital archives.

A further obstacle is the unstructured nature of personal information and the large volumes of it. There are however a number of technologies emerging that can deal automatically with voluminous and unstructured information, with varying but increasing success: text mining, data fusion, semantic technologies, natural language processing, and automated recognition of people, artefacts (eg clothing), and plant and animal species[142]. Chippendale and colleagues have outlined a system for creating a calibrated vision sensor network using socially generated and georeferenced photos as means of gathering environmental information not normally available *via* satellites[143]: the technique employs the kinds of photos

---

[137] Kanjo et al (2008), ibid

[138] L. Laursen (2009) A memorable device, *Science* **323**: 1422-1423; see also http://en.wikipedia.org/wiki/Microsoft_SenseCam

[139] A. Hopper and A. Rice (2008) Computing for the future of the planet, Philosophical Transactions of the Royal Society A 366: 3685-3697; J. Walko (2009) Researchers ready personal energy monitoring devices, EE Times, http://www.eetimes.com/showArticle.jhtml?articleID=218000032

[140] S. Turkle quoted in R. Fisher (2009) Welcome to Appland. Is there nothing a smartphone can't help you do better, New Scientist, 22 August 2009, pp 33-36

[141] G. Bell and J. Gemmell (2007) A digital life. New systems may allow people to record everything they see and hear - and even things they cannot sense - and to store all these data in a personal digital archive, Scientific American, March 2007; G. Bell and J. Gemmell (2009) Total recall. How the e-memory revolution will change everything, Dutton, New York

[142] For example: K. J. Gaston and M. A. O'Neill (2004) Automated species identification: why not? Philosophical Transactions of the Royal Society of London, Series B 359: 655-667; A. T. Watson, M. A. O'Neill and I. A. Kitching (2003) Automated identification of live moths (Macrolepidoptera) using Digital Automated Identification SYstem (DAISY), Systematics and Biodiversity 1: 287-300

[143] P. Chippendale, M. Zanin and C. Andreatta (2009) Re-photography and environment monitoring using a social sensor network, 5th International Conference on Image Analysis and Processing (ICIAP 2009), Vietri sul Mare, Italy

that people typically take rather than carefully directed ones. These aspects are discussed further in §9.

8.4 Benefits: Humanities, Humanitarianism and Humanity

"I like very much people telling me about their childhood, but they'll have to be quick or else I'll be telling them about mine", Dylan Thomas[144]

(1) In turn, this greater scientific use of personal data obtained from people generally would potentially benefit the arts and humanities enormously, for it will help to direct funding towards the securing of personal digital archives, thereby strengthening and expanding historical, literary, philosophical, linguistic and political analysis.

(2) The 'digital capture imperative' means that archives need attention sooner rather than later; but if curators and archivists approach people very early in their careers it is difficult to reliably identify the most eminent statesmen and stateswomen, the most successful writers, the most influential scientists. By seeking to help sustain the archives of all people, the 'high profile' archives will be secured. Thus it is necessary to provide advice and engage with all individuals at a range of levels.

(3) There is a fundamental need in people for creativity, for witnessing creativity including their own (as this project helps to show, see §3). Equally there is a desire in many people for recognition of the hurdles overcome, accounts of past events and learning experiences to be captured. Research indicates that these phenomena are not trivial, involving as they do self esteem and psychological well being[145].

(4) Individuals gather and retain personal materials for their portfolio, for furthering their education and advancing their career.

(5) The personal library and archive can also serve as immediate sources of reference, accumulated during the life of an individual. An essential tool in learning and personal advancement is for an individual to know what he or she has consulted in the past, and having access to personally accumulated references with his or her own annotations and mark up.

(6) Families will pass on their personal digital archives through the generations. The enabling of longstanding personal and family memory, of keeping of personal and family archives, is fundamental to the process of emancipation, allowing digital objects of varying and diverse kinds of value to be passed from generation to generation.

(7) With more and more technologies and new forms of personal information emerging it is essential to monitor them, to continually research and try to understand their impact.

(8) With the emergence of personal fabricators such digital objects will represent not only text and image but also artefacts. Even analogue objects may be digitised using 3D scanning techniques, and enter the personal archive.

---

[144] D. Thomas (1954) Reminiscences of Childhood (Second Version), pp 8-14 in Quite Early One Morning, broadcasts by Dylan Thomas, J. M. Dent & Sons, London

[145] J. Etherton (2006) The role of archives in the perception of self, Journal of the Society of Archivists 27: 227-246

Box: Digital materiality

The study of literary and historical manuscripts and personal papers is rich with references to the sensation of paper, parchment and vellum, to the idiosyncratic markings unique to the individual and to the passing moment. In recent times the refrain has commonly been heard that the digital era means an end to all this immediacy, this physical intimacy. The antidote to this sense of loss is provided by Matt Kirschenbaum and his archival colleagues. The excellent "Mechanisms. New media and the forensic imagination" provides a window to the interests of future digital humanities scholars and of those in the vanguard today[146]. It leaves no doubt that future scholars will pore over hexadecimal code, will be schooled in the intricacies of ancestral word processing systems, and will be familiar with obscure as well as common disk filing systems and the physical nature of magnetic, optical and flash media. Volatility, transformission, original acts of erasure, extreme inscription, the digital work and poem Agrippa, the excitement of digital personal objects: these will be the parlance, emotions and landmark exemplars of future digital scholarship.

In 2001, Christine Finn, a contemporary archaeologist studying the history of computers, especially personal computers, from the perspective of enthusiastic often private collectors, pondered: "Maintaining the historical theme it may not be long before museums display computers that were used for writing a particular document. An award-winning novel, say, or a major film-script. It would be like seeing the pen used to sign a contract or a declaration, or a quill used by a writer"[147]. So it came to pass. Kirschenbaum and colleagues recently listed 17 notable authors with at least some born-digital material in US collections, and some of it has been put on display[148]: eg at the Technologies of Writing 2006 and The Mystique of the Archive 2008 exhibitions at the Harry Ransom Center, University of Texas, Austin, which included digital material from the Michael Joyce Papers and also a laptop from Norman Mailer, with the keys of the keyboard stained and worn with use (reportedly by his 'personal assistant', his digital amanuensis)[149]. Some of the digital media of the W. D. Hamilton Archive have been widely shown, and this personal archive has been joined by others - both scientific

---

[146] M. G. Kirschenbaum (2008) Mechanisms. New media and the forensic imagination, The MIT Press, Cambridge, Massachusetts

[147] C. A. Finn (2002) Artifacts. An archaeologist's year in Silicon Valley (originally published in 2001), The MIT Press, Cambridge, Massachusetts, pp 147-148

[148] M. G. Kirschenbaum, E. Farr, K. M. Kraus, N. L. Nelson, C. Stollar Peters, G. Redwine and D. Reside (2009) Approaches to managing and collecting born-digital literary materials for scholarly use, White Paper to the NEH Office of Digital Humanities, May 2009; M. G. Kirschenbaum, E. L. Farr, K. M. Kraus, N. Nelson, C. Stollar Peters, G. Redwine and D. Reside (2009) Digital materiality: preserving access to computers as complete environments, iPRES 2009 Conference, The Sixth International Conference on Preservation of Digital Objects, Stanford University, California, preprint

[149] See also the announcement, 11 January 2010, by Emory University concerning the opening of the Salman Rushdie Archive complete with his digital life and several computers, on 26 February 2010, http://web.library.emory.edu/node/358. As Naomi Nelson puts it: "The emulated environment is as close as we can get to recreating Rushdie's desktop for the researcher". "Instead of isolated files on floppies, we have the entire context in which he worked. We can see how he used technology and the growing Web as part of his creative process"; and a further announcement with comments by various members of Emory Library including its director for born-digital initiatives, Erika Farr: "As an academic and researcher herself, Farr is a big believer in preserving the whole ecosytem, or 'biostructure,' of the author's digital archive: the hardware, software, programs, and applications, all the files and file names, search histories - even the order in which everything was installed. 'There is something fundamentally interesting about the computers themselves,' she says, 'as the medium between the user and the digital media.'", http://www.emory.edu/EMORY_MAGAZINE/2010/winter/authors.html

and literary - at the British Library in recent years[150].

With the cloud computing becoming more commonplace, will "access to the original storage media" become a thing of the past? It is indeterminable but even besides the possibility of invited and permitted remote acquisition over the network, the era of personal digital materiality might not be a transient one. It is possible that people will still want to have at least a copy of personal information held locally (as the Digital Lives project advocates); and with the crucial role played by the hand, information may continue to exist in handheld and wearable devices. Time will tell.

On the other hand, materiality alone clearly does not suffice, as Kirschenbaum and colleagues make plain. The context, the holistic whole, is essential. Thus as Christine Finn observes in conversation with Sellam Ismail, the founder of the Vintage Computer Festival, a collector who has amassed more than a thousand computers: "He reminds me that hardware is just the shell of the vehicle, and without the software it is lifeless. It provides Sellam with a reminder of the experiences he had with it. 'But still having the software allows me to make the experience alive and real again. It is the soul of the machine. It almost sounds like a person, doesn't it?" "'...having the program listings and being able to replicate the work that I did sixteen years ago really brings back the experience, more so than just the hardware itself'".

The physical object does convey information, of course, even grandeur, but in isolation there is also loss and regret, as observed in the quotation from Reverend Francis Kilvert's diary, 29 March 1871, at the beginning of this synthesis[151].

In the contexts of an electronic computer system, the concerns of physicality, materiality blend into the experience of using the computer. The longterm preservation of digital technology is expensive and typically impractical, and on occasion it is probably impossible even for specialist computer museums; and so accurate emulation is fundamental for getting as close as possible to the original experience. The evocative title of a paper on preserving the 'look and feel' of ancestral digital objects and systems emphasises again the centrality of the original behaviour: "'The Old Version Flickers More': Digital Preservation from the User's Perspective"[152].

8.5 Locations: Virtual versus Real

(1) So there is a vast source of fascinating and useful personal data; but there are also challenges – technical, ethical and legal, and motivational. Three specific ones are to ensure (i) that personal digital archives are created, retained and preserved, (ii) that their contents do not lose value due to uncertainty of authenticity and provenance, and (iii) that personal digital objects can be subject to use and reuse without impinging on privacy.

---

[150] eg J. L. John (2005) Because topics often fade. Letters, essays, notes, digital manuscripts and other unpublished works, in Narrow Roads of Gene Land. The Collected Papers of W. D. Hamilton, Volume 3, Last Words, edited by Mark Ridley, pp 399-422; see also J. Andrews (2009) Save, get, delete, Times Literary Supplement, 13 March 2009, p 15; J. Andrews (with J. L. John) (2009) Save the ephemeral, The Author. Journal of the Society of Authors, Summer 2009, pp 70-71; J. Andrews (2010) 'Laid aside'? Collecting contemporary literary archives and manuscripts, Archives (British Records Association), in press

[151] W. Plome, editor, (1977) Kilvert's diary 1870-1879. Selections from the diary of the Rev. Francis Kilvert. Chosen, edited and introduced by William Plomer, Penguin Books, Harmondsworth, Middlesex, p 22

[152] M. Hedstrom, C. A. Lee, J. S. Olson and C. A. Lampe (2006) 'The old version flickers more': digital preservation from the user's perspective, The American Archivist 69: 159-187

(2) There are essentially two principal *physical* locations for personal digital content, for an individual's own content and personal archives: (i) the vicinity of individuals and their families; and (ii) more distant institutions (commercial or otherwise; online or otherwise).

(3) Institutions in turn may be either (i) not-for-profit repositories (public and nongovernmental organisations); or (ii) commercial organisations such as most online service providers. Historically, it has proved difficult for commercial organisations to combine longterm security of large diverse collections with profit-making viability. It is sometimes possible with small, highly valuable and desirable collections.

(4) Commercial entities sometimes become not-for-profit organisations or at least the collections join a not-for-profit organisation; and many of the world's finest and most comprehensive collections originate from profit-making philanthropic benefactors.

(5) One issue is the future relationship between the origin of the personal digital objects (potentially people all over the world including the UK and Europe) and their ultimate geo-legal destination (conceivably a few institutions in one, two or a few countries). Consider for example the location of many of today's most successful online service providers.

(6) Museums and libraries today are faced with requests for the return of cultural objects to their local origin. Perhaps even commercial organisations might want to avoid future demands of this kind on their time and resources (perhaps involving restoration of digital rights and monetary compensation), and work sooner rather than later with local institutions, and find favour with people for respecting the individual origins and cultural heritage of personal objects and content.

(7) Closely related to this matter, of course, is the issue of personal information pertaining to an individual living and originating in one country (of many countries across the world) being subject to the laws and ethics of another country, typically one or a few other countries.

(8) It is clear that the issue of privacy and the holding of personal information by globally commercial organisations such as Google, Microsoft and Facebook, and the potential power that this grants a commercial organisation that controls such information, continues to attract the close attention of news media and the press as well as of specialist academics and rights organisations[153].

(9) Recent events with online service providers show that it still remains to be seen that they can or even want to offer longterm sustainability of personal digital objects. Many online service providers categorically insist (i) that they will not take responsibility for short term preservation let alone longterm preservation, (ii) that individuals must backup and look after their eMSS themselves, and (iii) that individuals retain key digital rights to the objects as well as responsibilities.

(10) On the other hand, some appear to covet the rights to the user created content. Facebook, for example[154], has been criticised for making it difficult for users to export much

[153] Magazines and newspapers that have recently highlighted this issue in the context of Facebook and Google include Wired and the Guardian newspaper: see Vogelstein (2009); Anonymous (2009) Medical privacy. Dr Google will see you now, Guardian newspaper, leader, p 26, 28 July 2009

[154] See Vogelstein (2009)

of their personal content (including contacts, email, images and movies), and moreover recently changed the terms of service in such a way that it seemed to many users to be giving itself permanent ownership of everything uploaded to its social networking site; following the complaints of many users the terms of service were amended.

(11) Nonetheless, whatever ultimately transpires it is plain for all to see that online service providers are highly influential operators in the digital universe, and are in numerous ways showing the way forward, and most especially are revealing some of the inherent value of personal digital content. Equally, not-for-profit organisations and movements have explored and pioneered new paths with the Creative Commons, Internet Archive, and open access (see §4).

(12) There are potential synergies to be explored in the requirements and aims of online service providers and those of public repository institutions.

8.6 Trusted Repositories as Trusted Mediators

(1) The first step would be to raise awareness of the value of personal archives and to help to motivate the keeping of archives in the first place.

(2) There are two fundamental ways that a repository can promote the security of an archive once it exists: (i) by taking the archive in and looking after it directly; or (ii) by facilitating its sustainability indirectly through advice and services.

(3) One approach would be to offer: (i) a bespoke system for the most influential and creatively productive individuals including scientists, writers, political reformers, artists, environmentalists, architects, engineers and so on; and (ii) services and advice for people who are sustaining 'archives in the wild', with a repository supplying crucial authenticating facilities, digital preservation services, archival tools, and occasional or emergency intervention.

(4) Regarding research access to 'archives in the wild'[155], trusted repositories (public or nongovernmental or even commercial) would make it possible for registered individuals to make their archives (or elements of them) available to *bona fide* researchers *via* the internet just as volunteers already leave their personal computers online in order to allow research programmes (eg MalariaControl.net and Climateprediction.net)[156] to harness the processing power of these computers remotely.

(5) The fact that individuals would in this scenario hold the information themselves, and be in

[155] The vast and rich information of the publicly available world wide web will in the years, decades and centuries ahead be potentially an enormous treasure of information for historical, cultural and scientific endeavour (if it is successfully captured in the first place). Yet, while some individuals are placing highly personal material on public websites, most people do not put anywhere near all aspects of their lives on the public web, partly due to a sense of modesty or privacy, partly because interest is perceived to be limited, and partly because people tend to present themselves in a moderately good light if not the best light (a mild instance is found in the simple fact that people are inclined to put forward their best photos, taken by them or of them, but retain in their personal collection less good photos that are nonetheless of some value to them); and it is for these kinds of reasons that the personal archive can be expected to remain a fundamental, incomparable, research resource. It is worth bearing in mind too that many websites start and continue life in draft form on personal computers

[156] The most famous example is SETI, Search for Extra Terrestrial Intelligence, a serious scientific programme that monitors radio signals emanating from space: SETI@home

ultimate control of their own archive, would mean that people would be free to choose to make it available to researchers. The extent to which people do so will vary but nonetheless many might well want to contribute to research in this way - research that could advance scientific solutions for many pressing medical, social, environmental and cultural concerns.

(6) This touches on the future value of the content - monetary, scholarly, familial, individual - and the changing nature of the value through time. Monetary value will depend on the technical, economic, cultural and legal circumstances but in any case the potential value of the information to research in science and humanities will be demonstrably very high.

(7) Trusted repositories could mediate appropriate access for registered researchers while helping to protect privacy on behalf of the individual and helping to ensure authenticity on behalf of the researcher. This would clearly require the development of rigorous, accountable and transparent mechanisms and agreements.

(8) In some cases tools for anonymising the personal information might be applied by the repository prior to access by researchers, or access could be restricted through selective encryption, for example, to sections of an individual's archive.

(9) To further authenticity, repositories could make it possible for people to submit files for hashing, with the repository retaining the hash values but not the files themselves; many years later a file's longstanding existence and integrity could be vouched for by the repository based on the hash value and other supporting provenance data. (There are many possibilities for additional provenance securing information; for example, the music database company Gracenote uses start and end times for tracks measured in minute fractions of a second as unique identifiers, since CDs do not have 'exactly the same sequence of track lengths[157]; in short it uses existing information within the object[158].)

(10) There might also be a registry for individuals to notify a repository – if they choose - of any pseudonyms and other alternative identities (eg avatars) that he or she has adopted.

8.7 A Role for Personal Archives in Personalised Usability?

(1) The concept of product usability can be applied to the universal behaviours of people generally, to the archetypal behaviours of types or statistical classes of people, or to the personalised behaviours specific to individuals. A device might be designed so that it can suit either right handed or left handed people or both classes of people. Among the most advanced forms of usability are those that attempt to specifically match and adapt to the preferences, past activities and idiosyncrasies of the individual.

(2) Research (including the present project) has shown much variation in the way individuals obtain, manage and use their information. This realisation has led to software that incorporates a usability that enables the identification and application of individual preferences based on past behaviours and activities. In some cases it makes use of the preferences of like-minded individuals (and thus mirrors some of the processes underlying

[157] E. Dyson (2003) Online registries: the DNS and beyond... Release 1.0. Esther Dyson's Monthly Report 21, p 29

[158] Certain significant properties of the technical nature of files and media might be employed too; this is picked up again in §10

social networking sites)[159].

(3) Looking into the future, it is a highly salient point that a copious source of personal information for developing 'personalised usability' is an individual's personal archive. A dynamically accessible personal digital archive might be seen as the ultimate resource for identifying suitable personal characteristics and preferences that matches a device's (or a service's) function to an individual in order to optimise its effective use.

(4) This has three conceivable implications: (i) the personal digital archive supports usability which further induces the existence and sustenance of personal digital archives; (ii) indeed for this reason dynamically available personal information may well become essential for modern life; and (iii) personalised usability might also in turn mean that online service providers would be required to allow individuals access to all (and ultimate control) of their own personal content.

(5) It seems likely that even those individuals who agree to a repository storing their digital archives will want to keep copies at hand too, and might need to do so anyway because of personalisation requirements. In short the repository will add an important layer of security and authenticity but will operate in parallel with the existence of material outside the repository. It is difficult to anticipate with any success let alone predict accurately the longterm future but an interesting question concerns the fate of these parallel digital objects outside the repository[160]. The prospects of those personal digital objects held by archival repositories that have a public longterm remit seems more assured, and these objects may end up helping to authenticate any that survive outside the repository.

8.8 iCuration: from I to i

(1) Web technologies have, of course, made it increasingly feasible for members of the public to access the content of repositories remotely. Much less attention has been given to how curatorship and archiving can be conducted remotely.

(2) Traditionally, curators and archivists have undertaken many of their activities offline, in person, travelling across the country in order, for example: (i) to examine archives of interest, (ii) to collect donated manuscripts, (iii) to give lectures, (iv) to provide training, (v) to record an interview with an eminent writer or scientist, and (vi) to provide expert opinion regarding authenticity and provenance; or alternatively members of the public have been expected to travel to the curator's institution for consultation.

(3) Clearly, the other side of adopting web technologies is for curators and archivists to undertake more and more of their work online. An online system is required in order to help members of the public in creating and maintaining personal digital archives of their own.

---

[159] V. Bellotti, B. Begole, E. H. Chi, N. Ducheneaut, J. Fang, E. Isaacs, T. King, M. W. Newman, K. Partridge, B. Price, P. Rasmussen, M. Roberts, D. J. Schiano and A. Walendowski (2008) Activity-based serendipitous recommendations with the Magitti mobile leisure guide, Conference on Human Factors in Computing Systems (CHI 2008), Florence, Italy, pp 1157-1166; Y. Takeuchi and M. Sugimoto (2009) A user-adaptive city guide system with an unobtrusive navigation interface, Personal Ubiquitous Computing 13:119-132

[160] For an evolutionary perspective see §9.5; J. L. John (2009) The future of saving the past, Nature 459: 775-776, 11 June 2009

(4) An essential approach therefore is that of iCURATION: curation over the network.

(5) This form of online curation would embrace (*inter alia*) the networking of (i) archival institutions and repositories and their content and services (international, national, regional, local and expert or 'high profile' individual), as well as (ii) personal archives everywhere, that is the curation of archives in the wild, supporting authenticity, privacy protection, and the provision of mediated access.

(6) Core elements of iCURATION include: (i) motivating and researching sustainable personal archives, (ii) providing online advice, training and services, (iii) enabling remote capture and reception of personal digital objects and content, (iv) encouraging online participation and engagement, (v) making available suitable tools and research resources, (vi) sharing content in accordance with wishes of individuals, and (vii) making access to content possible adapted in accordance with the preferences of researchers subject to the legal and ethical environment.

(7) Unpacking the concept of iCURATION a little further, it could embrace cloud computing, without being restricted to it, with for instance longterm digital storage being offered by quality repositories (as a costed service if nothing else). It would also encompass digital preservation services of the kind devised by the Planets programme: remote emulation, characterisation, testbed experimentation, planning, migration and so on. Authenticity and utility could be supported not only through hash values and provenance data but through the maintenance of personal digital object registries, the use of digital watermarking, and digital object identifiers and handlers, and also the permitted tracking of the way digital objects are used; forensic analysis of digital objects such as digital documents could be conducted to assess likelihood of purported origins. Curatorial services would include guidelines, tools and services about how to manage personal information for longterm usability, how to anticipate future value, how to research personal archives through mediated access, what research qualities and quantities of information exist in accessible portions of personal archives, and so on[161].

[161] Some of these notions are elaborated further in §9 and §10

Chapter 9: Enabling Archives (for Everyone) with Future Research[162]

9.1 Perspectives for the Future

(1) This chapter aims to outline key areas of research in furthering the curation of personal digital archives generally, and iCURATION specifically. The emphasis is placed on helping every individual to create, maintain and sustain a personal digital archive, and in furthering digital literacy and inclusion, and the role of repositories in these activities.

(2) The chapter explores eight topics appertaining to a field of personal informatics: (i) authentication of personal information; (ii) an archivally oriented personal information management (archival PIM) shared by creators and curators; (iii) a functionality, based on personalised and archetypal usability, that is designed to benefit users and curators as well as creators; (iv) an evolutionary and phylogenetic approach to the analysis of the dynamics of personal archives and of historical events; (v) the central role of infoethics, personal rights and digital value in the future economies of personal objects; (vi) the advancement of automated and supervised cataloguing and personally created metadata and added contextual information; (vii) the requirement for new research techniques and access in the future, from complex visualisation through to new forms of numerical and statistical analysis; and (viii) the design and adoption of adaptive systems and technologies for curation.

(3) Suffusing all of these perspectives is an appreciation of both the variation and similarities among individuals in their sociocultural, ecological and behavioural relationships with information, in the ways they search, organise and manipulate information, in their sense of privacy and openness, in the financial resources and time available to them, in the extents of their desire to leave an archive for future generations, and in the quantities and qualities of objects which they feel inclined to retain.

9.2 Authenticity and Forensics: from Hand to Network

(1) The present study has advocated and demonstrated the use of forensic techniques in meeting the need for authenticity.

(2) Three key aspects are the abilities (i) to interpret the dates and times in personal digital objects (sometimes with considerable effort), (ii) to forensically capture the disks and files in a sound way, and (iii) to extract metadata and earlier versions as reliably as possible.

(3) A pressing requirement is to establish the most suitable ways of informing, and securing the permission, of originators and families with various options being offered, and to explore ways of ensuring that such forensic processing is monitored and audited in ethically and curatorially appropriate ways.

(4) Establishing and adapting suitable forensic procedures for identifying the correct sequences of draft versions, and doing so in a scholarly context, and for diverse word processing and other creative applications, requires careful research and documentation.

(5) In some circumstances, it is possible - albeit usually with significant effort - to identify patterns of editing and styles of composition. Clearly this kind of research would be

[162] This chapter was prepared by Jeremy Leighton John

potentially very interesting to the study of creativity and working practices.

(6) The use of handheld devices and gesture interactions brings new requirements and it is possible to anticipate some convergence with biometrics.

(7) The identification of digital fakes and forgeries will be an important topic for future archivists.

(8) Historians and other researchers including scientists will need to be aware of the possibilities of digital deception in the legacies of digital lives.

Advanced forensic format[163]

In addition to being a standard format for storing hard drive images, AFF is associated with a software package (AFFLIB) and several tools including an Advanced Disk Imager (aimage), a program that produces AFF metadata in XML (afxml), and a program that converts raw 'dd' images into AFF images and *vice versa* (afconvert).

The widely used 'dd' program is limited in its ability to read bad sectors, in being prone to accidental misuse, and in its failure to capture metadata such as the serial number of a hard drive, time of imaging and name of person doing the imaging. The new aimage program can create a disk image as a raw 'dd' file or as an AFF file or both at the same time, and it incorporates error and bad disk handling procedures, including the ability to go to the end of the disk and try imaging the disk sectors in reverse[164].

AFF is an open, multiple platform, extensible format that is designed to be able to store: (i) disk images with or without compression; (ii) disk images of any size; (iii) metadata with in disk images or separately; (iv) disk images in a single file or as multiple split files (the popular but proprietary EnCase file format, for example, is limited to 2 GB and splits the files); and (v) arbitrary metadata defined by the user as name-value pairs. Significantly, it incorporates the ability to authentic the files with advanced digital signatures certificates along with hash functions (this helps to address a concern raised by the Digital Lives user focus group). Nearly all of these flexible features potentially serve the purposes of the digital curation and preservation community. The compression is effective (one test showed AFF to perform notably better than EnCase) and the compressed versions of the AFF disk image are seekable[165].

---

[163] S. L. Garfinkel (2006) AFF: a new format for storing hard drive images, Communications of the ACM 49(2): 85-87; S. L. Garfinkel, D. J. Malan, K-A. Dubec, C. C. Stevens and C. Pham (2006) Disk imaging with the advanced forensic format, library and tools, in Research Advances in Digital Forensics, Second Annual IFIPWG 11.9 International Conference on Digital Forensics, Springer; S. L. Garfinkel (2009) Providing cryptographic security and evidentiary chain-of-custody with the advanced forensic format, library, and tools, International Journal of Digital Crime and Forensics 1:1-28; M. Cohen, S. Garfinkel and B. Schatz (2009) Extending the advanced forensic format to accommodate multiple data sources, logical evidence, arbitrary information and forensic workflow, Digital Investigation 6: S57-S68

[164] http://www.garloff.de/kurt/linux/ddrescue

[165] See also M. Cohen (2008) Pyflag: an advanced network forensic framework, Proceedings of the 2008 Digital Forensics Research Workshop (DFRWS 2008), http://www.pyflag.org. Pyflag uses a gzip format (specifically sgzip, a seekable variant of gzip) to store a compressed disk image that is seekable, allowing random access (ie direct access) to information within the compressed file. The significance of this functionality of Pyflag and AFF is that the computer does not need to decompress the compressed captured information in order to access the sought information; thus it can behave as if it is accessing information on a physical disk through the file system

By means of a program called affuse, AFF integrates with the open source program FUSE (Filesystem in Usespace)[166] which means that a compressed AFF file can be seen as a raw file in a computer's own file system, a procedure that can be combined with the use of virtualisation software such as VMware. It is also possible to manage cryptographically the disk image: for example providing different forms or levels of access to specific parts of the acquired disk image, which is useful for controlling access according to privacy policies and requests. These aspects of AFF have great relevance to the provision of disk images to scholars in providing the writing and research digital environment of the creator.

(9) One general activity would be the selective integration of forensic approaches to the curation of digital media and objects with digital preservation strategies, and an exploration of these techniques operating together.

(10) It will be beneficial to continue to monitor these technologies and perspectives. Forensic techniques are continually advancing. One effect of this change is that some forensic techniques and understanding become less pertinent to forensic scientists in the contemporary environment. But this forensic understanding of earlier software and hardware and the effects that these have on digital objects represent an invaluable resource that will - if secured for future posterity - be of very great benefit for future historians and archivists.

9.3 Archival PIM

(1) Personal information management from the perspective of personal digital archives is a surprisingly under researched area[167].

(2) The project has proposed a model that adopts the lifecycle approach of professional archivists working with personal archives and transfers it to the context of personal information management by individuals. Significant research is necessary to bring about this whole life approach to the curation of personal objects based on OAIS (or an advanced form of it), and in integrating at a basic level personal curation by individuals with the processes of repositories.

(3) The archival perspective looks at an individual's entire life and even beyond his or her lifetime, thereby providing the original creator, with the opportunity to form a lasting personal archive. A central role is envisaged for an archival PIM, a personal information management that is directed not only at the use of current or nearly current personal digital objects but also with future retrieval, use and reuse[168].

(4) Another requirement is for a better understanding and exploration of the way files can be

[166] http://fuse.sourceforge.net, Filesystem in Usespace

[167] A notable exception to the omission is the research of Cathy Marshall: C. C. Marshall (2008) Rethinking personal digital archiving, part 1. Four challenges from the field. D-Lib Magazine 14; C. C. Marshall (2008) Rethinking personal digital archiving[,] part 2. Implications for services, applications, and institutions. D-Lib Magazine 14; C. C. Marshall, S. Bly and F. Brun-Cottan (2006) The long term fate of our digital belongings: toward a service model for personal archives, IS&T Archiving 2006. Ottawa, Canada: Society for Imaging Science and Technology, p 25-30; other researchers working in this field include E. Churchill and J. Ubois (2008) Designing for digital archives, Interactions:10-13

[168] P. Williams, J. L. John and I. Rowland[s] (2009) The personal curation of digital objects: a lifecycle approach, Aslib Proceedings. New Information Perspectives 61:340-363

used to the individual's benefit in the future: such knowledge could perhaps be incorporated within archival PIM systems. Lifetracking is in a sense an exploration of future benefits from personally archived information[169]. Techniques can be developed for enabling better access and use of his or her own archive, and for extending the 'life tracking' ethos with its provision of detailed feedback and temporal comparison of an individual's activities (for those who want it).

(5) A key necessity is for individuals to have control over what they keep and what they discard, and what they are prepared to make available for research, and for what types of research (eg perhaps scientific but not biographical). In addition to technical capability there is a need for a psychological understanding of what influences retention and disposal. One option that has been raised as a way of preventing the dissemination of personal and sensitive information recently is that of automatic deletion. People could set their computer devices to delete files according to certain rules based on the passage of time and *a priori* choices indicated to the software[170]. There are of course risks to this strategy in its most unrefined form, most notably in terms of the quandary of predicting future desirability. Other options may well be feasible. Nonetheless it emphasises that people need to be given control: otherwise many people may simply delete files wholesale with less regard to possible longterm benefits. It should be possible to devise archivally-sound ways for individuals to manage their personal files without undue effort and while feeling comfortable about their holdings. (This issue is explored further in the following Box.)

(6) A specific aim therefore is to reinforce information so that it is interpretable, authentic and significantly useful even (perhaps especially) after long periods, and even to future generations.

(7) Some people may be quite accepting of researchers looking for important relationships between life history and, say, medical conditions; but, nonetheless, not wish to be unduly exposed to some past memories, or wish for their most personal feelings to be made identifiable in public within 30 years of their demise or ever. Mechanisms might be devised for allowing information to be retained and made available for *bona fide* research securely without invoking and imposing repeated unwanted memories.

(8) The research into PIM of today will be an invaluable resource for future historians and scholars wishing to understand how computers were used and how personal information was organised and managed.

Box: Psychological consequences and information hazard

Many of the concerns relating to sustained memory, or digital persistence of information, relate to social and political implications: the potential for example for institutions to examine an individual's past in detail, and no doubt socio-technical-legal devices and structures for ensuring the right balance will be sought. Some commentators have noted, however, that there are psychological concerns arising too, from either too much, incomplete

---

[169] See also G. Bell and J. Gemmell (2009) Total recall: how the e-memory revolution will change everything. Dutton, Boston

[170] V. Mayer-Schönberger (2009) Delete. The virtue of forgetting in the digital age, Princeton University Press, Princeton

or overly persistent information. In a thoughtful account Viktor Mayer-Schönberger[171] brings to the fore an important point, the view that the way "humans manage memory and forgetting is not something that can be changed easily and at will but is deeply rooted"[172], and the idea that "how humans deal with time (especially through forgetting) is the result of a remarkably effective and useful adaptation over a long period of time to the contexts and environments they live"[173].

Mayer-Schönberger strongly promotes the use of deletion, in the hope of avoiding the risk of harm due to a technology that ostensibly runs counter to deep rooted psychological adaptations. It is certainly true that these kinds of concerns warrant careful consideration, research and policy. On the other hand, the ability to use technology to go beyond existing human capabilities is also a longstanding, even ancient, phenomenon. Technologies almost by definition impose themselves, disturbing the existing relationship between humanity and its environment, and inducing yet more change, adjustments and remedies that take many forms: social, legal, technical, emotional.

The solution seems to lie in finding the balance, usually a new balance. After all, the psychological artificiality of modern technologies is sometimes used to good advantage. Virtual reality therapy is being used to give burn patients relief from pain, and to offer victims of extreme phobias ways to overcome their fears[174]. Recurring unwanted memories certainly pose a risk; but recent early research suggests that novel ways of ameliorating them continue to emerge, eg by eliminating fear reactions to traumatic reminders as result of a better scientific understanding of how memory operates[175]. It is not only real memories that can be harmful; so can false ones, as in a case where an old soldier became convinced that he had caused the death of a comrade, whereas archival resources demonstrated that he could not have been responsible, to the heartfelt relief of the elderly subject[176]; any people who have inadvertently contributed to a friend's death are unlikely to forget and might want or need therapy anyway.

Deletion, culling or at least deaccessioning is normal practice for archives, and has been incorporated in research into archiving of digital belongings[177]. The discarding of digital

---

[171] V. Mayer-Schönberger (2009) Delete. The virtue of forgetting in the digital age. Princeton University Press, Princeton; V. Mayer-Schönberger (2009) No thanks for the memory, The Independent newspaper, London, 1 October 2009

[172] V. Mayer-Schönberger (2009) citing, in Delete: D. L. Schacter (2001) How the mind forgets and remembers. The seven sins of memory, Houghton Mifflin, Boston

[173] V. Mayer-Schönberger (2009) citing, in Delete: J. R. Anderson and L. J. Schooler (1991) Reflections of the environment in memory, Psychological Science 2: 396-408

[174] H. G. Hoffman (2004) Virtual-reality therapy, Scientific American: 58-65; see also J. Difede and H. G. Hoffman (2002) Virtual reality exposure therapy for World Trade Center post-traumatic stress disorder: a case report, CyberPsychology & Behavior 5: 529-535

[175] G. J. Quirk and M. R. Milad (2010) Editing out fear, Nature 463 :36-37

[176] J. Etherton (2006) The role of archives in the perception of self, Journal of the Society of Archivists 27: 227-246

[177] C. C. Marshall (2008) Rethinking personal digital archiving, part 1. Four challenges from the field. D-Lib Magazine 14; C. C. Marshall (2008) Rethinking personal digital archiving[,] part 2. Implications for services, applications, and institutions. D-Lib Magazine 14; C. C. Marshall, S. Bly and F. Brun-Cottan (2006) The long term fate of our digital belongings: toward a service model for personal archives, IS&T Archiving 2006, Society for Imaging Science and Technology, Ottawa, pp 25-30

objects is likewise anticipated in the lifecycle approach to archival personal information management[178]. Automated deletion with each digital object having an expiry date (which might be subject to the wishes of the creator, for instance, in the case of a personal digital object) is favoured by Mayer-Schönberger. This is one option, one useful tool. But there will be others too, other ways of coping with the artificial nature of memory technology.

There can be no doubt that it is important to research and understand the human psyche and the effects that online and digital technologies have on it[179]. A simple instance is provided by the 'dark side' of emails, some properties of which reportedly can fuel incivility[180].

The evolutionary anthropologist Robin Dunbar has noted that the "capacity to discover novel facts about the environment has a very ancient basis", and this has been offered as an explanation for the modern, sometimes excessive, craving for information in the information rich age[181]. Attention has been drawn (naturally) towards a variety of 'information hazards', a term coined by Nick Bostrom for dangers posed by information[182].

Mayer-Schönberger clearly understands the desire in many to leave some legacy of their lives behind them, as this passage shows: "By using digital memory, our thoughts, emotions, and experiences may not be lost once we pass away but remain to be used by posterity. Through them we live on, and escape being forgotten. As fertility rates plummet in modern societies, our desire to preserve the memory of our lives outside of the traditional context of intergenerational sharing may get stronger. It is a very human strategy to ensure that we haven't lived in vain, that we aren't forgotten after our deaths as if we've never lived." The value and necessity of personal and family memory has been elaborated by a number of practitioners[183]. It seems likely that for such memories to be valuable and valued they need to be grounded in some level of reality. Misleadingly selective documented memories might fail ultimately to provide the desired benefits. A real understanding is required of how and why individuals and institutions use information to serve their ends, how information is selected and manipulated. Neither retention nor deletion are necessarily entirely neutral acts, and neither of them is dispassionate in its effects on veracity. Equally, individuals may wish, quite naturally, to keep their private lives to themselves while making available a selection in a way that is not misleading: eg in providing, say, drafts of a novel for literary scholarship, while not making available a series of love letters that remain close to the heart.

---

[178] P. Williams, J. L. John and I. Rowland[s] (2009) The personal curation of digital objects: a lifecycle approach, Aslib Proceedings. New Information Perspectives 61: 340-363

[179] eg A. Barak (2008) Psychological aspects of cyberspace. Theory, Research, Applications, edited book, Cambridge University Press, Cambridge

[180] V. K. G. Lim, T. S. H. Teo and J. Y. Chin (2008) Bosses and their e-manners, Communications of the ACM 51: 155-157

[181] P. Parsons (2010) Ignorance is bliss, New Scientist 205: 38-39

[182] N. Bostrom (2009) Information hazards: A typology of potential harms from knowledge, Draft 1.11, Future of Humanity Institute, University of Oxford, Oxford, pp 1-35, http//www.nickbostrom.com/; P. Parsons (2010) Ignorance is bliss, New Scientist 205: 38-39

[183] eg J. Etherton (2006) The role of archives in the perception of self, Journal of the Society of Archivists 27: 227-246

9.4 Personalised and Archetypal Usability

(1) Some initial analyses (by the present project although details are not reported here) have provisionally identified some archetypes (or classes) of personal information management that are also relevant in this context. The same individuals can show behaviours that they share with others in the same archetypal class as well as behaviours that are highly individualistic. Both types of behaviour are potentially important and useful in developing tools and approaches that enhance the organising, finding and storing of personal information by people.

(2) Equally, individuals may use different functionalities at different times. A diversity of approaches is already widely seen in much computer software with various routes to finding or acting being offered: from menus, tool bars, mouse right clicking and so on. These various routes are quite deliberately retained by software designers.

(3) An archivally-based system for personal information management needs to embrace these personal and archetypal variations. How would this be brought about in practice?

(4) The same goal of ensuring usability through individual or archetypal differences can be extended to users of archives, with researchers being offered various means of conducting research and obtaining access. This would require research in the archive and library context.

(5) Moreover, institutions might consider following suit, by allowing their own people to work in diverse ways. Thus the application of personalised or archetypal usability principles to curators and archivists as well as to users of repository institutions and to creators might be explored.

(6) A notable tool in personalising devices and tools involves the production of mild variation (through judicious randomisation), allowing the user of the product to select the most suitable version from the resulting variants. The selected variant is subject to the same process, with iterative randomisation and selection, yielding hopefully a product that matches more closely the tastes and preferences of the individual. This is in essence a version of the technique of evolutionary usability.

9.5 Evolutionary Dynamics and Phylogenetics of Archives

Complex systems

(1) A fundamental change arising from the digital nature of personal archives is manifested in their passing from one generation to the next. Self-archiving by families will likely result in a child receiving (perhaps typically in adulthood) two sets of digital manuscripts, one from each parent[184].

(2) In the case of eMSS, in contrast to a paper letter, all siblings can inherit exact digital replicates of the file.

---

[184] J. L. John (2009) The future of saving the past, Nature 459: 775-776; J. L. John (2008) Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives and tools, iPRES 2008 Conference, The Fifth International Conference on Preservation of Digital Objects, The British Library, London.

(3) On the other hand, siblings can be expected to differ in what they choose to keep, and in turn their descendants (biological, intellectual or even institutional) will not necessarily retain all of the eMSS received. Over a number of generations some digital manuscripts would be lost while others would be favoured, and some would be modified either inadvertently or deliberately.

(4) After many generations, there will be highly diverse digital archives, and identical replicate eMSS (as well as modified variants) distributed throughout the population of individuals and families (and of institutions and repositories). It is worth remembering in this context that eMSS might include the code for personally created 3D artefacts as well as text and image.

(5) This will be of considerable interest to researchers of complex systems, networks, and ecological and evolutionary change.

(6) Obviously the system will show differences from more familiar evolving systems of organic biology: for one thing people will routinely acquire, create and replicate large numbers of eMSS throughout their own lives, and of course the circumstances of the personal information will reflect the prevailing cultural, economic, legal and ethical environment, which is continuously changing and subject to amendment. It offers a novel and complex system for investigation.

Box: Adaptive and pervasive computer networks and human social networks

A strong insight into the future impact of ubiquitous computing is given by reflecting on its highly distributed nature and in particular the extent to which users will communicate in new ways beyond the conventions of existing network systems. The internet is a distributed system itself with no single centralised ownership or single locus of control, and indeed this is a source of its success and emancipating attractiveness to individuals manifested in today's world wide web. However, conventional networks generally depend on always-on and end-to-end connectivity. Future networks of mobile and ubiquitious computers promise even greater decentralisation, transience and dynamism based on adaptive, temporary and opportunistic connexions. There are significant sociotechnical challenges which are being addressed by projects such as SOCIALNETS and TESS (Developing Theory for Evolving Socio-Cognitive Systems)[185].

The approach seeks to take as its model and even to harness directly the flexible and networking nature of human relationships, "exploiting the core characteristics of human behaviour that lead to highly effective and dynamic interpersonal relationships and social networks". For example, conventional routing protocols such as MANET (for wireless communication between devices) aim to provide an internet-like stable view of the network, but opportunistic protocols accept that the path through the network cannot be completely established before starting the forwarding process and instead allow messages to proceed

---

[185] S. M. Allen, M. Conti, J. Crowcroft, R. Dunbar, P. Liò, J. F. Mendes, R. Molva, A. Passarella, I. Stavrakakis and R. M. Whitaker (2008) Social networking for pervasive adaptation, Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshops, pp 49-54; C. Boldrini, M. Conti and A. Passarella (2008) Exploiting users' social relations to forward data in opportunistic networks: the HiBOp solution, Pervasive and Mobile Computing 4: 633-657; P. Hui, J. Crowcroft and E. Yoneki (2008) Bubble rap: social-based forwarding in delay tolerant networks, 9th ACM International Symposium on Mobile Ad-Hoc Networking and Computing (MobiHoc08)

towards the destination making use of temporary bridges. In short, the message is sent on its way before the route to its destination has been determined. In order to ensure efficiency, context-aware routing based on users' social behaviours is being explored. This entails, for example, making use of personal information about the individual ("information that describes the reality in which the user lives, and the history of social relations among users").

Equally, researchers have been investigating social processes of altruism and selfishness in the forming of cooperative relations and in establishing and enabling trust and security and also network size[186]. Other evolutionary and ecological phenomena that seem potentially relevant are ornamental attraction, signal selection and host-parasite relationships.

Information passing through time: genes, memes, eMSS

(1) It may be possible to analyse the movement and influences of eMSS with their content of ideas, observations of nature and accounts of events as never before. It will be multidisciplinary and interdisciplinary, involving scientific analysis of systems alongside social and humanities scholarship for detailed interpretation[187].

(2) Phylogenetics has been applied already to paper manuscripts as might be expected given its close correspondence with scholarly stemmatics. It is likely that future researchers will be able to create phylogenetic networks or trees from extant personal digital archives, and to determine the likely composition of ancestral personal archives and the ancestral state of the personal digital objects themselves[188].

(3) The approach - that is to say, the focus on personal objects and their content - provides in principle an enriched form of analysis of networks, enabling not only the finding of connexions between people and digital objects but also the examination of the contents of digital objects enabling a more sensitive and fuller interpretation (and microanalysis) of the exposed interrelationships. There are, however, a number of challenges to be examined in practice.

(4) One advantage of analysing the passage of eMSS, or combinations of them, through time is that they are less nebulous than 'ideas' or 'concepts' (and are therefore more readily susceptible to objective analysis), and reside in external media, physically independent of the brain, and so are available for direct examination at various scales.

Theoretical and practical foundations

(1) One sees in the phenomenon a naturally occurring though liberal form of LOCKSS (lots of

---

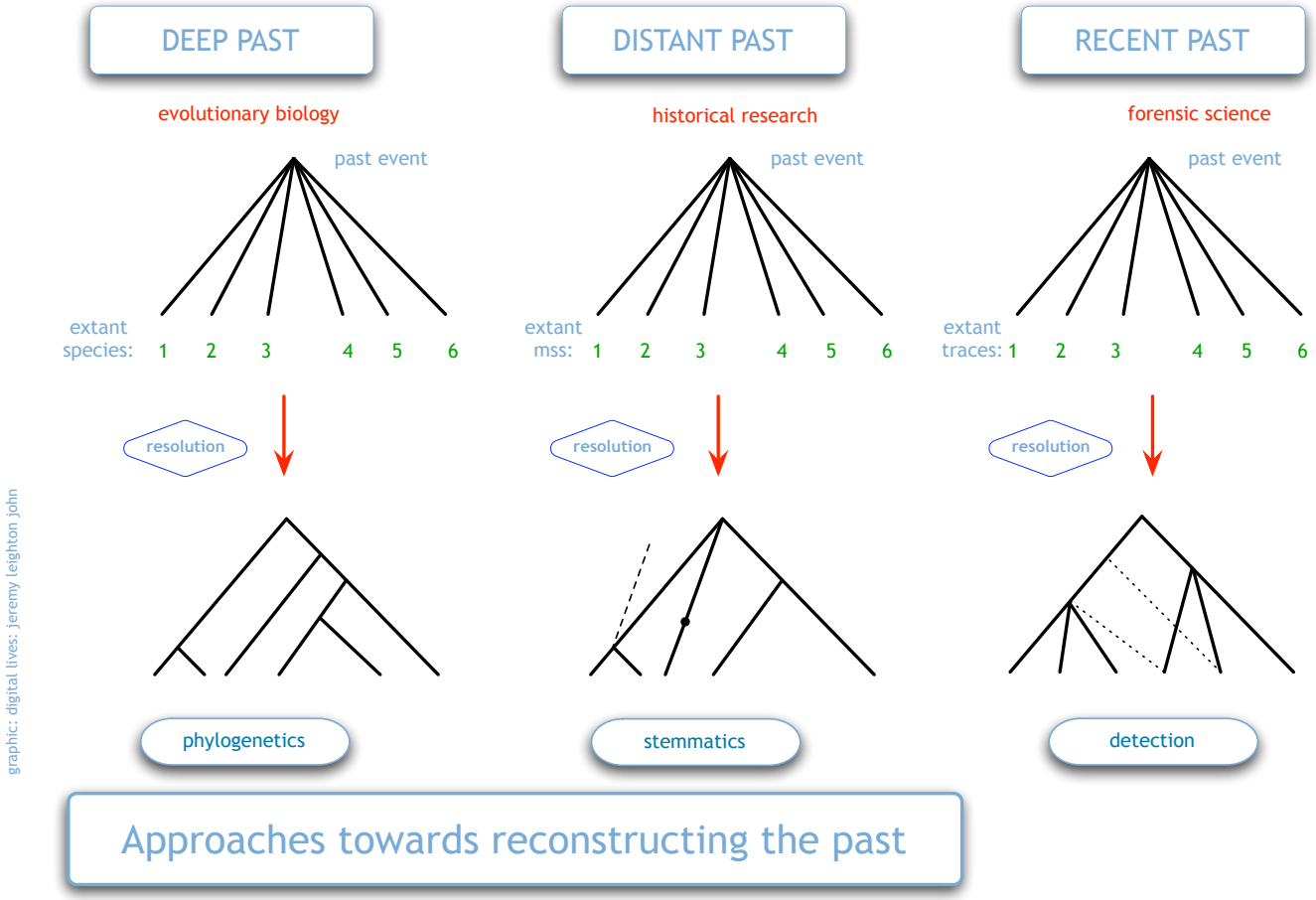
[186] K. Xu, P. Hui, V. O. K. Li, J. Crowcroft, V. Latora and P. Liò (2009) Impact of altruism on opportunistic communications, First IEEE International Conference on Ubiquitious and Future Networks (ICUFN09), Hong Kong, China; Y-E. Lu, S. Roberts, P. Liò, R. Dunbar and J. Crowcroft (2009) Size matters: variation in personal network size, personality and effect on information transmission, IEEE International Conference on Social Computing (Social COM), Vancouver, Canada

[187] J. L. John (2009) The future of saving the past, Nature 459: 775-776

[188] A. C. Barbrook, C. J. Howe, N. Blake and P. Robinson (1998) The phylogeny of The Canterbury Tales, Nature 394: 839-840; M. Spencer, B. Bordalejo, L.-S. Wang, A. C. Barbrook, L. R. Mooney, P. Robinson, T. Warnow and C. J. Howe (2003) Analyzing the order of items in manuscripts of The Canterbury Tales, Computers and the Humanities 37: 97-109; C. J. Howe, A. C. Barbrook, L. R. Mooney and P. Robinson (2004) Parallels between stemmatology and phylogenetics, in Studies in Stemmatology II, editors P. van Reenen, A. den Hollander and M. van Mulken, John Benjamins Publishing, Amsterdam, pp 3-11; M. Spencer, E. A. Davidsion, A. C. Barbrook and C. J. Howe (2004) Phylogenetics of artificial manuscripts, Journal of Theoretical Biology 227: 503-511

copies keep stuff safe), a well known concept within the field of digital preservation[189].

(2) Of particular significance, is the route it provides for analysis of both digital preservation and digital change in the face of an unpredictable future. It offers the prospect of a theoretical foundation for both digital preservation and digital curation as well as furthering the understanding of evolutionary, historical and creative processes.



9.6 Value of Objects, Digital Rights, Infoethics: Key Factors in a Digital Economy

Value

(1) The phenomenon that lies at the heart of personal curation is the inherent value of eMSS. It is intimately dependent on the prevailing social, technical and legal environment.

(2) Predicting future value is difficult. Even the identification of future use is not trivial. Nonetheless, there is a strong perception that some of the information will often be valuable in the future. Some information is retained without justification: just in case, a powerful motivator, and an interesting psychological circumstance worthy of study in its own right.

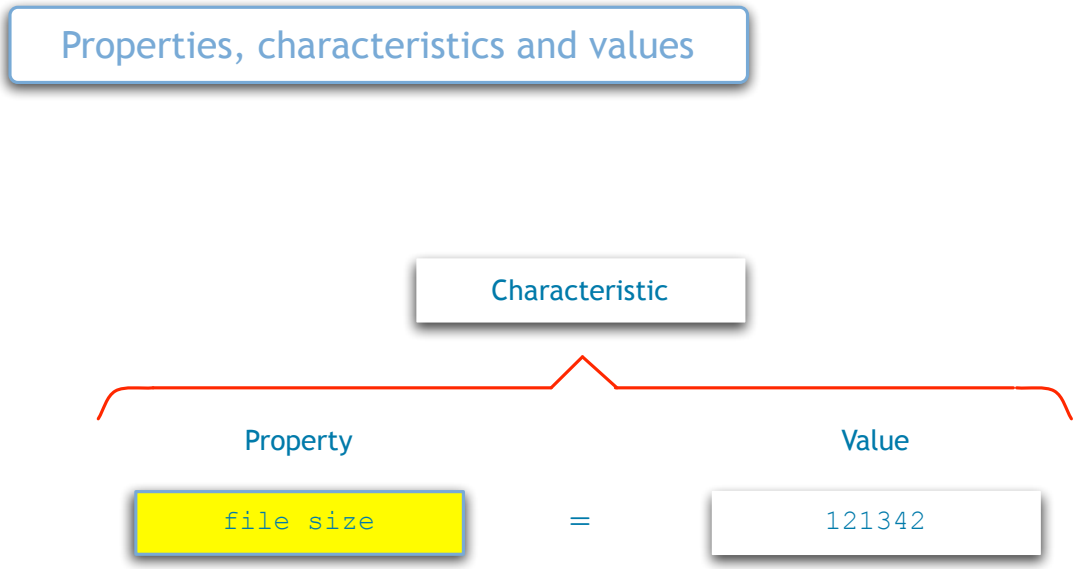
(3) It is possible for personal digital objects to have monetary value although it may take a form that is not usually experienced by conventional manuscripts dealers and auction houses, and more akin to the kinds of value exploited by online service providers in supporting advertising campaigns and marketing activities. Some eMSS may be desirable despite not

[189] http://lockss.stanford.edu/lockss/Home

being unique or rare. Some eMSS may be made more valuable by not being shared widely; or through strictly limiting digital rights.

(4) As now, there may be a distinction between the sale of the (digital) rights to an object and the (digital) object itself.

(5) Closely allied to the topic of 'value' is the digital preservation concept of 'significant properties' which concerns the identification of those characteristics of a digital object and its environment(s) that influence the object's usefulness and the way it is used[190]. It is especially interesting because it is proposed that quantities and qualities represented by the significant characteristics of a digital object and environment should be selected for preservation according to the wishes of its user community, eg commonly the researchers who would like to consult it.



graphic: digital lives: jeremy leighton john

adopted from Dappert and Farquhar (2009)
Significance is in the eye of the stakeholder

Sharing and exchanging

(1) The ability of eMSS to provide indirect benefits to an owner depends on the nature of the economic system.

(2) People who create or otherwise possess desirable eMSS and share them may be rewarded through means other than money: (i) widespread recognition of generosity and creativity; (ii) mutually returned sharing of other eMSS to which access is not universal; (iii) private charitable feelings from contributing to worthwhile research; and so on.

---

[190] A. Dappert and A. Farquhar (2009) Significance is in the eye of the stakeholder, 13th European Conference on Digital Libraries (ECDL 2009), Research and Advanced Technology for Digital Libraries, Corfu, Greece, 27 September - 2 October 2009, M. Agosti, J. L. Borbinha, S. Kapidakis, C. Papatheodorou and G. Tsakonas, editors, LNCS 5714, pp 297-308

(3) The impact of digital rights on future creativity (as well as preservation) may be counterintuitive. Copyright is often justified as a means of rewarding and encouraging creativity, and as a means of encouraging production, dissemination and distribution of creative objects in high volumes. Classically, where enforced, copyright rests with one individual or intermediary organisation, and the replication of the object relies on the intermediary organisation. Due to the nature of the costs, predigital systems tended to concentrate resources on the very most desirable items (the world's, region's or culture's favourite songs, pictures, poems, stories, classical music). In the digitally networked era, replication of useful digital objects can be driven by many individuals, which could promote the object's longterm sustainability as well as its reach; and, as has been widely witnessed, this can happen even without the formal backing of an intermediary organisation. Crucially, it also means that the less universally desirable objects are replicated due to the ability of the internet to reach even small communities of individuals that will value a less universally desirable object.

(4) Is it possible to find a way to reward creators of useful objects without artificially constraining replication? It is necessary to cater not only for the very best digital objects - the ones with the highest artistic quality, financial and popular appeal. It may be beneficial to society to cater for the very good and good and even marginally good or fair, to enable mass replication. Intellectual property right or copyright has not served the creator if it simply controls and limits replication. Even moderately good or specialist objects may be beneficially replicated, deserve to be replicated.

(5) One possibility would be to devise a scheme for registering the creation of objects and attributing them to an individual, and then finding ways of measuring their replication and usefulness, and rewarding the creator according to these metrics.

(6) Identifying and testing possible systems that could register and reward (in various ways) individuals with useful or aesthetic digital objects should be a priority. Of course, informal systems in the guise of social networking sites are manifestations of this process, rewarding individuals with (potentially) replication and repeated viewing of their digital objects such as travel photos. The reward is, however, directed at the organisations that lie behind the websites rather than the individuals. What would be the effect of having systems that rewarded first and foremost the individuals, the creators? Would people be even more inclined to share and proffer digital objects which they have created? (This is taken up again in §10.5.)

(7) Many people, notably scientists, have compilations of data in their personal archives that lie neglected and rarely shared, if at all. Scientific institutions are now exploring ways to encourage scientists to share such data. One interesting observation that has been made[191] is that attitudes towards data sharing vary among different scientific disciplines, and this stems from longstanding differences in the cultural tradition of disciplines. Mathematicians, physicists, molecular biologists, geophysicists and astronomers have developed an ethos of openness and sharing where other fields of science have not done so[192]. An understanding of where the convention of openness and sharing came from and what fuels it would be helpful in attempting to spread such attitudes.

---

[191] "All too many observations lie isolated and forgotten on personal hard drives and CDs, trapped by technical, legal and cultural barriers...", B. Nelson (2009) Empty archives, Nature 461: 160-163

[192] Nelson (2009), ibid

(8) Beyond scientists, what criteria would help to reassure and encourage individuals to share personally created information and objects?

(9) This discussion is, perhaps, a long way from the considerations of traditional archival repositories but with the likely proliferation of personal digital archives, archives in the wild, it is indispensable for digital curators to understand and possibly influence the environment in which they operate. The value of personal digital objects or eMSS will be fundamental.

Infoethics and novel technologies

(1) With new technologies such as haptics, high performance internet, personal fabricators, immersive virtualisation and enhanced performance of humans, many new forms of personal information will emerge.

(2) Future use of personal information and personal digital objects, and the way information technology impacts the individual prompts a field of study akin to 'bioethics': namely, 'infoethics'.

(3) It has been demonstrated that even where data are carefully anonymised powerful techniques designed for decryption can be used to reveal identities[193]. Aggregated data can be unpacked too. In the context of genomics, a new statistical technique has emerged that can resolve individual genotypes within a dataset containing aggregated data: in short it can determine whether a specific individual's genomic data are part of the dataset[194]. The impact of these possibilities needs to be examined both from the point of view of privacy and of the ability to conduct research[195].

(4) Systems need to be designed that make it easy for permissions to be sought and granted, for technologies that monitor privacy protection, and so on.

Deletion diffidence

(1) Perceived value influences the likelihood that people will retain versions of a creative work, and the efforts made in selection, retention and deletion.

(2) The present study revealed (see §3) that many people do not produce interim versions (but write over the existing version) and, even if they do, a number of respondents later deleted the interim versions that they had produced. This represents a challenge for the study of creativity.

(3) At the same time, it seemed that personal digital collections in many cases increase in volume organically with little appetite for deletion due to the effort required to select and delete 'valueless' or 'less valued' objects: evidently useful information is mixed with information that is less likely to be useful.

(4) How often should versions be made and retained for the creative process to be optimally

---

[193] Bachstrom, L., C. Dwork and J. Kleinberg (2007) Proceedings 16th International World Wide Web Conference

[194] E. A. Zerhouni and E. G. Nabel (2008) Protecting aggregate genomic data, Science 322: 44-45

[195] See also L. Backstrom, C. Dwork and J. Kleinberg (2007) Wherefore art thou R3579X? Anonymized social networks, hidden patterns, and structural steganography, International World Wide Web Conference (WWW 2007), 8-12 May 2007, Banff, Alberta, Canada

captured? Is it possible to identify automatically the timely moments, junctures, for capturing creative variants?

(5) What strategies for deletion can be devised? Would an automated ranking based on an individual's past preferences allow the quantity of digital objects retained to be matched to the individual's resources?

(6) Will it be possible for families to keep everything? The technological and economic advances in digital storage are expected to continue to be impressive. Undoubtedly more information will be stored by and on behalf of individuals than ever before. However, the expansion in capacity may be matched by the proliferation of possible objects and content to be stored by an individual. Highly detailed lifelogs are possible already for some people. Yet some form of selection seems inevitable. (An obvious candidate selection today might be retention of low resolution video instead of high resolution video.)

Box: The analysis and context of creativity: literary engineering

Writers vary greatly in the ways that they go about their practice of writing, in writing alphabetical code that stimulates, moves, inspires, illuminates and exposes. Scholars have long been interested in the process of creativity, of writing, of art and indeed of scientific insight and deliberation.

The primary path to understanding the writing process is the careful examination of the preceding drafts that led to the final version, the fair copy in hand or beyond to the publication *via* annotated proofs, and also the notebooks, the scraps of paper with inspired moments: examination of these papers, which have been termed 'genetic documentation' with the subject being termed 'textual genetics'[196]. There is a longstanding tradition of studying interim versions that goes back to ancient and medieval times. Critical analysis took a leap forward with Richard Bentley in the 18th century and Matthew Arnold in the 19th century, and today textual research employs highly sophisticated techniques. Just one example of the depths to which modern scholarship goes is provided by recent analysis of early drafts of James Joyce's *Ulysses*, newly uncovered in May 2002[197]. (See references to phylogenetic approaches applied in seeking to determining the most ancestral copies of Canterbury's Tales extant today[198].)

The complexity gets even more profound in the digital era. Umberto Eco[199] has reflected on an apparent paradox that on the one hand in printing out and annotating each version of a

---

[196] P.-M. de Biasi (2007) What is a literary draft? Towards a functional typology of genetic documentation, online, http://www.item.ens.fr/index.php?id=13599, last edited 19 January 2007. The prestigious "L'institut des textes et manuscrits modernes" has highlighted the following essays: "Genetic criticism and the creative process. Essays from music, literature and theater", http://www.item.ens.fr/index.php?id=491314

[197] P. Sopčák (2007) 'Creations from nothing': a foregrounding study of James Joyce's drafts for Ulysses, Language and Literature 16: 183-196

[198] A. C. Barbrook, C. J. Howe, N. Blake and P. Robinson (1998) The phylogeny of The Canterbury Tales, Nature 394: 839-840; note, moreover, C. Fletcher with R. Evans and S. Brown (2003): "Of course, the problem would be at least partially solved if a copy of the Tales had come down to us in Chaucer's hand but... in an age of professional scriveners the autograph manuscript was rare indeed"

[199] U. Eco (2002) Literary game of drafts, Guardian newspaper, online, http://www.guardian.co.uk/books/2002/mar/30/scienceandnature.philosophy, 30 March 2002

piece of writing being created on a computer it is possible to retain a sequence of versions, and on the other hand it is likely that in transferring the contents of the amended printout to the computer again the writer will make changes in the course of the transfer[200], and thus unless both printouts with amendments and digital versions are retained there will be losses in the sequence. Indeed if a writer simply edits on the computer (some or all of the time), there will be significant losses unless versions are carefully saved under a series of version names. In truth, and as Eco notes, it can be done. It is primarily a matter of motivation, for with the right incentives and benefits can be encouraged to keep and retain versions. Indeed every keystroke can be retained, quite cost-effectively. With increasing demand for versioning and editing to be captured, software will likely become increasingly convenient. Since Eco wrote the article Apple's Time Machine has of course taken significant steps in the retention and restoration of versions at the will of the computer user.

If the versions are retained, the authenticating and time extraction capabilities offered by the forensic approach to digital manuscripts make it possible to time and date the objects, which might be nigh impossible with paper that has not been explicitly dated, notwithstanding the possibility of dateable watermarks and paper[201]; and even if the versions are not retained, earlier drafts and fragments can sometimes be forensically recovered from word processed digital objects such as Microsoft Word documents, these being compound files with much latent and stranded information embedded unseen within them, not simply valuable metadata, but portions of, if not entire, earlier drafts[202].

One specific difficulty lies in integrating digital and analogue drafts; placing them in order may not always be straightforward. There may be instances where digital and paper versions of a draft essay or playscript are exactly the same; but in other instances there may be unique digital objects and unique analogue printouts, as it is all too easy to save over an earlier version that was previously printed out.

To put it in perspective, a reminder of how precarious the keeping of unique versions let alone interim versions has been in the predigital era is provided by the recent publication by the British Library on the manuscripts of the romantic poet John Keats: "Copies engendered further copies and, frequently, further revision on Keats's part"[203]. With his manuscripts being part of his "daily conversation", Keats was often careless with them, burning them on at least one occasion. The existence of many of his manuscripts to this day, is due to his friends and family who during his short life would take care to transcribe and share his literary outpourings, not to mention the subsequent zeal and energy of collectors and scholars who tracked them down over the years and made them available for scholarship. Yet it could so easily have not been so, and has not been so for other poets and writers. The history of existing literary and historical manuscripts is densely littered with close run survivals; and most have simply not survived or been replicated, as Johann Wolfgang von Goethe lamented:

---

[200] Depending on the precise procedure followed by a writer something akin to the process described must surely happen with handwritten drafts too, although it may happen more frequently with the computer.

[201] Kelliher and Brown make the observation that some writers, looking for somewhere to write a poem or get down a germ of literary text, "slit open envelopes and write on the inside, others seize paper which provides a record of where they are at the time - Keith Douglas at the Middle East RAC Base Depot, for instance, or John Drinkwater in a New York hotel": H. Kelliher and S. Brown (1986) English literary manuscripts, The British Library, London

[202] This is not simply a theoretical possibility and is being done in practice

[203] S. Hebron (2009) John Keats. A poet and his manuscripts, The British Library, London

"How little of all that has happened has been recorded in writing, how little of this corpus of writings has been preserved".

Decades and centuries after Keats's death, Stephen Hebron provides a portrait of what has remained: "Chronologically arranged, the manuscripts gave a wonderfully detailed account of Keats's short, intense life. Accurately transcribed and thoroughly edited, they traced his astonishingly rapid literary development and provided some clues to his creative processes. The most insignificant poems and the most personal letters all helped to complete the overall pattern of Keats's life and work, even, some thought, to the point of intrusion". "When looked at together, Keats's manuscripts can be seen as linked steps in his personal and literary progress..."[204]. This is a goal to which eMSS can realistically aspire and - with diligence - far surpass. More demanding is the creativity of Keats.

9.7 Advanced Cataloguing for Contextual Information

(1) The compilation of metadata and description and its association with the digital object will remain a core activity of the curation of personal archives, effectively adding value to the object by outlining its provenance and linking to it additional contextual information.

(2) Increasingly this activity will be supported by automated, supervised or augmented metadata extraction and annotation involving natural processing and other techniques[205]. There will remain, for a considerable time, a need to monitor these processes and to supplement them. A principal advantage is likely to be in enabling the processing of greater numbers of personal archives and objects.

(3) Techniques specifically for personal digital archives need to be researched and developed. Much existing research has been into publications and relatively structured databases.

(4) The extent to which technologies can be borrowed from other disciplines such as bioinformatics needs to be explored.

(5) A particularly valuable step would be the ability to produce or enlist ontologies suitable for personal archives, and to begin substantially to introduce semantic metadata[206].

(6) The characterisation and identification of the core metadata for the full diversity of personal digital objects is still emerging, and remains to be consolidated. If deemed desirable by the archival community, designs of suitable icons for metadata categorie and of possible procedures for their application (as recommended by the Digital Lives project) will need to be prepared.

---

[204] For an account of the British Library's literary manuscript treasures see C. Fletcher with R. Evans and S. Brown (2003) 1000 years of English literature. A treasury of literary manuscripts, The British Library, London. For more recent collecting of literary manuscripts by the British Library and a digital context see J. Andrews with J. L. John (2009) Save, get, delete, Times Literary Supplement, 13 March 2009, p 15; J. Andrews (2010) 'Laid aside'? Collecting contemporary literary archives and manuscripts, Archives (British Records Association), in press. See also the excellent blog of Kate O'Brien centred around the archive of the playwright Harold Pinter, http://britishlibrary.typepad.co.uk/pinter_archive_blog

[205] R. Feldman and J. Sanger (2007) The text mining handbook. Advanced approaches in analyzing unstructured data, Cambridge University Press, Cambridge

[206] Aspects of semantic and deep linguistic processing are being addressed by the major project SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg), http://shaman-ip.eu

(7) Systems for collaboration between institutions and for participation by diverse contributors that are again suitable for personal archives need to be tested for interoperability and effectiveness.

(8) A number of activities under the umbrella of 'enhanced curation' have been identified including panoramic photography of the writing and laboratory environments of authors and scientists; but there are many other possibilities.

Box: Automated and quasi-automated description and processing

The future potential for automated indexing, annotation, identification and description is illustrated by recent advances in unsupervised content-based indexing of video and the transcription of audio recordings[207]. A system, TrackMarks, designed to enable collaboration between human and computer in tracking the movement of humans in video has been devised, providing for the annotation of the location and identity of people and objects in video; it is concluded that TrackMarks is much more accurate than fully automated annotation but is much more efficient than a fully manual processing of video[208]. There have also been steps to allow natural language queries to be addressed at a video corpus based on spatial prepositions such as 'along the hallway' and 'across the kitchen' using encoded visual models for semantic interpretation; the system of video retrieval is being tested with naturalistic video as part of the Human Speechome Project (HSP) which is following the language development of a child[209]. Already the HSP has yielded more than 75,000 hours of audio and 35,000 hours of video which have been recorded even when there is no speech or significant activity taking place, requiring new techniques and facilities for transcribing the speech, and annotating the position and head orientation of multiple individuals. In developing the TotalRecall system the researchers at the Massachusetts Institute of Technology are already looking beyond the immediate requirements of HSP and anticipate the huge databases that are emerging in the context of medicine, security and personal applications[210].

The automatic mining of text is even further advanced. As a modest instance, email communication has been successfully modelled using PLSA (Probabilistic Latent Semantic Analysis); thus instead of simply analysing email traffic based on the 'from' and 'to' fields, the semantic content of the emails has been analysed. Email metadata such as 'sender', 'recipient' and 'date' were automatically extracted from emails in the corpus, the text content of the body of the email was subjected to text processing, and the PLSA was directed at each email yielding the essence of the topic conveyed by the email. The initial study was

---

[207] M. Fleischman, H. Evans and D. Roy (2007) Unsupervised content-based indexing for sports video retrieval, 9th ACM Workshop on Multimedia Information Retrieval (MIR 2007), Augsburg, Bavaria, Germany, 23-28 September 2007, pp 473-474; B. C. Roy and D. Roy (2009) Fast transcription of unstructured audio recordings, Interspeech 2009, Brighton, UK

[208] P. DeCamp and D. Roy (2009) A human-machine collaborative approach to tracking human movement in multi-camera video, ACM International Conference on Content-based Image and Video Retrieval (CIVR 2009), Santorini, Greece, 8-10 July 2009

[209] S. Tellex and D. Roy (2009) Towards surveillance video search by natural language query, ACM International Conference on Content-based Image and Video Retrieval (CIVR 2009), Santorini, Greece, 8-10 July 2009

[210] R. Kubat, P. DeCamp, B. Roy and D. Roy (2007) TotalRecall: visualization and semi-automatic annotation of very large audio-visual corpora, 9th International Conference on Multimodal Interfaces (ICMI 2007), Nagoya, Aichi, Japan, 12-15 November 2007

able to reveal influences of individuals on each other, and the dynamics of the topics associated with email users, and to model interactions between groups of individuals[211].

The processing of still photographs has long presented challenges in identifying locales and people. Not only are digital images accompanied by automatically created metadata including - increasingly - GPS location, but tools are being developed for recognising people, their clothes and artefacts, and for identifying locations in the images themselves. One of the most striking advances is the ability of computer vision technology to explore diverse, unstructured collections of photographs of a scene (eg community collections found on the internet of well known places[212]), computing the viewpoint of each photograph, estimating 3D models of the scene and compiling a 3D photographic reconstruction of the scene from the photographic images[213] (an approach that is available for video images too and no doubt comparable techniques will be applied to sound landscapes).

An important implication of this capability is that as the databases and collections are gathered that record all of the world's notable landmarks and cities, it will be possible to ascertain the location of a photo "not using GPS, but by matching it to a massive collection of georegistered photographs"[214]. This technology has been tested largely in the context of the internet but it is extendable to personal collections. Moreover, annotations applied to images of historic places for instance can be automatically transferred to other relevant images.

As a concluding example: even without face recognition techniques there are tools being developed that assist in the indexing of large volumes of social photos and the eliciting of social networks from them based on the propagation of information from photos that have already been indexed[215].

9.8 Future Access and Visualisation and New Research Techniques

(1) There are deep challenges in the processing of indefinite quantities of digital information and in the supplying of secure access and storage for personal data[216].

(2) New techniques need to be developed and tested for enabling researchers to make sense of complex data and freestyle information: not least in numerical and statistical analysis.

---

[211] D. Zhang, D. Gatica-Perez, D. Roy and S. Bengio (2006) Modeling interactions from email communication, IEEE International Conference on Multimedia & Expo (ICME 2006), pp 2037-2040

[212] N. Snavely, I. Simon, M. Goesele, R. Szeliski and S. M. Seitz (2009) Scene reconstruction and visualization from community photo collections, preprint. This reference mentions as an example that a Flickr search for 'Trafalgar Square' yielded nearly 100,000 photos, April 2009

[213] N. Snavely, S. M. Seitz and R. Szeliski (2006) Photo tourism: exploring photo collections in 3D, ACM Transactions on Graphics 25(3): 835-846; N. Snavely, S. M. Seitz and R. Szeliski (2008) Modeling the world from internet photo collections, International Journal of Computer Vision 80(2): 189-210

[214] N. Snavely et al (2009), ibid

[215] M. Crampes, J. de Oliveira-Kumar, S. Ranwez and J. Villerd (2009) Visualizing social photos on a Hasse diagram for eliciting relations and indexing new photos, IEEE Transactions on Visualization and Computer Graphics 15(6): 985-992

[216] S. Green (2009) The digital library programme at the British Library: goals and priorities, Interlending & Document Supply 37: 136-139

(3) In the context of the analysis of creativity, and the placing interim versions in their most probable sequence, perhaps with new metrics and techniques automated analysis can yield hypothetical sequences, with probability levels determined, for expert assessment by the scholar.

(4) Historians in decades to come will be confronted by questions of identity, as indeed has long been the case. Does this person have multiple online identities? Is this the same person using a pseudonym? Who is behind this avatar? Is this the same online event described by people who are residing in different time zones and places? The challenge lies in automating the process as far as possible. Data fusion and forensic techniques can be expected to continue to help as these continue to advance[217].

(5) Visualisation will be a central plank for initial exploratory analysis and for concluding distillation of findings, and a specific requirement will be extensive chronological mapping of relationships.

(6) Allied to the topic of visualisation and computer mapping techniques are the topics of data fusion and data mashing, where seemingly unpromising sources of data are combined in revealing or unexpected ways.

Box: Visualisation and life information

Visualisation is commonly seen as a means of condensing large or complex volumes of information in a way that makes interpretation possible or provokes new insights.

Visualisation has been used in various ways. In the context of personal archives, work with emails is of special interest. Attention has been directed at emails because of the tendency of people to retain most of them for storage and retrieval and their inherent richness of contextual information; for example the *PostHistory* visualisation project showed that users were fascinated by the overall histories of interaction in their messages, while the subsequent tool *Themail* reveals patterns in exchanged messages by identifying words that characterise an individual's correspondence with another individual and the way this pattern of word usage changes in time; in contrast to other email visualisation projects that exploit the email header information, *Themail* depends on the content of email messages[218].

Another important topic pertaining to personal digital archives is the study of collaboration. Visualisation that portrays the way communal documents are created and edited offers a better understanding of the collaborative dynamics. "When visiting a wiki, one is greeted with what looks like a conventional static Web site. Yet this serene façade conceals a more agitated reality of constant communal editing", and so *history flow* was devised for visualising the patterns of change to the document over time, revealing the social mechanics

---

[217] S. Garfinkel and D. Cox (2009) Finding and archiving the internet footprint, Conference Paper, First Digital Lives Research Conference, Personal digital archives for the 21st century, the British Library, London; J. Bleiholder and F. Naumann (2008) Data fusion, ACM Computing Surveys 41:1-41. For a perspective on the limitations of data fusion in practice, in contemporary situations, see S. Garfinkel (2008) Data fusion: the ups and downs of all-encompassing digital profiles, Scientific American, August 2008

[218] F. B. Viégas, d. boyd, D. H. Nguyen, J. Potter and J. Donath (2004) Digital artifacts for remembering and storytelling: PostHistory and Social Network Fragments, 37th Hawaii International Conference on System Sciences, Hawaii; F. B. Viégas, S. Golder and J. Donath (2006) Visualizing email content: portraying relationships from conversational histories, CHI 2006, Montreal, Quebec, Canada

of the community, the importance of fora for conflict resolution, and the necessity of prompt notification in the surveillance of changes. Beyond the wiki, the tool can be used in other contexts such as software version development and control[219].

A third area of activity has been in the analysis of literary and historical texts[220]. This shares its past with document analysis, critical scholarship and literary and historical forensics.

The strong and wide appeal of visualisation suggests that it will become an integral part of an archivally oriented personal information management in private life and in professional collaboration[221]. (Many of the intricacies of information management during professional collaboration have been elucidated recently by Cathy Marshall[222].) Much of the emphasis at the moment in visualisation research is directed at the insightful condensation and high level abstraction of information as a means of making the most of burgeoning volumes of data; but visualisation will be increasingly important too in offering ways to navigate in fine resolution along pathways through an individual's and a family's life information, and to explore digital objects, events and interactions, chronologically and otherwise.

As the origin of the term 'visualisation' suggests, this approach to presentation innately involves the visual sense; but recently the term has been co-opted in the context of haptic sensations, it becoming possible therefore to refer to *haptic visualisation*[223]. Even beyond acknowledging the requirements of people with sensory disabilities of some kind this conceptual extension has the merit of embracing a multisensory approach to the distillation and exposition of information.

9.9 Adaptive Curatorial Systems and Technologies

(1) Future curation involving the capture, storage and access of eMSS will depend on the nature of emerging technologies and the legal and ethical environment.

(2) A persistent challenge - for both individuals and repositories - arises from the unceasing and escalating flow of technological change. Alongside benefits, there are dilemmas and concerns that continually demand new or adjusted remedies that may be legal, ethical, institutional, social or technical in nature.

(3) Many curators and archivists encountered during the project expressed varying degrees of disillusion, even despair, due to an inability to take advantage of the benefits offered by

---

[219] F. B. Viégas, M. Wattenberg and K. Dave (2004) Studying cooperation and conflict between authors with history flow visualizations, ACM CHI 2004, Vienna, 24-29 April 2004, pp 575-582

[220] M. Wattenberg and F. B. Viégas (2008) The Word Tree, an interactive visual concordance, IEEE Transactions on Visualization and Computer Graphics 14(6): 1221-1228; F. van Ham, M. Wattenberg and F. B. Viégas (2009 Mapping texts with phrase nets, preprint

[221] J. Heer and M. Agrawala (2007) Design considerations for collaborative visual analytics, Visual Analytics Science and Technology, VAST 2007, IEEE Symposium, pp 171-178; S. Bresciani and M. J. Eppler (2009) The benefits of synchronous collaborative information visualization: evidence from an experimental evaluation, IEEE Transactions on Visualization and Computer Graphics 15(6): 1073-1080

[222] C. C. Marshall (2008) From writing and analysis to the repository: taking the scholars' perspective on scholarly archiving, ACM JCDL 2008, Pittsburgh, Pennsylvania, USA, 16-20 June 2008, pp 251-260

[223] S. Panëels and J. C. Roberts (2009) Review of designs for haptic data visualization, IEEE Transactions of Haptics, preprint

digital technology or to move forward effectively. These sentiments ranged from a feeling that resources are to be explicitly directed away from the archiving of personal papers to their digital counterpart, or *vice versa*, through to a feeling that digital objects are to be processed along with analogue with little or no additional resources. Some felt that their managers or institutions did not understand the implications for staff time and other resources.

(4) There was a sense also - strongly felt - that while larger institutions might be beginning to cope with the digital era, the smaller institutions are being left high and dry. Most especially, existing curators and archivists should be granted timely training and clear directions towards chosen new practices.

(5) These findings clearly speak for the importance of research into techniques for effective and timely change and redirection.

(6) It is tempting to adopt *ad hoc* once-only solutions in response to new and pressing circumstances but these risk either being unsuitable or short-lived, and indeed might never be put in practice properly as more change keeps coming, resulting in a continuing sense of playing catchup. The approach may ultimately be wasteful in resources and in time.

(7) Unless changes are addressed quickly the continual stream of them tends to produce a perpetual state of crisis (and despondency) that risks not only immediate costs but lost opportunities which can be longlasting in their impact in a world where innovators and leaders tend to be the ones that establish a new niche quickly.

(8) An open culture combining transparency and accountability; modules of activity that can be brought together in diverse and new combinations quickly and effectively; small pods of employees with special skills being brought to bear on different tasks as occasion and changing circumstances demand: these may all be helpful notions but need to be worked out in practice. There are unlikely to be simple evident solutions - otherwise these would probably have made themselves known by now.

(9) It must surely entail special training for curators (and everyone else) in order to be adaptable, anticipating and even welcoming change and novel ways of working. Conversely, some continuity is clearly essential so that people can build on the skills already developed.

(10) The requirement may not be only occasional wholesale changes of direction. Instead what is need are dynamic systems of operation that absorb and expect change: a highly adaptable approach to the inexorable and unpredictable flow of technological and social evolution.

Box: Dynamically anticipating change: incremental, autonomic and innovative

In information science, the setting up and running of systems and architectures suitable for the longterm storage and preservation of immense numbers of highly varied digital objects is one of the great challenges of the digital era. In particular, as Cal Lee of the University of North Carolina has summarised[224], there is a profound tension between forces that push for

---

[224] C. Lee (2006) Never optimise: building & managing a robust cyberinfrastructure, History & Theory of Infrastructure Workshop, Distilling Lessons for New Scientific Cyberinfrastructures, Ann Arbor, Michigan, 28 September - 1 October 2006, http://icd.si.umich.edu/~cknobel/?q=node/40

the most modern high performance computing and forces that demand reliability, trustworthiness and stability. There is a balance to be attained but it is one that has to be maintained in the face of continuing change, with an infrastructure that "is effective in the short-term but also *sufficiently flexible to remain effective in a wide range of possible future contexts*" (added italics for emphasis). A robust and flexible cyberinfrastructure is crucial for the entire lifecycle of digital curation, from acquisition to access.

Critically, as Lee points out, the longterm curation of information "requires not only robust artifacts and computer systems but also *social systems that can both withstand and benefit from changes in the environment*. The professions and organizations involved in data curation should also be cautious not to fall into a competency trap... of only being able to solve yesterday's problems" (added italics). In short, the pace of technology change is such that individuals need to be intensively and actively supported in developing and maintaining a repertoire of skills, a diversity of capabilities, and, most crucially the ability to embrace new concepts and activities while finding new contexts for and sustaining existing skills.

It is not just individuals that need to be trained for ongoing change, for continuing innovation in a shifting environment. Institutions too need to be designed for change, to be evolvable, with flexible structures and processes. Similarly, distributed archival networks not only need to develop and be integrated to the benefit of researchers in the present, but continue to operate effectively in the face of change. How can sustained cooperation between institutions be fostered?

An important parameter in organisational and network flexibility surely lies in their overall structures, and the nature of the paths of communication and decision making. Should organisations be arranged mono-hierarchically or not? How should the relationships between institutions be structured? How do networks of institutions behave when there is a mix of large and small institutions? What structures allow streamlined, cooperative and judicious decision making within institutions? Does it make sense to separate day-to-day communication and decision making from career development and advancement decisions? There are no easy answers, and in the context of a fast moving technological environment it is an area in urgent need of much more research.

In computer science and information systems practice, agile methods have been increasingly advocated in recent years: methods that are founded on the notion of 'iterative and incremental development'. Rather than attempting to specify in advance all the complex requirements (the so-called 'waterfall model'), the specifications are allowed to emerge and evolve in a gradual way with each iteration. Yet despite widespread recognition of the benefits of the iterative approach, the 'waterfall model' has proved to be surprisingly persistent perhaps due to people being set in their ways (and perhaps not)[225]. Agile techniques have gained significant ground in the software engineering community for small and intermediate projects but not for large scale projects or large institutional operations. Some computer scientists such as Peter Denning and colleagues have argued that an evolutionary system is applicable (instead of preplanned processes) to large operations even where reliability and risk-avoidance are prime concerns[226]

---

[225] P. A. Laplante and C. J. Neill (2004) 'The demise of the waterfall model is imminent' and other urban myths, Queue, February 2004

[226] P. J. Denning, C. Gunderson and R. Hayes-Roth (2008) Evolutionary system development. Large system projects are failing at an alarming rate. It's time to take evolutionary design methods off the shelf, Communications of the ACM 51: 29-31

In addition to a modular and incremental approach, they suggest the need for a common platform to which all participants in the project subscribe along with a shared vision and set of interaction rules, with this common ecosystem supplying sufficient constraints for a loose management to operate successfully.

To support their case for what they term 'evolutionary development', they cite a comparison: "In 2004, the Office of Secretary of Defense [USA] sponsored the launch of W2COG, the World Wide Consortium for the Grid (w2cog.org) to help advance networking technology for defense using open-development processes such as in the World Wide Web Consortium (w3c.org). The W2COG took advantage of a provision of acquisition regulations that allows Limited Technology Experiments (LTEs). The W2COG recently completed an experiment to develop a secure service-oriented architecture system, comparing an LTE using evolutionary methods against a standard acquisition process. Both received the same government-furnished software for an initial baseline. Eighteen months later, the LTE's process delivered a prototype open architecture that addressed 80% of the government requirements, at a cost of $100K, with all embedded software current, and a plan to transition to full COTS software within six months". "In contrast, after 18 months, the standard process delivered only a concept document that did not provide a functional architecture, had no working prototype, deployment plan, or timeline, and cost $1.5M."

Denning and colleagues are not only considering the size and cost of projects but also the pace of change in the environment: : "To avoid obsolescence, ...a system should undergo continual adaptation to the environment".

This may sound genuinely promising; but in truth there is still a very great deal to be learned and understood about such techniques and perspectives. Particular care is required when dealing with the demands of complex archives, from storage and preservation in the longterm[227] to integrated access over extensive networks. Yet evolutionary processes do warrant further analysis and investigation, especially in the context of dealing with a moving and uncertain technological future.

A related research activity that is directed at dealing with complexity and change is autonomic computing, a grand challenge initiated by IBM in response to what has been termed "a looming software complexity crisis"[228]. It is not simply a concern about the difficulty of developing and engineering software or of producing upgrades efficiently. Problems are arising even in the installation, configuration, protection, optimisation and management of software and hardware. Systems are so complex that it takes large teams of computer engineers and programmers months to prepare them, and intensive round-the-clock

---

[227] It is interesting, however, that OceanStore which is an architecture that uses the global network as a utility to store information, employs 'introspection' "an architectural paradigm that mimics adaptation in biological systems" because "manually tuning a system so large and varied is prohibitively complex": J. Kubiatowicz, D. Bindel, Y. Chen, S. Czerwinski, P. Eaton, D. Geels, R. Gummadi, S. Rhea, H. Weatherspoon, W. Weimer, C. Wells and B. Zhao (2000) OceanStore: an architecture for global-scale persistent storage. In: Ninth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2000), ACM, Cambridge, Massachusetts

[228] Essentially anticipating that software will become too complex for direct and timely development and management, efforts are being directed at adopting biologically inspired processes such as self-healing and protection, immunity, self-management, robustness, learning and the use of genetic algorithms, J. O. Kephart and D. M. Chess (2003) The vision of autonomic computing, Computer: 41-50

attention to look after them[229]. Autonomic computing aims to create self-managing systems, and draws inspiration from biology including immune and self-healing systems[230]. The emerging field of web science has also promoted systems biology as a useful source of ideas and models, and in particular the need for new evidential methodologies for anticipating "how human behavior will affect development of a system that is evolving at such an amazing rate" is emphasised, as is the need to understand how to engineer web systems with desirable properties at scale[231].

Innovation and research are crucial in coping with change: research because it is helpful to anticipate as far as possible what is coming or might be coming, and innovation because it offers solutions, stimulates agility and adaptability and may give an institution or field of activity a head start in a new emerging niche[232]. In some ways quirky ideas or just small but genuinely novel ideas can play a role akin to randomisation in evolutionary processes[233]. The trick may be to welcome and test ideas actively and repeatedly, while not to letting them prevail without frequent reiterative testing.

Ed Catmull, cofounder of Pixar[234] and the president of Pixar and Disney Animation Studios has offered the following principles under the heading, Pixar's Operating Principles: (i) everyone must have the freedom to communicate with anyone; (ii) it must be safe for everyone to offer ideas; and (iii) stay close to innovation happening in the academic community. Catmull, incidentally, incorporates iteration as a key process; he also discusses the importance of breaking down barriers within organisations. It seems worth quoting Catmull at length[235].

"Getting people in different disciplines to treat one another as peers is just as important as getting people within disciplines to do so. But it's much harder. Barriers include the natural class structures that arise in organizations: There always seems to be one function that considers itself and is perceived by others to be the one the organization values the most. Then there's the different languages spoken by different disciplines and even the physical distance between offices. In a creative business like ours, these barriers are impediments to producing great work, and therefore we must do everything we can to tear them down".

"Walt Disney understood this. He believed that when continual change, or reinvention, is the

---

[229] In some sense cloud computing itself can be seen as a means of injecting greater agility into systems in the face of growing complexity

[230] L. A. Segel and I. R. Cohen, editors (2001) Design principles for the immune system and other distributed autonomous systems, Santa Fe Institute Studies in the Sciences of Complexity, Oxford University Press, Oxford; L. N. de Castro and J. Timmis (2002) Artificial immune systems: a new computational intelligence approach, Springer, London

[231] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt and D. J. Weitzner (2006) A framework for web science, Foundations and Trends in Web Science 1: 1-130; J. Hendler, N. Shadbolt, W. Hall, T. Berners-Lee and D. Weitzner (2008) Web science: an interdisciplinary approach to understanding the web, Communications of the ACM 51: 60-69

[232] It is well known in the commercial and marketing context that it is the early starters that tend to occupy the niche in the marketplace

[233] One aspect of the merit of randomness is highlighted by L. Tuck, F. Vetere and S. Howard (2008) Abdicating choice: the rewards of letting go, Digital Creativity 19(4): 233-243

[234] Pixar is, of course, the highly successful and ground breaking computer animation company that Steve Jobs nurtured when he was not at Apple Computers

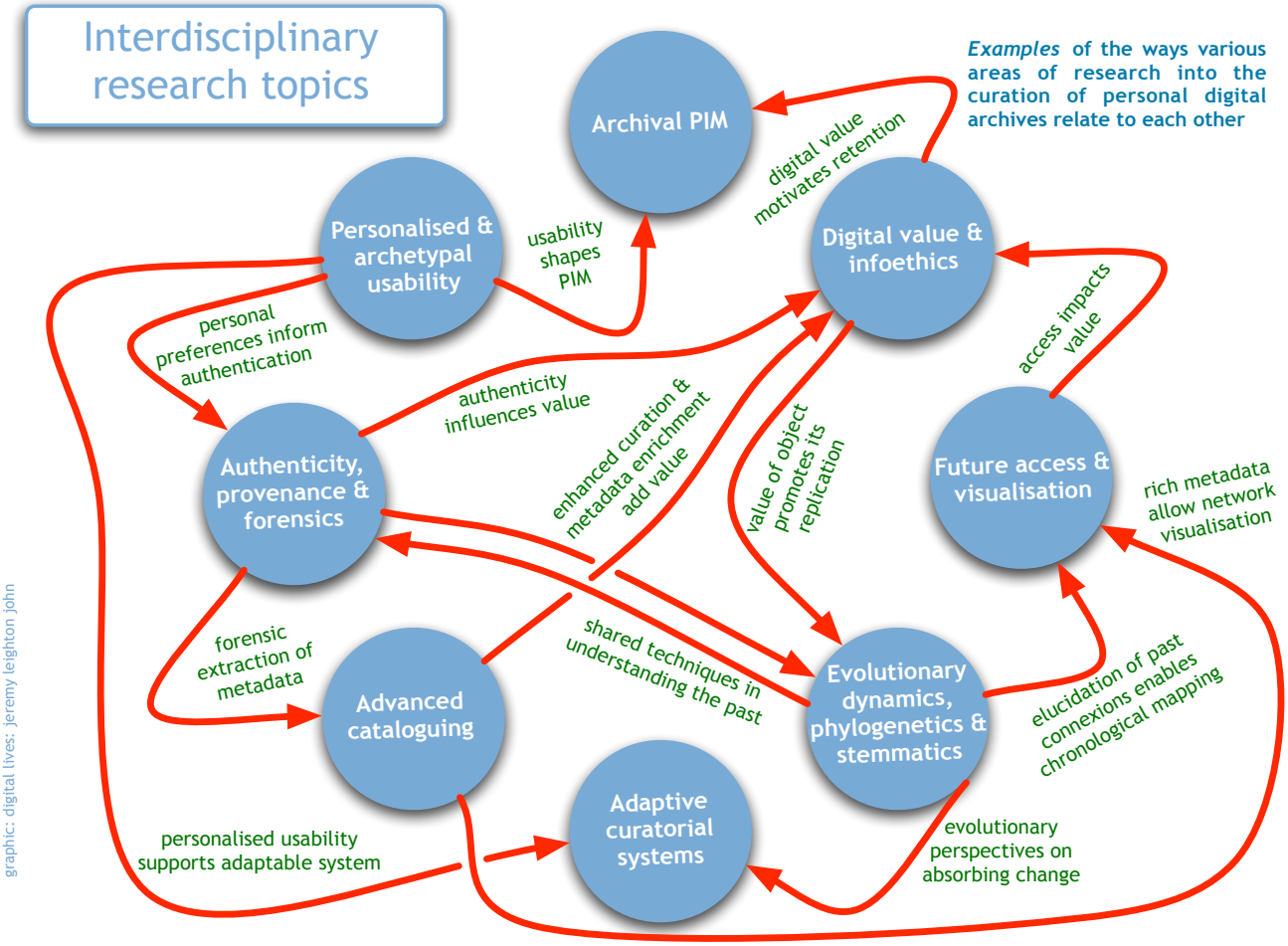
[235] E. Catmull (2008) How Pixar fosters collective creativity, Harvard Business Review: 64-72

norm in an organization and technology and art are together, magical things happen. A lot of people look back at Disney's early days, and say 'Look at the artists!' They don't pay attention to his technological innovations. But he did the first sound in animation, the first color, the first compositing of animation with live action, the first applications of xerography in animation production. He was always excited by science and technology".

This is one view from one organisation. There are very many types of players, participants, potentially involved in the curation of personal digital archives, both online and in the real world, within and beyond a repository or other institution, and so on. To take the procedure forward will require some research into organisational options and implications.

It seems possible, nonetheless, that a good place to start is with combinations of small teams, adopting modular structures and incremental and iterative processes, along with open discussions and a respect for unplanned meetings, shared vision and understanding across all teams, and with individuals who combine a recognition of the value of novel technologies and activity metrics with an appreciation of art, literature, science and scholarship.

Technology in its broadest sense lies at the heart of much of what humans do. As new technologies arise, artists take them to the limit, scientists reach for previously unaccessible spaces in the natural universe, writers explore ancient and modern themes in the new society that accompanies the new technologies, and decision makers seek to manage and take advantage of the complexities of some organisational unit as it engages with the surrounding culture and its new technologies.


Interdisciplinary research topics
Examples of the ways various areas of research into the curation of personal digital archives relate to each other
Archival PIM
Personalised & archetypal usability
Digital value & infoethics
Authenticity, provenance & forensics
Future access & visualisation
Advanced cataloguing
Evolutionary dynamics, phylogenetics & stemmatics
Adaptive curatorial systems
digital value motivates retention
usability shapes PIM
personal preferences inform authentication
authenticity influences value
enhanced curation & metadata enrichment add value
value of object promotes its replication
access impacts value
rich metadata allow network visualisation
forensic extraction of metadata
shared techniques in understanding the past
elucidation of past connexions enables chronological mapping
personalised usability supports adaptable system
evolutionary perspectives on absorbing change
graphic: digital lives: jeremy leighton john

Chapter 10: Notes Towards a Strategy for Personal Digital Archives

10.1 Proposed Actions and Model Strategy

(1) This chapter seeks to sketch the beginnings of a strategy for ensuring the longterm use and availability of personal digital archives, centred around a series of mutually supporting activities. These will need to be founded on a judicious combination of (i) a wide cooperation between repositories that is both international and national, (ii) the occupation by repositories of specialist niches of operation and expertise, and (iii) the aid and participation of people outside the repository itself: creators, curators (of other institutions) and consumers.

(2) Over the duration of the Digital Lives project, many observations and suggestions emerged and this chapter aims to document a number of examples grouped according to the proposed strategic activities. These notes help to ground the presentation in practicalities that might support the achievement of strategic goals and vision. It is not intended to be comprehensive, covering every aspect of digital curation of personal archives, but to give a sense of what is required and what is becoming possible.

10.2 Mission

Three aims can be identified for repositories operating at various scales:

- to advance the excellence of collections of personal digital archives;
- to offer exceptional high-quality advocacy, advisory and specialist support to other organisations in the holding of personal digital archives; and
- to help to make it possible for each and every individual to have a sustainable personal digital archive and to engage with repositories and their curatorial and research activities.

10.3 Justification

There are six basic justifications:

- advanced humanities scholarship with enriched primary content and powerful historical, literary and contextual analysis;
- far reaching scientific research in social and natural sciences through digital collections and mediated access to personal digital archives;
- digital literacy and digital economy;
- widespread public engagement and digital inclusion;
- continuity and expansion of the historical record; and
- theoretical and practical foundations for personal informatics.

10.4 Strategic Activities and Modules

At every stage of the lifecycle there are obstacles and novel challenges to be met.

Ten strategic areas of essential activity can be identified, and these can be conveniently organised as five pairs, or modules.

(1) Facilitation & Motivation
(2) Advocacy

>> STRATEGIC MODULE FOR **PUBLIC AND ORGANISATIONAL ENGAGEMENT**

(3) Advice
(4) Training

>> STRATEGIC MODULE FOR **SKILLS**

(5) Collections & Capacity Building
(6) Participation

>> STRATEGIC MODULE FOR **CONTENT & ACCESS**

(7) Collaborations & Partnerships
(8) Services

>> STRATEGIC MODULE FOR **PROFESSIONAL INTERACTION**

(9) Tools & Workflows
(10) Research & Development

>> STRATEGIC MODULE FOR **PERSONAL INFORMATICS**

10.5 Actions for Engagement Module

Central to the success of the entire strategy will be the ability (i) to motivate creators in sustaining their personal archives, and (ii) in inspiring decision makers and legislators through effective advocacy.

Facilitation and motivation

Promoting capture and retention

>> Many creators are already striving to maintain a personal archive. More creators need to be alerted to the benefits of retaining items for themselves and for future research; and motivated at this time to keep a sustainable personal archive.

>> Even those people who intend to retain files for the duration of their working lives do not necessarily have in mind subsequent generations.

>> Attitudes towards personal digital objects seem not to be the same as those towards analogue objects such as diaries and notebooks, and so the value of research and reuse of digital objects should be highlighted.

>> Everyone should be encouraged to download files from their online accounts to their local computers; and made aware of the implications of the diverse terms of the services offered by online service providers.

>> An institution also needs to pay close attention to the interests of third parties such as those individuals who are represented in the archives but are not the depositors of the archive. It must be readily possible for people who wrote to the originator of an archive to get in contact with, and be engaged by, the repository: to give permissions, or to indicate wishes regarding intellectual property or privacy, or to provide further materials. Perhaps more than any other factor this points to the need for a system of effective engagement with the public.

>> Reception for distributed archival objects, 'recovery of the dispersed': some people may have only a few archival materials – such as emails or word documents or paper letters received – that would fit well into existing national or local collections. Channels for receiving these relatively isolated items should be could be made more effective.

Box: Registries and incentives

Personal digital objects or eMANSUSCRIPTS might be shared for a variety of reasons. Some might be made available widely and openly as objects of personal creativity for others to enjoy, appraise or witness; others might be made available to researchers as a personal digital archive in its entirety or in parts through a mediated access programme with judicious anonymisation and other privacy guards. Some people might be glad to receive financial reward, in token or in full market value, for the use of these objects; some might enjoy the benefits of an overtly enhanced reputation; indeed some might receive rewards in kind, through greater access to the created objects of others. Other people might be gratified purely by a private sense of charity.

Many people might welcome the opportunity to contribute to important research, especially longitudinal social research over a number of years - with regular interaction with the research programme giving volunteers satisfaction. Nonetheless, the topic of information sharing raises a broader matter of incentives and how the sharing of personal digital objects and content might operate in a wider context.

In the digital era with personal objects existing in vast numbers and with timely delivery likely to be favoured, some degree of automation is essential along with networked access to enable the possibility of controlled sharing. As the highly regarded computer consultant Esther Dyson has observed in the context of online delivery and the marketplace of

copyrighted materials "reliable and comprehensive registries of copyrighted materials, combined with digital-rights management software" are central to the process[236].

Registries provide for the cross referencing from an object's identifier to the object itself. An identifier might be a web domain name or an RFID tag (eg for a manufactured item of physical goods). One of the most well known identifiers is the digital object identifier (DOI) used with monographs and serials, books and periodicals. Some identifiers adopt the Handle System whereby the identifier, handle, is incorporated within the digital object itself.

Dyson cites the company Content Directions, Inc (CDI) as an example of a DOI registrar that has applied the Handle System in order to meet the requirements of the scientific publishing community, enabling the placing of high value content online in a controllable way that ensures "proper credit in the academic world and also for the usual commercial reasons of protecting intellectual property"; thus it provides a link from a research citation in a publication of one publisher to the content in the cited publication of another publisher. The system was designed to cope with very high volumes of registrations and references, and with any language or alphabet, and be permanent. The approach can be used in various ways to track the histories of an individual item[237].

Identifiers are being used in various contexts for individuals, organisations and licenses as well as textual works, catering for resolvability, context-sensitivity and semantic interoperability[238]. Of particular interest is the ability of the Handle System to provide for trusted resolution using public key infrastructure[239], as this might well be usefully applied to personal digital objects.

A useful parallel is to be found in the context of the personal archives of scientists. A major obstacle for the sharing of scientific or scholarly data is the concern that individuals who have carefully compiled the information will gain little benefit from making the data available. Research councils are increasingly insisting on archival measures in order to ensure that scientific data are shared[240]. This is an important step but it alone is unlikely to lead to the capture of scientific information that is compiled by scientists in a personal capacity, and mechanisms are required anyway.

Further solutions are emerging. James Boyle of Duke University and founding board member of Creative Commons is cited by Nelson[241]: "He points to a music site associated with Creative

---

[236] E. Dyson (2003) Online registries: the DNS and beyond..., Release 1.0, Esther Dyson's Monthly Report 21,16 September 2003

[237] Dyson (2003), ibid, pp 27-28

[238] N. Paskin (2010) Digital Object Identifier (DOI) System, in Encyclopaedia of Library and Information Science, Third edition, Taylor & Francis, preprint

[239] Paskin (2010), ibid

[240] For a digital preservation and cost perspective of research data, see Keeping Research Data Safe project, including follow up project 2, http://www.beagrie.com/jisc.php

[241] B. Nelson (2009) Empty archives, Nature 461: 160-163

Commons known as ccMixter[242], in which users can upload... a trumpet solo or other musical samples. Users are free to remix the samples into new tracks. But when they do, the program automatically keeps a continuous credit record. So why not implement a similar system that would add a link back to a database every time a researcher repurposed some data". Could this kind of approach be extended to all the personal digital objects of individuals more widely, and not just scientists?

In an intriguing paper[243], Jonathan Schull of the Rochester Institute of Technology has outlined ways in which the recording of the use of digital information objects could be combined with systems of reward, as an extension of the 'superdistribution' concept that is used to distribute and sell software. "Consumers who recommend and distribute products to their friends are providing marketing, distribution, sales, and technical support services to their recipients. Why should they not be compensated? And if we are going to compensate them, why not compensate them with something that we can 'manufacture' at no cost - the right to consume other digital products". Individuals who similarly share their personally created and useful objects might be rewarded with the right to use other objects within recognisable communities of individuals.

Schull also discusses the concept of a Personal Information Trust (PIT) which would reward individuals when marketers use their personal information. This notion might be extended to the use - by scientists, scholars and other *bona fide* researchers - of all kinds of personal information held in an individual's personal digital archive[244].

Of course, organisations such as Google and Yahoo! already reward individuals for using their personal information in the form of access to search capability. However, this relationship might be made more transparent (at least to the individual), with the value, the usefulness, of an individual's personal information being quantified and the record made available to the individual.

With a Personal Information Trust system, an individual could register with the PIT; and, from their side, organisations could check with the PIT (or a network of PITs) to see if a person is registered and thus can be rewarded or at least provided with a record of the extent to which their personal information has been used. The onus would be on the individual to register and to provide contact details to the PIT. Individuals who do not engage with the system would not benefit in this way even though some of their personal information (eg browsing habits or search keywords) would presumably still be used by organisations. A key issue would be scalability and this should be researched. As Schull notes the tracking of the use of an object invokes privacy issues which will - as in other contexts - need to be addressed.

This digital and networked system would in some ways be reminiscent of the much older and

---

[242] The software behind ccMixter (http://ccmixter.org) is ccHost which is open source and designed to facilitate the sharing and remixing of multimedia content; ccHost employs metadata tags and tag extractors in order to ensure that: a mix is linked to the samples and vocals which are incorporated in the mix; remixes link to the source track;, and there are comprehensive links to the creators (http://wiki.creativecommons.org/Cchost; see also http://www.linux.com/archive/feature/49565?theme=print); ccHost is not restricted to music and has been used in other 'remixing' contexts

[243] J. Schull (2007) Predicting the evolution of digital rights, digital objects, and digital rights management languages, in Digital rights management: an introduction, editor D. Satish, ICFAI Books, Andhra Pradesh, India

[244] See also http://www.openprivacy.org and its annotated bibliography http://www.openprivacy.org/bibliography.shtml

more limited system that has operated within the United Kingdom for rewarding writers according to the use made of their books by users of the Public Libraries.

In many cases, with precise tracking and measurement of the use of personal information by researchers, individuals may feel sufficiently rewarded by being explicitly and quantifiably aware of how much they have contributed to the important research.

Also of interest is the emergence of the whuffie, a virtual currency based on an individual's online reputation which is ascertained by tracking an individual's activity[245]. By helping other people one can increase one's whuffie rating.

Reputation phenomena manifest themselves in various ways in the digital era. In an illuminating paper Donath and boyd[246] examined the individual profiles that people prepare and supply as participants in social networking sites and the connexions which these people establish and display. Donath and boyd ask: "Why do people display their social connections in everyday life, and why do they do so in these networking sites? What do people learn about another's identity through the signal of network display?"; and explore various interpretations including the display of connections as a way of verifying personal identity, ensuring cooperation, signalling success, trustworthiness[247], making new connections, making suitable connections, validating social ties, expanding the personal network while keeping separate the contexts of an individual life, and the cost as well as the benefits of maintaining displays of social ties. As Donath and boyd observe there is a concept in the field of evolutionary biology (with a parallel in economics), under the umbrella title of 'signal selection', that suggests that ornamental signals are (or need to be) costly in order to be valid signals of quality; but in an emerging phenomena such as social connectivity on the web, simple measures of connections may be (in present circumstances) misleading, and the potential for deception is recognised.

Advocacy

Promoting an understanding of personal archives

>> It is strongly recommended that the value and requirements of personal digital archives be vigorously and publicly promoted.

>> In order for resources to be made available for archives generally, there is a pressing need to persuade decision makers and funding agencies of the increasingly immense value of personal digital archives for individuals and for future research. A number of interviewed archivists felt that there has been a general failure of resources to meet the initial set up costs.

[245] http://thewhuffiebank.org/; Whuffie is calculated according to (i) public endorsements (eg a message by the subject individual is retweeted), (ii) level of influence, (iii) endorsement by a whuffie-rich person, (iv) content of retweeted message is exclusively derived from the subject individual, see http://thewhuffiebank.org/static/faq

[246] J. Donath and d. boyd (2004) Public displays of connection, BT Technology Journal 22(4): 71-82

[247] See also J. Donath (2007) Virtually trustworthy, Science 317: 53-54

>> Illustrative instances and demonstrations of how personal archives can benefit individuals and their families would be useful in advocacy.

>> A register should be maintained of examples of research using personal digital archives and sources of personal information, including new types of research. The establishment of a register of good examples of research that has benefited from the kind of information that is to be found in both analogue and digital contemporary personal archives is highly recommended.

>> Publishers of popular books and ebooks should be approached in order to encourage mention of primary sources in acknowledgements. Frequently popular and even academic books cite secondary sources with little or no indication of the primary sources and of the location of those sources that ultimately made the book possible. Conversely, researchers working on behalf of the register of research might ascertain for popular and other works (as necessary) the ultimate primary sources for historical observations.

Personal control

>> In view of the impact of technologies on privacy and the increasing use of personal information by commercial and other organisations, it seems possible that a desirable trend for the future would be for individuals to have increasing control of their own information, and, perhaps, be provided with details of the way their personal information is used. Is this feasible, tenable? Would this entail new legal rights or simply greater awareness of what is already possible? Might this be an effective antidote to any future ethos of overzealous surveillance, to any holding and using of personal information to excess by governments, commercial operations or anyone else? It would be useful to contemplate further the pros and cons of these kinds of possibilities.

>> Tracking of the way personal information is used raises privacy issues of its own but it might be possible to have oversight by a number of trusted organisations involved in monitoring human and digital rights. Whatever the specifics, it ought to be possible to design systems that incorporate checks and balances.

>> Legal reform or clarification would be helpful if aimed at: (i) reducing the risk to repositories in collecting the widely accessible and public components available on the web; and (ii) reducing any possible legal risk to a repository when attempting to access a creator's social networking account following his or her death when permission of the family or executor has been granted.

>> Should there be an obligation for every individual to have full downloadable access to information that concerns them such as the personal profiles and other information on their social networking accounts? Is it feasible for people to be granted more details of the way their personal information is being used? Should individuals be free to pass their information to a longterm repository or to be able to grant mediated access to their personal archive?

10.6 Actions for Skills Module

Training and advisory services are essential if the population of archivists and curators are to respond effectively to the wide ranging demands made upon them.

Advice

Interactive advice and awareness

>> More and more people are already documenting their own lives, and others - creators, curatorial colleagues and consumers or researchers - might be encouraged to do so through advice that is tailored for diverse audiences and levels of expertise.

>> Writers and their representatives frequently work independently of any institutional support and warrant special attention: even the more technically minded often need to be guided in procedures for longterm preservation and access, and in their choice of strategies.

>> Specific advice is required with regard to priority in selection of eMSS for retention. What kinds of digital objects serve different research purposes? How can this utility be protected?

>> A meta-advice service may be useful, giving advice about the giving of advice: enabling local curators and archivists in the passing on of guidance to others, including people generally.

>> Mutual awareness and appreciation of the concerns of creators, curators and consumers needs to be increased. Meetings between creators and users may help each group to understand privacy issues and also the kinds of digital object that are useful for research.

>> Routes for interactive advice enabling enquiries, feedback and occasionally unsolicited approaches. Channels for communication include emails, VoIP with or without webcam; and also one-to-one surgeries and group discussion meetings.

Guidelines

>> There is a requirement for instruction manuals and guides that are straightforward and are regularly updated. These need to be directed at the three audiences: creators, curators and consumers, and should embrace not only legal matters but also technical issues.[248]

>> Guides to the supplementing of metadata and contextual information by creators and their families and others in order to enhance a personal digital archive undoubtedly would be welcomed.

>> It would be important to facilitate the development and understanding of new research techniques, and to maintain a register of these techniques for users, scholars and scientists.

>> In addition, to conventional text guidelines, audio and video podcasts can be produced.

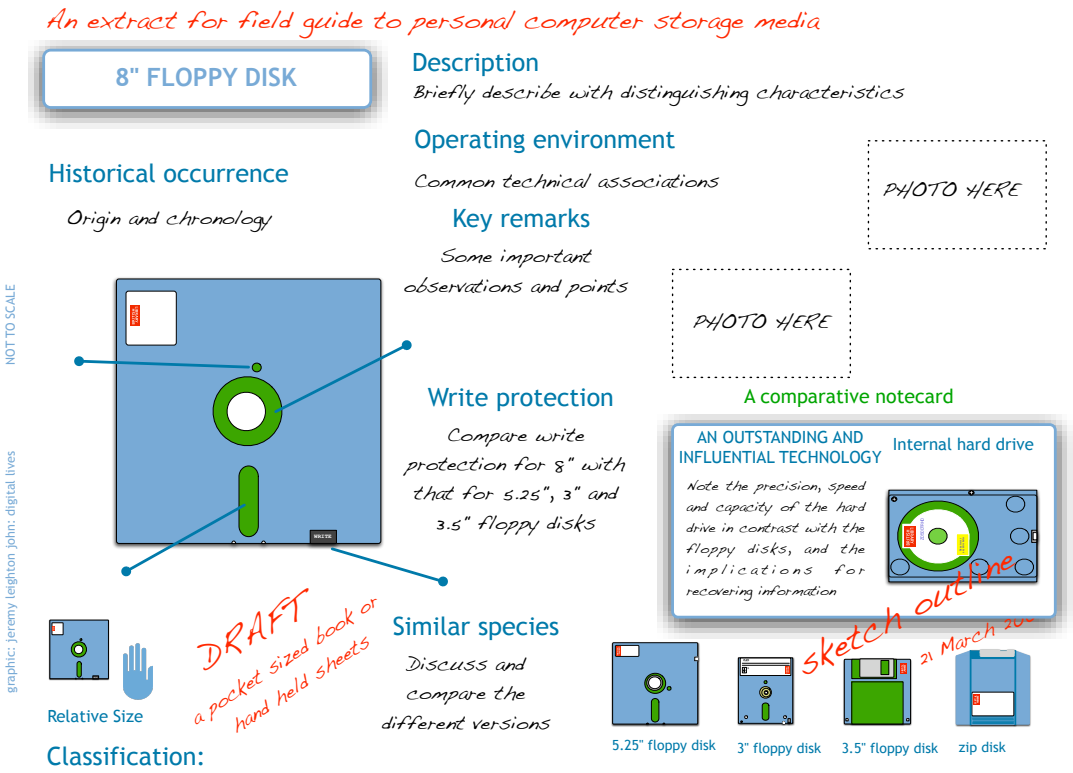
Training

Creators, curators and consumers

>> The interviews conducted by this project suggested that many if not most people are self-taught or received informal training. Many misconceptions were apparent especially in the realm of backing up, storage and preservation, and there is a widespread lack of awareness of the distinction between backup and digital preservation.

---

[248] There are already useful resources including the excellent Paradigm workbook, http://www.paradigm.ac.uk/workbook/index.html (see also its successional FutureArch website), and the Library of Congress website. G. Bell and J. Gemmell (2009) cite a book by A. Bainbridge (2009) Organize your digital life. How to store your photographs, music, videos, and personal documents in a digital world, National Geographic, Washington DC. Some conformity and design for interoperability would be helpful. Most importantly, there is a need for guidelines to be reliably up to date, and this a burden that is likely to be best shared through the combined attention and efforts of many archivists and institutions, perhaps through a suitable web hub



>> Training is important for maintaining standards across the community of creators, curators and researchers: enabling consistency and accuracy.

>> There is a pressing demand for training among curators and archivists. For example: in forensically-sound and archival standard digital capture and preservation; in creating an inventory on reception with fast digitisation of analogue objects; and in enhanced curation activities involving use of panoramic equipment, portable lighting, video cameras, audio recorders and graphical software.

>> Institutions and individual curators would benefit considerably from an ongoing training of individual curators that is determined to produce and sustain suitably skilled people. The pace of technological change is such that it is essential to empower existing people by continually and intensively developing their versatility and adaptability.

>> It may be most practical for smaller institutions to have multitasking archivists and curators who have wide understanding of digital aspects, with complementary specialisms. Accordingly, significant effort must be directed at the extra training or retraining of existing, often conventionally trained, archivists as well as training young archivists.

>> It is important that curators are fully aware of the power of digital and networked technologies in terms of data fusion and forensic analysis. Curators need to keep abreast not only of digital preservation and curation but of the kinds of research activity that are being undertaken by scholars and scientists.

>> Specialist topics for curators will include: (i) emerging technologies and possible impacts; (ii) the identification of fakes and the inauthentic[249]; (iii) photographic and video techniques; (iv) use of graphical and multimedia software, (v) oral history methodologies[250]; (vi) exercises in social, legal and ethical change; (vii) resource implications for managers and curators.

>> Researchers correspondingly need to gain new research skills and knowledge: (i) computer techniques for working with new media, new digital objects, for working with hexadecimal code; (ii) ancestral computing and the history of the use personal computing, (iii) the application of tools for tasks as diverse as documenting performance art, analysing body language in moving images, and new methods of textual analysis in authorship attribution; along with (iv) the ethics and implications of potentially intrusive forms of statistical and numerical analysis.

Didactics and certificates

>> Career paths and currricula need to be developed and be responsive to further change[251].

>> A significant amount of the training will need to be hands on; but much of it might be done through virtual interaction, video conferencing, and other remote teaching methodologies.

>> Master class sessions and tutorials can be conducted with world renowned experts.

---

[249] H. Malik and H. Farid (2010) Audio forensics from acoustic reverberation, International conference on acoustics, speech, and signal processing, Dallas, Texas; H. Farid (2009) Image forgery detection. A survey, IEEE Signal Processing Magazine 26(2): 16-25, March 2009; W. Wang and H. Farid (2009) Exposing digital forgeries in video by detecting double quantization, ACM Multimedia and Security Workshop (MM&Sec 2009), 7-8 September 2009, Princeton, New Jersey. Of particular interest is the paper describing a technique for authenticating printed and scanned text documents: E. Kee and H. Farid (2008) Printer profiling for forensics and ballistics, ACM Multimedia and Security Workshop (MM&Sec 2008), 22-23 September 2008, Oxford, United kingdom

[250] Training for oral history provides a useful starting model: http://www.ohs.org.uk/training

[251] The DigCCur project at the University of North Carolina at Chapel Hill is undertaking significant work in the development of digital curation curricula. Within the UK, the Digital Curation Centre is doing likewise

>> Virtual worlds and the metaverse offer research, conference and training at online venues or digital centres. Although not a panacea, the approach has a number of advantages including global reach, reduced travel costs, and, where appropriate, anonymity.

>> Mobile learning technologies make it possible to extend teaching and training beyond the classroom, providing opportunities for complementary learning experiences.

10.7 Actions for Content & Access Module

The core role of curators and archivists and their institutions remains in the holding and description of personal objects for research. In principle it will be possible to capture, store and preserve more personal information than ever before; but it will require highly efficient and adaptable processes to do so.

Collections and capacity

Complex archives

>> Bearing in mind the 'digital capture imperative', it is sensible to locate and engage with 'high profile' creators early in their careers. Long before retirement, influential writers, scientists and sociopolitical reformers can be approached with advice and support.

>> Along with the proliferating quantities of contemporary digital media and files, there is also a growing legacy of neglected personal digital objects that require retrospective processing urgently. The population of readable floppy disks in the United Kingdom and elsewhere can be expected to fall rapidly in the coming years.

>> One of the most profound distinctions between analogue and digital archives is the potential volumes involved. Analogue archives can also be large, and indeed could be almost limitless in size; but this is extraordinarily impractical for most organisations and individuals. With digital personal archives, large volumes are technically feasible. For the extensive benefits to be realised, and for research expectations to be met, it is essential therefore that repositories address the ability to handle large volumes of information either internally or through shared or delegated responsibilities.

>> Hybrid collections remain, and will predominate for the foreseeable future, with even the paper components continuing to burgeon.

>> An integrated approach to paper and digital archives can be enabled through fast digitisation on reception of analogue objects such as papers, photographs, ciné and analogue sound and video recordings.

Capture

>> A series of tools already exist (as highlighted by the present study) that are relevant to personal digital archives for digital capture, authentication, audit control, emulation, metadata extraction, file type characterisation, significant properties and values extraction, conversion and normalisation, synchronisation, planning, testing, integration, curatorial examination, annotation.

>> Enhanced curation activities directed at adding value to the core personal archive can be initiated even by small repositories. The more careful and complete the capture of ancillary contextual information, the greater the potential use and value of the original archive itself. Activities include: 3D graphic imagery of the creator's environment, encompassing adjacent rooms or even entire cottage, flat or garden with writer's shed, and the compiling of a life inventory and 3D images of artefacts.

>> Archives can be examined with care and detail in the presence of the originators in order to facilitate selection and the addition of metadata.

>> Many curators routinely conduct site visits, and mobile technologies will be helpful in a number of ways such as the preparation of preliminary documentation.

Optimising requirements

>> Repositories are required to meet diverse demands while being subject to numerous constraints and uncertainties. The process of doing so while seeking to maximise benefits within the operating environment is especially complex for public repositories that are obliged to consider the wishes of a very wide variety of stakeholders.

>> Procedures for managing risks and securing licenses, for example, need to be balanced by the need to encourage individuals with personal digital archives to engage with a repository. At the same time it is necessary to meet the research imperative of effective access to personal digital objects and their contents.

>> A major enabler is the establishment of a highly transparent and accountable system based on straightforward and clear policies combined with a prompt, efficient and sympathetic system for addressing any concerns or queries.

>> To a significant extent the onus can be passed on to the users, the researchers, with access agreements, as well as making the processes for deposit clear to originators. Ethical committees that give a mandate as well as direction to policies might make it possible to avoid potentially labyrinthine practices.

>> Institutions have long needed to find ways to deal with a disjuncture between legal requirements and public expectations. This has intensified in the digital era. In balancing the costs and benefits it is important to develop a finely tuned awareness of contemporary practices beyond the repository in other spheres of digital life.

>> In sustaining and improving its reputation, a repository is required to provide for effective intake of archives, protect the privacy of individuals, maintain digital rights and make possible scholarly access to the archives. An essential goal is to engender the trust of originators over the longterm and it may be better in some crucial matters to occupy a higher ethical standard than is required by law. At the same time this has to be communicated effectively to the user community. An institution's reputation may be impacted by the perception (or reality) of unjustifiable and excessively cautious regulations that deter access and deposit.

Access

>> Clear benefits to be gained from digital archives are remote and wide access to the personal digital objects, by multiple users, across collections with more flexible and powerful analysis of information becoming available.

>> A single machine with ports blocked and printer connection disabled in a Reading Room is an option for where there are copyright and other access restrictions in place.

>> Emulation can be used to allow a scientist's early homemade computer programs to be run and witnessed by researchers. Furthermore, researchers could be given the ability to experiment with the program in ways that the originated did not do: eg modify the parameters of a model.

Box: Virtual archival computing

The traditional approach towards digital archives and libraries has been to focus on the individual digital objects, making them available independently albeit with (hopefully) rich complementary information and metadata. In the context of personal archives, there is the possibility of making available an entire disk and even the entire personal digital archive, so as to allow the researcher to experience the original creative and operating environment of the originator. Virtual forensic computing offers a route to this functionality.

The virtual experience is highly portable meaning that it can be made available on workstations with restricted access, a single reading room or, where appropriate and permitted, on the web. The computer forensics scientist Michael Penhallurick has researched many practicalities underlying virtual forensic computing[252]. The conventional way to be able to interact with the original system is to restore (create a duplicate of) the original disk (with its operating system, applications and files) from the disk 'image', and boot up this clone. As the investigator examines the system, it of course changes and is no longer the original. A dynamic write blocker such as the Shadow Drive of Voom can be used[253] to prevent changes, and write protect the restored disk while allowing dynamic interaction. Nonetheless there are disadvantages such as space requirements and time taken for preparation; there are further complexities and constraints when dealing with a number of disks.

Another more versatile approach, however, is to make use of virtual computing tools whereby virtual hard drives can be accessed by a virtual machine, a virtual computer. The software Mount Image Pro of GetData[254] is able to mount multiple 'dd' disk 'image' files as well as virtual file systems such as those of VMware independently of any forensic software; it also works with the Advanced Forensic Format 'image' file as well as the proprietary one of Encase. Having mounted the disk image as an emulated physical disk using Mount Image Pro, a suitable virtual machine can be created (eg VMware) and the 'raw disk' can be added to the virtual machine, which is booted up using the original operating system that resides in the 'raw disk' that has been added to the virtual machine. Sometimes it is helpful to first create a virtual disk (eg using VMware) and then to clone the 'raw disk' to this virtual disk[255].

One of the key advantages of using a virtual environment is that it allows for highly repeatable, controlled and referable experiences suitable for the scholar or scientist who is required to study the subject in an academically accountable way. A significant advantage of capturing entire disks - or at least capturing the software (and profiles of the hardware and services) along with the focal files - becomes clear: the captured disk 'image' comes ready-made (in many cases) with the appropriate settings and preferences of the archive's originator; moreover, the captured system is authenticated by means of the hash values, and software can be identified in detail by means of hash libraries.

---

[252] M. A. Penhallurick (2005) Methodologies for the use of VMware to boot cloned/mounted subject hard disk images, report, Cranfield University, 26 pp, March 2005; M. A. Penhallurick (2005) Methodologies for the use of VMware to boot cloned/mounted subject hard disk images, September 2005, Digital Investigation 2(3): 209-222

[253] The use of this device is being explored by the Digital Manuscripts Project at the British Library in the context of personal digital archives

[254] http://www.getdata.com, http://www.mountimage.com

[255] M. A. Penhallurick (2005) Methodologies for the use of VMware to boot cloned/mounted subject hard disk images, report, Cranfield University, 26 pp, March 2005; M. A. Penhallurick (2005) Methodologies for the use of VMware to boot cloned/mounted subject hard disk images, September 2005, Digital Investigation 2(3): 209-222

It enables scholars to explore and immerse themselves in an environment that matches that of the creator, with a virtualised equivalent of the original folder structure and computer desktop arrangement in place and files available for examination[256]. It is even possible, depending on availability and condition, to set up this environment on the original computer with a new hard drive. Where a person retains system snapshots of the entire contents of a computer throughout its life, it will be possible for researchers to follow its evolution through time.

Often the originator's setup will be working properly and therefore usable but sometimes the original system will be awry and prone to malfunction; in which case some accountable tinkering or in depth manipulation of a replicate of the archival disk image may be necessary to produce an access version of the system.

There are some licensing issues that need to be explored. May the licence to use the software be transferred from the originator to the archival institution? Is it necessary to limit access to the creator's system (with its software) to one researcher at a time? There are many questions of this kind to be resolved.

Participation

Metadata creation

>> It must be a high priority to bring about the appropriate participation in metadata creation of creators and users of personal digital archives, as well as curators at other institutions. The involvement of creators, family members and scholars in various aspects of curation provides a way to empower individuals outside of the repository.

>> Even before the archive is destined to come to a repository, creators can take some responsibility for cataloguing, describing or annotating material in their digital archives.

>> One special concern is the identification of issues of privacy which can be greatly assisted by the involvement of the creators themselves.

>> Tagging is already happening in several contexts such as social networking (eg delicious, Flickr) and the creation of folksonomies. While these provide a motivating model and practice, it may be possible to design participation in ways that favour even more sophisticated metadata for little extra effort, especially with curatorial engagement.

---

[256] See J. L. John (2008) Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives and tools, iPRES 2008 Conference, the Fifth International Conference on Preservation of Digital Objects, the British Library, London, http://www.bl.uk/ipres2008, for discussion of the use of 'dd' files and Encase's Virtual File System and Physical Disk Emulator in this context

>> Creators can be motivated by: (i) a simple metadata scheme with layered meanings, (ii) easy-to-use, standardised metadata icons; (iii) free advice; and (iv) the privilege of acquisition by a secure repository[257].

Box: Metadata icons and notations for life

Complex systems have long inspired the creation of visual languages, iconic symbols and graphical notation or glyphs for concise and revealing representation of the system: from from genealogical networks documented by anthropologists to the convoluted traces of aerobatic maneouvres in the sky[258], from electrical circuit diagrams to architectural and engineering plans. More recently the constructs necessary for a visual language for describing, archiving and analysing aspects of biological systems have been devised[259]. There are several requirements in these efforts: (i) to initiate a set of icons or glyphs for conveniently representing entities, events and processes; (ii) to enable systems involving simple relations or complex interactions to be modelled by suitable arrangements of these icons or glyphs; (iii) to devise a visual language that can lead to new insights, explorations and discoveries as well as providing a means of documenting what is already established; and (iv) to make it possible for diagrams to be comprehensible to machines as well as humans.

A motivation for devising visual languages in biology has tended to be either (i) to enable pathways to be modelled and to explore hypotheses where pathways are uncertain[260], or (ii) to support biological engineering based on representations of biological parts and their behaviour[261]. In the context of the personal archive and historical events the outlook is largely retrospective (without precluding the testing of historical hypotheses and simulations), with icons providing a means for visualising unfolding sequences of events and communications and influences, and with an ontological foundation allowing queries to be submitted, with responses taking visual as well as verbal forms.

The approach is closely allied to the concepts of object oriented programming and ontologies, drawing in the benefits of visualisation and the possibilities of qualitative as well as quantitative description and analysis, while seeking to ensure interoperability, usability and network communication.

Faced with the wide ranging, complex and intricate nature of personal and professional digital lives, no single visual language can be compiled from the outset. This has been

[257] A. Charlesworth (2009) Digital lives >> legal & ethical issues, Digital Lives Research Paper, 14 October 2009, http://www.bl.uk/digital-lives/index.html

[258] Notably, the notation devised by the Spanish aviator José Luis de Aresti Aguirre (1919-2003) recognised as the standard for denoting the elements of aerobatic flight by the Fèdération Aeronautique Internationale, http://en.wikipedia.org/wiki/Aresti_Catalog

[259] D. L. Cook, J. F. Farley and S. J. Tapscott (2001) A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems, Genome Biology 2(4): e12; D. L. Cook, J. L. V. Mejino and C. Rosse (2004) Evolution of a foundational model of physiology: symbolic representation for functional bioinformatics, MedInfo 2004, pp 336-340

[260] D. L. Cook, J. C. Wiley and J. H. Gennari (2007) Chalkboard: ontology-based pathway modeling and qualitative inference of disease mechanisms, Pacific Symposium on Biocomputing 12: 16-27

[261] Y. Matsuoka, S. Ghosh and H. Kitano (2009) Consistent design schematics for biological systems: standardization of representation in biological engineering, Journal of the Royal Society Interface 6: S393-S404

recognised in the preparation of biological visualisation languages. As Daniel L. Cook and colleagues have put it: “A major criterion for the utility of any technical language is whether it can fully express concepts within a domain of knowledge; that is, is the language ‘complete’? For a domain such as biology which is almost limitless in terms of physical (not to mention psychic, social and evolutionary) phenomena, it is unlikely that any language will suffice for all purposes...”[262]. Thus the approach is to make a start with a conceptual framework, initial lexicon and graphical grammar, iteratively tested and modified, and anticipate subsequent modification and expansion. Likewise, the designers of the Systems Biology Graphical Notation indicate: “It was clear at the outset of SBGN development that it would be impossible to design a perfect and complete notation right from the beginning. Apart from the prescience this would require..., it would also likely need a vast language that most newcomers would shun as being too complex. Thus, the SBGN community followed an idea used in the development of other standards, i.e. stratify language development into levels”[263].

Instead, it would involve a combination of bringing together and unifying existing ontologies as part of a general archival system, by enhancing existing ontologies and by designing any necessary additional elements.

It is not simply a matter of securing an agreement on a collection or lexicon of icons to be adopted. It is necessary to devise expressive icons that can (i) represent functional entities consistently and clearly, (ii) be aggregated to form compound icons, and (iii) be extensible in other ways so that new icons can be derived naturally from existing ones[264].

There are of course icons that can be designed or adopted for archival practice itself such as an object’s status within an archive, the nature of the information medium (eg paper, magnetic, optical), privacy requirements, copyright requirements, dates; but the use of icons and associated ontological metadata can be extended to the incorporation and standardisation of icons for key life characteristics such as birth, death, leaving home, marriage, engagement and retirement, rites of passage, professions and employments and so on.

In addition to the need to devise icons that can represent events, interactions and entities and specifically archival and life metadata, there are the unique or nearly unique digital objects themselves. Some thought is necessary about the nature of icons representing these items, perhaps uniquely. Of interest are content-based icons that have been used for text, image, vector graphics, video and even music files - largely as a visual aid to accelerating the finding of files (although the icons may have familiar and aesthetic qualities too)[265]. In the case of the music files acoustic parameters of the waveform are captured by a computer

---

[262] D. L. Cook et al (2001), ibid

[263] S. Moodie, N. Le Novère, A. Sorokin, H. Mi and F. Schreiber (2009) Systems biology graphical notation: process description language Level 1, version 1.1, 2 September 2009

[264] D. L. Cook, J. F. Farley and S. J. Tapscott (2001) A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems, Genome Biology 2(4): e12

[265] J. P. Lewis, R. Rosenholtz, N. Fong and U. Neumann (2004) VisualIDs: automotive distinctive icons for desktop interfaces, ACM Transactions on Computer Graphics 23(3): 416-423; V. Setlur, C. Albrecht-Buehler, A. A. Gooch, S. Rossoffa and B. Gooch (2005) Semanticons: visual metaphors as file icons, Computer Graphics Forum 24(3): 647-656; P. Kolhoff, J. Preuß and J. Loviscach (2008) Content-based icons for music files, Computers & Graphics 32: 550-560

'neural network' which has been briefly trained to reflect the user's personal taste in depicting the music using the acoustic parameters; variations of an icon are generated, from which the user selects the preferred icon. If the creation of these (more or less) unique icons becomes commonplace in modern operating systems, personal archives will be populated by them; accordingly, it will be a matter of repositories retaining their original nature.

In a sense the approach is to enable and put in place the elements of an evolving visualisation system *a priori*, rather than later *a posteriori* or in an *ad hoc* way.

>> On registering with a repository or entering a repository, curators will work with the creator who must complete some fundamental metadata requirements to ensure accurate and rich data.

>> The use of ontologies akin to those used in bioinformatics and related disciplines did not receive strong support from curators and archivists participating in the Digital Lives workshops. This was due to a combination of: (i) the likely complexity of any ontology for personal digital archives (which after all potentially encounter everything under the sun and beyond); (ii) the rapidly changing requirements for such an ontology; and (iii) the limited resources available compared with genome, bioinformatic and medical discplines.

>> Nonetheless, it is strongly recommended that this option be examined further, with the possibility that steps be taken in the direction of forming appropriate, perhaps high level, ontologies. Universally agreed tag standards suitable for semantic interpretation are likely to be essential in the future. Hopefully, the creation of ontologies will be increasingly automated or augmented but nonetheless it seems sensible to make a modest start soon. The approach can be a gradual one that focusses on clearly bounded areas within the whole topic of personal archives.

>> There are exemplars for community annotation in other disciplines: for example the Sloan Digital Sky Survey is a consortium that involves the collation of information by many members who are formally recognised as coauthors of published surveys. This kind of approach and the use of annotation jamborees is deemed to be especially useful where there is insufficient funding for dedicated curators. It is an approach that can also be adopted for increasing the volumes processed even by institutions with significant curatorial support.

>> GalaxyZoo engaged 80,000 astronomers and members of the public to manually classify the morphology of one million galaxies in less than three weeks. While the documenting of people's lives is not the same as the classifying of galaxies there is a degree of transferability. For example, it might well be possible to involve people in seeking together to find links between their family histories, or in jointly seeking to identify person, time and place of photographs in a community or district.

>> Such collaborations might include professional and intellectual ones too: literary peers, scientific colleagues, politicians of various persuasions who are spending more time with their families, with annotation taking place at real world gatherings or more gradually but collaboratively online.

Scholars and experts

>> Experts can contribute their knowldege about historic individuals or artefacts but in order to get maximum benefit it may be necessary to find ways to reward experts for imparting their knowledge for the direct benefit of the scholarly and informational richness of the digital objects. It needs to be seen as part of the academic output of an individual in the same way that a scholarly book might be. It might be possible to develop a system that rewards scholars who assist in the curation with academic prestige and funding[266].

>> WikiProfessional Life Sciences provides an illustration of a linking of community curation with research and reputation gains[267].

Selection and context with participants

>> The expertise beyond the repository needs to be recognised as part of the archival process: (i) expertise of close associates of creators; (ii) scholarly expertise; and (iii) experienced diligence of interested but specialist enthusiasts. For example, interviews of creators could be conducted by appropriate researchers and users of the repository.
>> User and creator participation has the added benefit of countering any perception that a repository represents a closed culture.

>> Colleagues and friends of creators can be involved in setting up and carrying out recorded conversations with each other.

---

[266] D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, D. P. Hill, R. Kania, M. Schaeffer, S. St Pierre, S. Twigger, O. White and S. Y. Rhee (2008) Big data. The future of biocuration, Nature 455: 47-50

[267] http://www.wikiprofessional.org/

>> Users of personal archives could be judiciously engaged as well as creators themselves in the selection process of collecting institutions. Clearly there are matters of privacy and confidentiality that remain the reserve of curatorial function but it is possible to make suitable arrangements.

>> To ensure full accountability and transparency, contributors can be registered and briefly described in the corresponding metadata (eg known expert, interested enthusiast, family or friend), and all contributions labelled according to participant category, name and unique identifier.

10.8 Actions for Interaction Module

Repositories need to interact with other institutions for five reasons: (i) to agree standards and policies; (ii) to collaborate in major programmes especially in the integration of archival and personal content; (iii) to share responsibility, with some repositories specialising in some activities, eg as a centres of excellence to the benefit of all; (iv) to advocate any necessary actions jointly to the wider community; and (v) to provide services.

Collaborations and partnerships

Standards

>> There is a disconcerting – not to say confusing and confused – variety of approaches to personal data and privacy: not just among repositories but diverse bodies that deal with personal information in some way such as data centres – both nationally and internationally.

>> Formal collaborations should be put in place for national and international archives to agree and set standards. No doubt some of the existing archival representative organisations have roles to play in this activity.

>> There is a need for cataloguing or indexing and searching procedures to be interoperable and integrated across repositories and individuals for these multisite resources to be of most benefit.

>> Standards for interoperability and exchange to be agreed among repositories include ethical standards and metadata schemes and icons, as recommended by this project. Interoperability not only allows interaction of objects but also ensures that researchers find similar tools in different scholarly repositories.

>> A difficult but unavoidable area of contention is the formation of universal or quasi-universal authority files for people: ie unique identifiers for individuals. Many people are in any case adopting means of standardising their online identity: openID, for example[268]. Some researchers are participating in a researcher ID scheme that will among other things make it possible for all their publications to be quickly identified and correctly attributed to them[269]. Such standards would of course facilitate linkages among repositories and complementary resources.

Box: Virtual International Authority File[270]

The Virtual International Authority File (VIAF) is an international project involving the Library of Congress (LoC), the Deutsche Nationalbibliothek (dnb), the Bibliothèque nationale de France (BnF) and the OCLC (Online Computer Library Center), and seeks to facilitate research across languages and institutions internationally by integrating virtually the name authority files of these three libraries and other organisations. By the autumn of 2009, 15 organisations were participating in VIAF, bringing into the fold 18 personal name authority files. Once set up, the linked and matched authority records for personal names will be shared and maintained for access by users wishing to make use of this international capability.

The OCLC is also engaging in a project in collaboration with the University of Illinois at Urbana-Champaign and the University of Maryland which is directed at automated name identification and disambiguation from unstructured text, as part of the Extracting Metadata for Preservation (EMP) project[271].

It is worth bearing in mind in this context the emerging capability of the semantic web expressed by services such as Garlik[272] that is able to glean identity information on the web and combined with the use of controlled vocabularies and semantic integration of information from diverse sources, is able to provide subscribers with a view of their online identity. Moreover, the Friends of a Friend (FOAF) project[273] has instigated a semantic web vocabulary for people's names, ages and relationships to each other[274].

[268] http://openid.net, the website of the OpenID Foundation

[269] M. Enserink (2009) Are you ready to become a number? Science 323: 1662-1664; see also http://www.researcherid.com and http://isiwebofknowledge.com/researcherid

[270] R. Bennett, C. Hengel-Dittrich, E. T. O'Neill and B. B. Tillett (2006) VIAF (Virtual International Authority File): linking Die Deutsche Bibliothek and Library of Congress Name Authority Files, 31 August 2006, World Library and Information Congress, 72nd IFLA General Conference and Council, 20-24 August 2006, Seoul Korea; OCLC (2009) VIAF (The Virtual International Authority File), http://www.oclc.org/research/activities/viaf/default.htm

[271] OCLC (2009) Name Extraction, http://www.oclc.org/research/activities/nameextract/default.htm

[272] http://www.garlik.com

[273] http://www.foaf-project.org

[274] L. Feigenbaum, I. Herman, T. Hongsermeier, E. Neumann and S. Stephens (2007) The semantic web in action, Scientific American 297(6): 64-71, December 2007

>> It seems likely that as the volumes of information in repositories and personal archives grow, the need for unambiguous identity will likewise increase. If widely adopted, standards can reduce errors, make curation less time consuming and greatly increase research effectiveness.

>> It is sometimes supposed that standards must be fully defined, detailed and agreed before benefits can ensue from them but in fact a gradual approach is feasible and may be the best way to balance privacy concerns with efficiency gains.

Equipment development, especially software

>> Much of the commercial software that incorporates personal information management is directed at limited functions that cater for information in the present or immediate future.

>> Ways of influencing the development of software and hardware that serves an archival PIM, and the sustainable use of eMSS, need to be explored. The development of technologies directed at archival PIM would play a crucial role in motivating creators.

>> Also desirable would be software that facilitates or partially automates the processes of adding metadata (eg legal, contextual) and of collating archives that reside in fragmented ways across diverse locations.

>> Existing forensic and associated software might be modified specifically for archival purposes: eg remote authenticated acquisition of files and folders from the computer of a donor with evident permission and conspicuous visibility to the donor.

Box: Authentication, magnetic tapes and the SCSI standards[275]

Archivists commonly encounter magnetic tapes, notably in the archives of scientists. The acquisition of data on magnetic tapes in a forensically sound and detailed way is more difficult and limited compared with the same task conducted with magnetic disks. Tapes store files sequentially. The SCSI interface (Small Computer System Interface) was developed in the 1980s for connecting devices to host systems: the QIC tape format (Quarter Inch Cartridge) was popular and data cassettes (two reels) and cartridges (one reel) replaced open reel tapes. Common tape technologies include the 4 mm DDS (DAT, Digital Audio Tape), 8 mm

[275] B. J. Nikkel (2005) Forensic acquisition and analysis of magnetic tapes, Digital Investigation 2(1): 8-18

Exabyte and 0.5” DLT (Digital Linear Tape) relying on the SCSI standard and its variety of high level commands determining functionalities such as compression, bit densities and data block sizes.

One implication of this dependency on SCSI, as Nikkel explains, is that it is not possible to access the tracks and physical blocks of the data: instead SCSI provides for access to the logical blocks or records, and the ‘dd’ command of Unix favoured for forensics can be used to acquire these logical blocks but not - unlike with disks - the lower level physical blocks. Moreover, it is not possible with SCSI to access any surviving blocks of information beyond the End of Data marker (which exists after the last current file on the tape which stores the files sequentially along the length of the tape). Thus it is not a simple matter to conduct a physical acquisition of the low level information of the entire tape (in contrast to a magnetic disk). For this reason it has been suggested that a proposal for a suitable amendment be made to the SCSI standards committee. (The current situation partly reflects the simplicity of the naturally sequential file system of magnetic tapes which can be contrasted with the much greater complexity of magnetic disk file systems where the data making up a file often exists as fragments in various physical areas of the disk, calling for more sophisticated (and allowing faster) access and mapping.)

In the case of hard drives it is necessary to use a write blocker to prevent the capture system from altering any information in the original disk. With tapes there is a physical read-only tab or switch that can be used for this purpose (in some cases this does not represent direct physical protection but rather invokes firmware - which in principle could mean that write protection is bypassed). Hash values can be created for the individual files and for the entire acquisition. There is a tape log record that can be captured: eg for information about tape history, error counts, numbers of files and partitions; such information may exist either before the Logical Beginning of Tape or on an EEPROM chip (Electrically Erasable Programmable Read-Only Memory) in the cassette or cartridge, and can usually be requested *via* the SCSI interface.

It appears to be an example of how far reaching the impact can be of a standards committee in digital curation and preservation. Might the existence of file remnants beyond the End of Data marker - crucial data or drafts of influential essays of celebrated scientists or literary figures - help to motivate an amendment to the SCSI standard?

Avenues for communication

>> A number of possible forms of communication emerged from project discussions: a specialist forum, workshops, and conferences dedicated to aspects of personal digital archives (eg Digital Lives Research Conference).

>> Some of these activities can be better advanced by established archival institutions but there does seem to be a place for a more informal means of communication specifically about personal digital archives. In this regard informal communication can be supported by online chat functionality, web conferencing or virtual world venues.

>> The possibility of establishing one or more wiki sites (and related options for a participatory reference source) is frequently met with enthusiasm coupled with concern about sustainability. Potential benefits are widely known: being up-to-date, integrated, and capable of growing with the community's own knowledge. The two key issues are the minimising of effort required by individuals and, perhaps more importantly, the recognition of authors for their contributions[276].

>> A potentially exciting development that has gone some way to resolving the matter is WikiGenes which tracks authorship in detail, linking and giving due recognition for every contribution to its author, "creating the first hybrid of traditional, scientific and collaborative, dynamic publishing". The technology behind WikiGenes makes it possible for readers to gain another benefit from authorship, which is the ability to appraise the "origin, authority and reliability of information"; and for authors to rate each other on the basis of specific contributions[277].

Communities of practice

>> Communities of practice can be established with professional colleagues: international and cross-disciplinary.

>> There is a overwhelming rationale for collaboration between computer museum specialists and archivists, and their respective institutions: access to computer equipment and technical knowledge, for archivists; and access to an understanding of how individual computers were actually used and their social context, for computer museum curators.

>> Archives are required to match objectives to resources, and make clear choices. Some archival institutions may want to focus on analogue, some on digital, while others embrace both. Bearing in mind the special vulnerability of digital media, it would it be useful to have a system that promptly alerts archives and curators to materials that are available, most especially digital ones; such a system could be founded on an understood and agreed division of labour and distribution of the burdens of responsibility. A degree of flexibility might be built in so that institutions can evolve in their relationships and responsibilities.

---

[276] R. Hoffmann (2008) A wiki for the life sciences where authorship matters, Nature Genetics 40:1047-1051; see also http://www.wikigenes.org/

[277] This ability to identify specific contributions shows, incidentally, how even where content (of a website, say) is derived from multiple authors it can be feasible for portions of content to be attributed to an individual as a 'digital object'.

Principal ongoing partners

>> As already emphasised by this project, the full potential of personal digital archives will not be realised until there are substantive moves to form a comprehensive and digitally networked integration of archives and their contents.

>> Universities, publishers, and learned societies have several key roles: not least in demonstrating the high research value of personal digital archives, and the benefits that could ensue for society and individuals; and in making available further personal digital objects derived from creators who have interacted with these institutions.

>> Similarly funding agencies can support the assessment of the impact of curated data, and the development of innovative curation methods.

>> Online service providers could be encouraged (by creators as well as curatorial repositories): (i) to allow harvesting of digital content for archiving and future research; and (ii) to participate in the search for mutually beneficial synergies.

>> Public archival repositories could form partnerships with organisations and public bodies that champion privacy and individual rights, in order to ensure that standards of privacy and other rights are rigorously maintained. This joint system of testing and vetting for individual rights and archival suitability might be offered as (i) a service to commercial online service providers such as social networking sites, or (ii) as the basis of an effective mutually beneficial partnership.

Services

>> Small archival repositories often cannot afford a fulltime digital specialist or extensive equipment. Larger institutions might offer consulting and technical services to other repositories, to universities, to the third sector, to organisations in the less developed world, and to individuals.

>> Personal digital objects supply: digital facsimiles and replicates can be made available where digital rights and data protection permit.

>> Authentication and provenance services might be offered to individuals, as already suggested by this project[278]. Digital objects could be registered concomitantly even if not actually held by a repository. Storage services will also be an option for some repositories, perhaps suited to specific communities of users such as writers.

>> Repositories could also provide – where helpful – mediated access to personal content that has been anonymised by the repository itself which is therefore able to vouch for its authenticity.

>> Other services might include visualisation tools facilitating exploratory research[279].

10.9 Actions for Personal Informatics Module

Personal informatics is the study of all aspects of personal information and the curation of personal digital archives.

Digital personal archives are still an emerging resource and there are many aspects of their curation - by individuals and repositories - that require new tools and workflows and ongoing research and development. Areas of research for the future have been outlined in §9, and tools have been considered in §4. It is useful, however, to briefly express an overall approach to the practicalities of personal informatics.

A natural place to start such an approach is by considering the Planets Programme and its theme of interoperable and networked tools and services integrated within a unifying framework. Planets is directed at digital preservation. An archival framework could be modelled on the digital preservation framework or it could be an extension of it: Services, Tools and Archival Research Systems.

From the perspective of personal digital archives there are considerations besides digital preservation that could benefit from having (i) a planning tool, (ii) a core registry of existing and emerging tools, (iii) access to a series of actionable tools and services, and finally (iv) a testbed with which to test rigorously and repeatedly aspects of archival science beyond digital preservation.

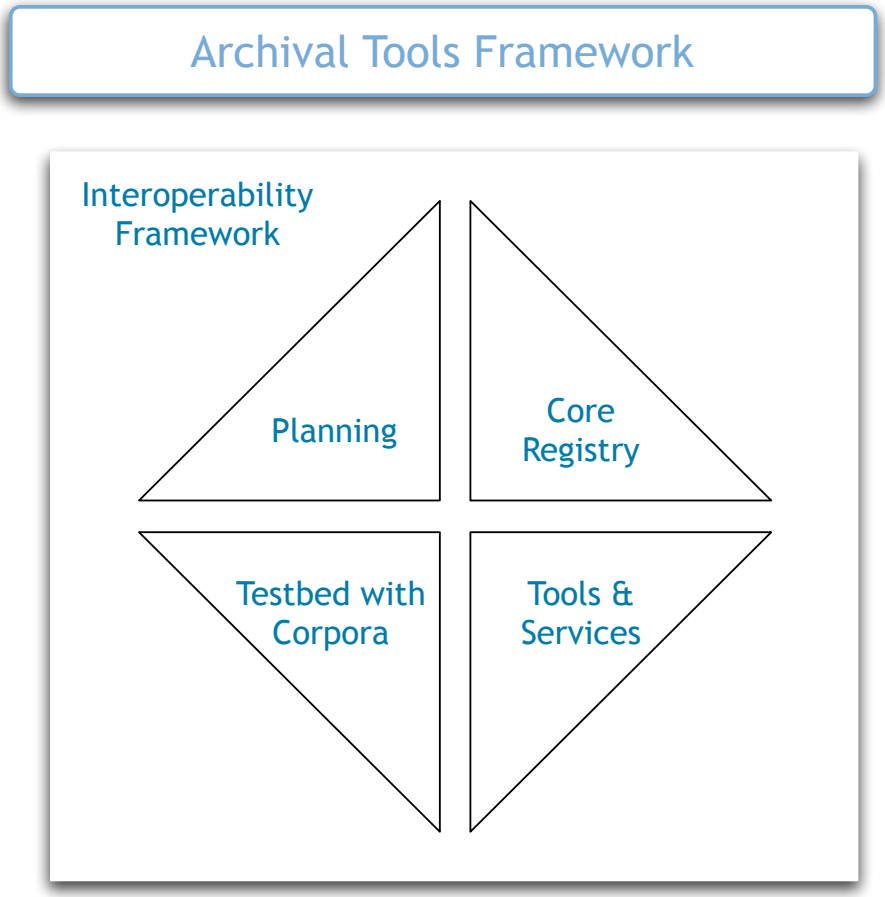
The testbed has an associated corpus or corpora of digital objects suitable for testing with regard to processes such as characterisation, migration and emulation. Equally, it would be useful for digital archivists to have access to a corpus of personal digital objects with which to test automated and semi-automated cataloguing tools, to pass a digital photo through a

[278] J. L. John (2009) The future of saving our past, Nature 459: 775-776

[279] M. McKeon (2009) Harnessing the web information ecosystem with wiki-based visualization dashboards, IEEE Transactions on Visualization and Computer Graphics 15(6): 1081-1088; F. B. Viégas, M. Wattenberg, F. van Ham, J. Krisse, M. McKeon (2008) Many Eyes: a site for visualization at internet scaLe; C. M. Danis, F. B. Viégas, M. Wattenberg and J. Kriss (2008) Your place or mine? Visualization as a community component, ACM CHI 2008, 5-10 April 2008, Florence, Italy

detector of digital tampering, to try out a new visualisation tool using a set of digital objects, or perhaps to subject some objects to encryption and decryption processes. Most especially it would allow archival science[280] to be truly scientific. The importance of having a standardised corpora of this kind has been well explained by Simson Garfinkel and colleagues in the context of computer forensics[281].

The archival planning tool would embrace enhanced curation activities such as video walks, oral history interviews, landscape photography and so on. The registry would include information about cameras, audio recorders, graphical software, various cataloguing packages and so on. Using the archival testbed a curator would be able to test a number of open source indexing and annotation tools for colleagues or a dashboard interface for researchers. A cataloguer could check to see if the metadata being produced meets the required standard. There would be a series of workflows for planning site visits, quantities of consumables required, and for managing software and other licenses.



graphic: digital lives: jeremy leighton john

adapted from context of digital preservation: Planets Programme

[280] The phrase 'archival science' is adopted by Seamus Ross, for example, in a keynote conference paper that reflects on the history of archival practices with the emergence of diplomatics, S. Ross (2007) Digital preservation, archival science and methodological foundations for digital libraries, 11th European Conference on Digital Libraries (ECDL 2007), Budapest

[281] S. Garfinkel, P. Farrell, V. Roussev and G. Dinolt (2009) Bringing science to digital forensics with standardized forensic corpora, Digital Investigation 6: S2-S11

Tools and workflows

Requirements

>> There are essentially three routes to obtaining a tool: find an existing one; amend an existing one; and design one.

>> First, however, it is necessary to identify the requirements, the necessary functionality.

>> It is helpful to design a modular system with a degree of compartmentalisation of the diverse functions. As suitable new technologies emerge these can replace existing ones in a module without unduly disrupting other modules.

Remodelling and testing

>> Current digital preservation and archival policies and processes are not necessarily optimally suited to personal digital archives.

>> Adapt planning and costing tools specifically for personal digital archives.

>> Explore scale and scalable solutions.

Awareness

>> Already there are some inexpensive digital preservation and forensic capture tools available. Awareness of existing tools, and of new ones as these emerge, and the way that they can help to ensure integrity and authenticity needs to be improved.

>> Compile and maintain a registry for existing tools and workflows, with functionality details.

>> Monitor of technology alerts, shared with community

Design and qualify

>> Enable design and development of tools for personal archives suitable for small archives as well as large ones.

>> Design straightforward and pragmatic workflows that can be scaled up or down. Many aspects of the personal curation lifecycle can be adapted from preservation workflows but many curation activities, eg enhanced curation, require brand new workflows.

>> Similarly, identify, test and design or adapt tools and services for researchers, the users of the repository.

>> Some functionality would be useful specifically for creators too. For example, it might be helpful if preservation software would automatically recognise the media format and recommend timely migration and transfer.

Research and development

Original research

The future research directions discussed in §9 included:

(1) Authenticity and Forensics
(2) Archival PIM
(3) Usability
(4) Infoethics and Digital Value
(5) Advanced Cataloguing and Context
(6) Evolutionary Dynamics, Networks, Phylogenetics and Digital Stemmatics
(7) Future Access, Visualisation and New Research Techniques
(8) Adaptive Curatorial Systems.

There are two further research activities that belong in a class of their own, representing as they do the shared intellectual environment and experience of the creator, curator and consumer: (9) on the one hand there is the understanding and experience of information technology and personal lives from the perspectives of past, present and future, referred to as Continuity and Change; and (10) there is the scholarly and scientific output that emerges from the entire process, an understanding of the content of and research into personal digital archives, referred to as Digital Scholarship and Life Information Research.

All ten research activities need to be shared and disseminated.

Dissemination

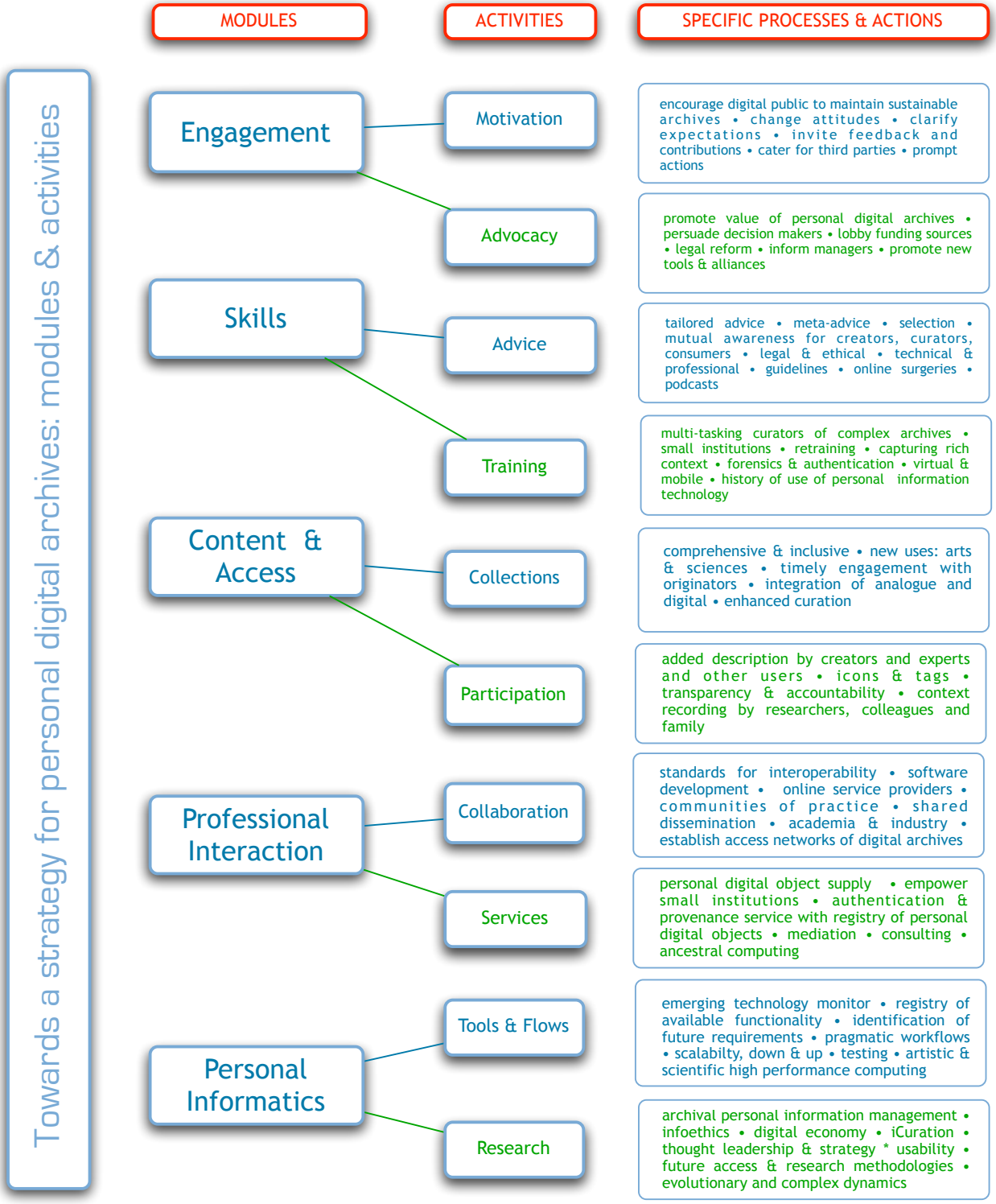
>> An online venue, portal or hub would allow the communication and discussion of research findings and ideas: eg newsletters and announcements, tools, guidelines, curatorial wikis, online meetings, chat sessions.

>> An interesting example of a community research system is nanoHub based at Purdue University, Indiana, USA, combining community annotation with a research community calendar, datasets and research tools, teaching materials and chat rooms, all in a one-stop location: "The site is built around a community-contributed series of nanoscience simulation, analysis and visualisation tools, which can be interactively run online on top of a compute farm maintained by Purdue University. These tools can be freely shared, combined in interesting ways, incorporated into online publications, and rated and tagged by community members. The site has 500 active contributors per year and serves requests from 60,000 users per year"[282].

>> A portal or hub can also be used for facilitating the collaborative and partnership activities, mentioned in §10.8: (i) announcing and registering emerging archives; (ii) ongoing gap analysis and opportunities regarding subject areas and themes that require more representation as personal digital archives; and (iii) seeking other archival requirements.

>> Research programmes should offer thought leadership and strategic as well as tactical analysis.

[282] A 'hub' is described by HUBzero as a web-based collaboration environment (i) founded on a series of well known open source packages such as Apache web server with LDAP for user logins, PHP web scripting, Joomla content management system, and a MySQL database, and providing (ii) interactive simulation tools, (iii) online presentations, (iv) mechanisms for uploading new resources, (v) a tool development area, (vi) content tagging, (vii) wikis and blogs, (viii) user groups for private collaboration, (ix) a user support area, (x) usage metrics, (xi) a place for news and announcements about events, and (xii) feedback mechanisms; http://hubzero.org/tour/features


Towards a strategy for personal digital archives: modules & activities
MODULES
ACTIVITIES
SPECIFIC PROCESSES & ACTIONS
Engagement
Motivation
encourage digital public to maintain sustainable archives • change attitudes • clarify expectations • invite feedback and contributions • cater for third parties • prompt actions
Advocacy
promote value of personal digital archives • persuade decision makers • lobby funding sources • legal reform • inform managers • promote new tools & alliances
Skills
Advice
tailored advice • meta-advice • selection • mutual awareness for creators, curators, consumers • legal & ethical • technical & professional • guidelines • online surgeries • podcasts
Training
multi-tasking curators of complex archives • small institutions • retraining • capturing rich context • forensics & authentication • virtual & mobile • history of use of personal information technology
Content & Access
Collections
comprehensive & inclusive • new uses: arts & sciences • timely engagement with originators • integration of analogue and digital • enhanced curation
Participation
added description by creators and experts and other users • icons & tags • transparency & accountability • context recording by researchers, colleagues and family
Professional Interaction
Collaboration
standards for interoperability • software development • online service providers • communities of practice • shared dissemination • academia & industry • establish access networks of digital archives
Services
personal digital object supply • empower small institutions • authentication & provenance service with registry of personal digital objects • mediation • consulting • ancestral computing
Personal Informatics
Tools & Flows
emerging technology monitor • registry of available functionality • identification of future requirements • pragmatic workflows • scalabilty, down & up • testing • artistic & scientific high performance computing
Research
archival personal information management • infoethics • digital economy • iCuration • thought leadership & strategy * usability • future access & research methodologies • evolutionary and complex dynamics
graphic: digital lives: jeremy leighton john

Chapter 11: Overview

11.1 Concluding Rationale

(1) Personal archives have greater research potential than ever before, for serving the individuals and families that create them as well as for endowing both scholarly and scientific communities with primary information.

(2) Comprehensive and widespread life information offers considerable benefits to medical, social and natural science.

(3) Personal archives are indispensable for the study of creativity and literary endeavour in following and mapping the workings and writings of inventive and original minds; and - in embracing as far as feasible people generally - archival repositories would allow the varied nature of creativity to be examined and compared, along with environmental circumstances.

(4) Among the raw materials for history, personal archives are supreme, allowing careful explications and interpretations of past events with independent historical witnesses.

(5) The phenomena of lifetracking, personalised medicine, context aware ubiquitous computing, personalised usability, biometric security and individual digital portfolios suggest that people will increasingly require a sustainable, dynamic and sophisticated resource of personal information and life history.

(6) Many people welcome and even psychologically depend upon family and personal memory, a testifying of individual creativity, a keeping of sentimental objects, and a securing of sensitive data. For a wide variety of reasons people appreciate the ability to hold, share, pass on and control personally created and acquired information and artifacts.

(7) Everyone should be encouraged to maintain a personal digital archive as soon as possible, and it is anticipated that many more people will seek to do so in the not too distant future.

(8) Personal digital archives may be held by individuals and by repositories. With the 'living' archive, individuals need to keep most or all of their personal archive to hand, and so it seems that a personal archive will either be maintained by an originator and his or her family alone or will be maintained both by the originator and the repository.

(9) The Digital Lives project has advocated a consistency in the methods and life cycle approach of professional archivists and of members of the public holding archives in the wild. This will help to ensure a degree of interoperability and a commonly shared language and understanding.

(10) Repositories can offer all individuals standardised guidance, interactive advice, and tools for ensuring the sustainability of personal digital archives and the digital objects within them.

(11) Depending on the precise nature and capabilities of emerging storage and digital library and archive technologies, repositories can also offer longterm storage for personal digital objects to some or many individuals. It seems likely that people generally will continue to accept and support the secure holding of the personal archives of the most creatively productive and influential luminaries by public repositories, although the precise criteria for

selection will always be debated. Few people will dispute a need to provide an additional layer of security or even a bespoke system for the archival materials that lie behind the greatest art and science, the most outstandingly influential events and creative acts - locally, regionally, nationally or globally.

(12) It is impossible to anticipate exactly how far public archival repositories will or should go in actually storing the personal digital objects of the digital populace. It is desirable that some representative archives should be collected across the social and cultural spectrum.

(13) In any case, if personal digital objects, eMSS, are as useful and desirable as is supposed, they have to be stored somewhere, and the growth of commercial providers of digital archival services is a distinct possibility. Depending again on the evolution of future technology and socioeconomic and sociopolitical circumstances, objects may be stored additionally or solely at remote places, away from an individual's home environment. These locales would presumably be effectively repositories: commercial, nongovernmental or public, or some combination of these possibilities.

(14) In the immediate and foreseeable future it is strongly recommended by the Digital Lives project that people hold replicates of their eMSS locally under their own control (whatever else they do).

(15) For the research and individual benefits of personal digital objects to be realised, there will need to be a coherent and professional standard of archival safe-keeping and secure access, as well as collaboration and coordination forming a network of archival-standard repositories.

(16) It would seem that the sooner that commercial, nongovernmental and public organisations begin to collaborate and agree standards the better. A key agreement that might be put in place would be that if a commercial repository fails or wishes to dispose of its collections, the archival materials would be automatically offered to another repository, most especially those that have been granted a long term remit; and agreement to this option would be factored into the original agreement with depositing individuals.

(17) Even beside storing objects of individuals, there are key roles that public longterm repositories can play, in addition to providing advice and guidance. These repositories could offer an authentication service, and in particular could receive digital objects *via* an automated reception, identify file characteristics such as type and size, create hash values (unique fingerprints) for them and compile other provenance supporting information even if not retaining the objects. This would allow the repository to corroborate at a later date an object's prior existence. In effect this would be a registry for personal digital objects, each of which would correspondingly be allocated a persistent identifier, akin to the Digital Object Identifier, and perhaps a handle, too, embedded within the personal digital object itself.

(18) Repositories might play a role in enabling and mediating access to the eMSS, through selective access, encryption and anonymisation as required and permitted.

(19) A further step would be for repositories to monitor and record qualitatively and quantitatively, through tracking and comparable technologies, the use made of unique personal digital objects. Permitted users of the eMSS such as a research programme would acknowledge and thank individuals for their contribution to the research, with news on progress made by the programme.

(20) At another level of involvement, it might - depending on the future cultural and legal environment - prove to be feasible in time for individuals to be rewarded in other ways for the use of their personal information, perhaps through other services, or through monetary equivalents. This would be analogous to a Personal Information Trust, or to the way the Digital Object Identifier system has been used by commercial and other publishers.

(21) Applied to eMSS, these kinds of systems would clearly require sensitive and rigorous checks and balances, and would only be viable with the full support of people generally. The enactment and enforcement of workable laws that grant to each individual ownership rights and a significant degree of control of their personal information might be helpful. Other possibilities will no doubt emerge. Whatever the solution, individuals would need to have freedom to exercise choice, in participating, in deciding what to keep and what to delete, and what to share. (Conceivably people might be informed, or granted some notion, of when and how their personal objects and content are being used.)

(22) Ten years into the twenty-first century and relatively little curation of personal digital archives has taken place. In the UK several institutions are active: the British Library, the Bodleian Library, the National Library of Wales and the Wellcome Library spring to mind. Some important work is being undertaken by local archives such as the county archives of Hampshire and of West Yorkshire[283]. Although important progress has been and is being made, quantitatively the volumes are relatively modest. This is more or less the situation in continental Europe, Africa, America, Asia, Australasia and the Pacific even though variation is not unexpected and exists.

(23) There is a need for a combination of continuing development and consolidation of processes and a vastly greater building of capacity and volumes actually processed by repositories across each country and continent.

(24) For this to take place requires a rapid and vigorous training of archivists, both young and experienced: in the basics of practical digital capture and digital preservation (from migration to emulation) and enhanced curation. The goal should be for many archivists to undertake the digital curation of many archives, rather than a few institutions attempting to do it almost alone.

(25) The major repositories may provide - where it is wanted - guidance, advice, tools, workflows and training as has already begun to happen.

(26) With more and more people embracing digital technologies, the overall population of personal digital objects will grow, although in the absence of a very rapid change in circumstances many of the longstanding eMSS (marooned on early floppy disks for example) may be lost.

(27) With some technical advancement and popular awareness it should be possible to ensure that people are able to retain personal objects for the foreseeable future (the longterm remains an imponderable). Such is the potential value of eMSS to many individuals and their families that it seems very likely that sooner or later the market will respond to the need for convenient and longer lasting personal information archival systems: steadily and gradually

---

[283] See also W. Kilbride and M. Todd (2010) The digital preservation roadshow 2009-10: the incomplete diaries of optimistic travellers, Ariadne 62, January 2010

improving systems that allow individuals to pass chosen personal objects from generation to generation. Arguably the market already is responding with the likes of the Time Machine from Apple, except for the need for greater digital preservation measures to be incorporated.

(28) At the present time, it is not possible to anticipate in the near future, public and typically not-for-profit repositories being able - without significant changes in policies and processes - to house personal archives and content of people generally on a vast scale.

(29) Yet, much personal information and numerous digital assets is being held in increasing quantities in the cloud by online service providers. The ultimate fate of these personal objects is hard to predict. On the other hand the major public repositories specialise in taking the long view. As at least one academic paper[284] has suggested, it may be that an online service provider ceases to exist and the contents are offered to an institution such as the Library of Congress or the Internet Archive. It is worth recalling that many of the foundation and major collections of public institutions were originally derived from independent collectors.

(30) With widespread capability for handling and preserving eMSS the public repositories would be well placed (one hopes) to accept the user generated content and objects if these ever become available. But there is no guarantee at this time that the contents would be offered; nor is it clear to which organisations any offers would be made.

(31) It brings the rationale back to advocacy and motivation. In many ways the most important requirement is that of emphatically and empirically demonstrating the research value of personal digital objects, to literary scholarship, to science, to social policy. The key to the longterm is to direct efforts at ensuring that people are able to look after the digital objects for themselves to a significant extent, with perhaps trusted repositories able - where individuals offer their support - to facilitate appropriate, selective and occasional access by *bona fide* researchers.

(32) There are three essential priorities: (i) advocacy and policy exploration and change; (ii) practical experience in digital capture and preservation by repositories; and (iii) empowering motivation and guidance for people with archives in the wild.

(33) Infoethics is the underlying consideration. Yet personal digital archives will not be alone in invoking ethical concerns. Individuals and society are going to be confronting and hopefully resolving many complex ethical issues anyway: gene technology, human enhancement, pervasive computing, surveillance, bionics, all of these issues are not far away if not already pressing today. Indeed one of the most helpful ways to deal with these issues will be through a better understanding of human behaviours and needs and the impacts of new technologies and policies on people.

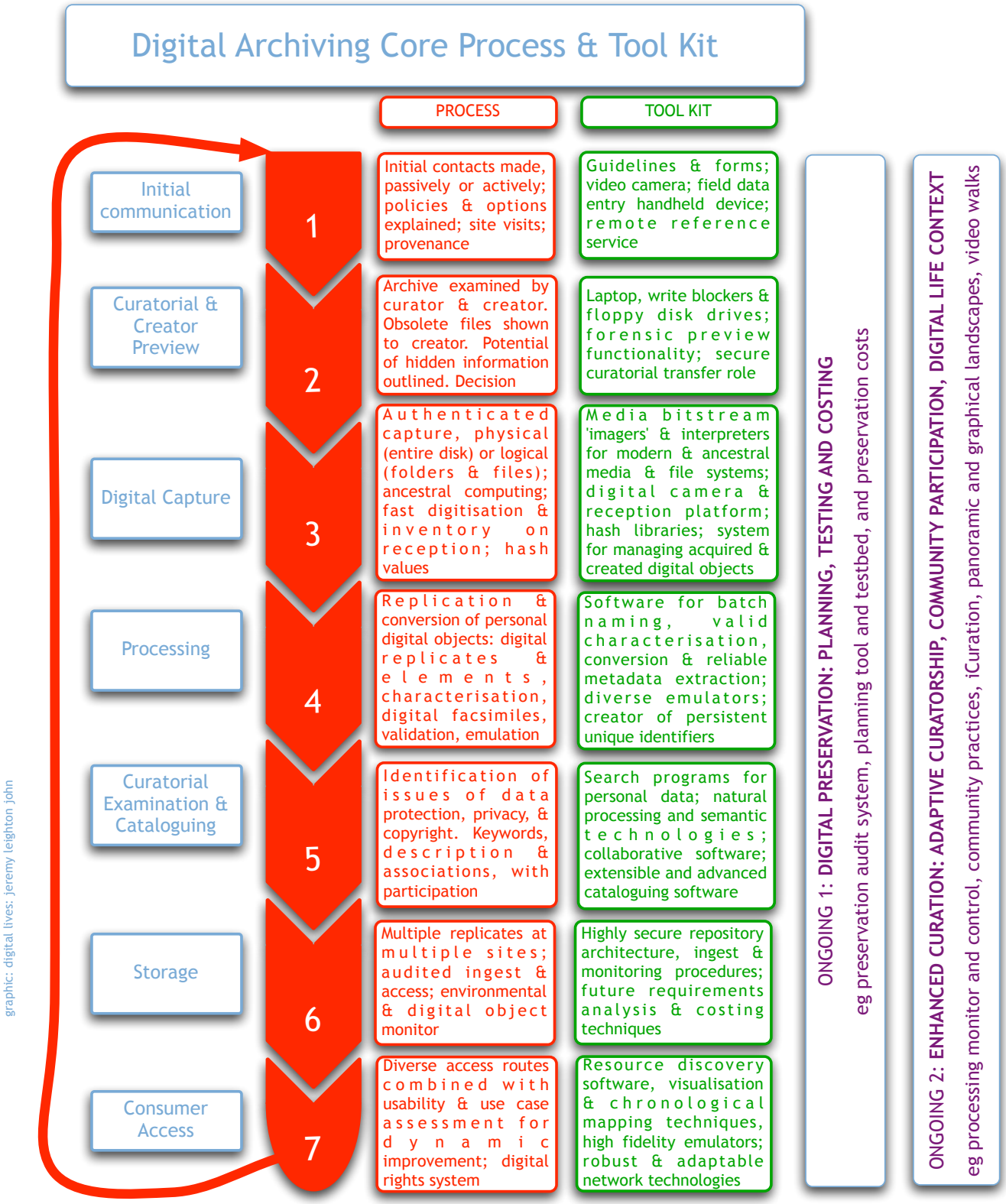
(34) These matters will compel the attention of everyone in the future. Today the first priority for repositories is the practical experience of digital curation and preservation of personal digital archives of writers, scientists, artists and social reformers.

---

[284] F. McCown and M. L. Nelson (2009) What happens when Facebook is gone?, 9th ACM/IEEE-CS Joint Conference on Digital Libraries, ACM, Austin, Texas, pp 251-254

11.2 Seven Steps in Contemporary Archiving, with Exemplar Tools

This section presents an illustration of the process undertaken by a digital curator or archivist. It is in a sense a portrait for the uninitiated. It is a slightly idealised scenario in that it omits the varying details that arise; but - if not quite 'a day in the life' - it does serve to provide a portayal of some days in the life of a curator as the profession of digital archiving is currently emerging. It is not intended as a formal workflow as such.



STEP 1: Initial communication

>> Active initiation by curatorial team or passive reception of enquiry

>> Website with contact details, along with collection development and advisory policies

STEP 2a: Examination of Archive, offsite and/or onsite

Preview modern media most specifically media that are no longer in active use (eg retired backup hard drives)

>> Preview examination (ie without acquisition and guarding against changes in dates and other metadata) of contemporary eMSS (typically derived from modern systems such as Microsoft Windows, Apple Macintosh, Linux and Unix)[285]

>> Forensic laptop with write blockers

Preview obsolete media (eg 3.5" or 5.25" floppy disks in the bottom drawer)

>> Preview examination of obsolete eMSS derived from ancestral computer systems (eg early Microsoft, Apple II, CP/M). Preview may entail temporary acquisition

>> Ideally modern laptop with appropriate reader for the computer media (eg 3.5" floppy disk drive connected by a modern interface such as USB). Viewers, readers with low fidelity may suffice (eg raw text). Desktop may be necessary in many cases (eg fitted with reconfigurable floppy disk controller such as Catweasel)

STEP 2b: Explain to creator the pros and cons of physical or logical acquisition

>> Establish wishes of creators and originators and confirm permissions

>> Open forms, deposit agreements and licences

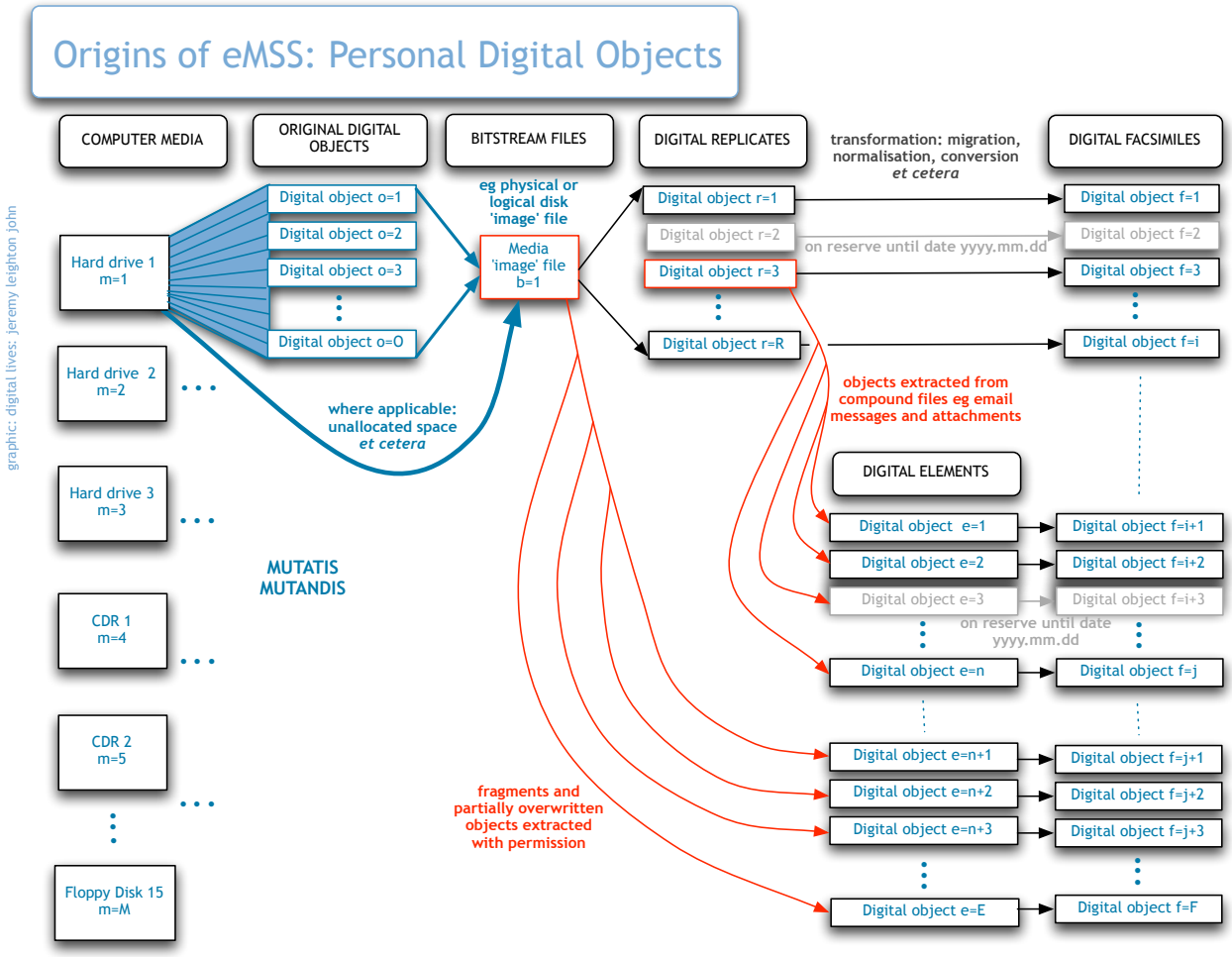
STEP 2c: Curatorial report to institution followed by decision and any transfer of media and hardware to repository

>> Secure and preservation protective packing and transfer

>> Transportation, with curatorial pickup or, in some circumstances, with highly trusted preservation-compliant courier service

---

[285] "Even for the IT savvy it can take a while to orientate oneself on an unfamiliar computer often while holding a conversation with office staff. For this reason, it is essential to gather as much information on the software and hardware being used by the depositor at the survey stage", S. Thomas and J. Martin (2006) Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives, Journal of the Society of Archivists, 27(1): 29-56



STEP 3: Acquisition with digital capture using forensically sound techniques, retained in holding repository

>> Forensic capture using write blockers and with hash values created. Virus, malware checking

>> For open, nonproprietary, 'image' files, the software options include Forensic Toolkit Imager and Advanced Forensic Format. Repeat using a second system of software and write blockers to ensure same hash values are obtained. Tableau write blockers are well known. Specialist disk 'imaging' software for systems such as Microsoft DOS, Amiga, Atari and other ancestral computers (eg ImageTool with Catweasel; Imgtool; Disk2FDI; RaWrite2 and RaWrite for Windows; WinImage)[286]. Some tools make it possible to convert from one media 'image' to another but such conversions should be subject to validation in the case of collection 'image' files

[286] http://mess.redump.net/reference:imgtool; http://disk2fdi.joguin.com/featreq.html; http://www.winimage.com; see http://www.fdos.org/ripcord/rawrite and http://en.wikipedia.org/wiki/Disk_image

STEP 4a: Digital replicates

Institute the digital replicates and extract any necessary digital elements

>> Export the replicate files from the authentic 'image' file

>> Various forensic software and other systems including 'imaging' tools

Characterise and identify the digital replicates

>> Identify and characterise file formats and the nature of the digital replicates

>> Planets will offer a convenient framework[287]: along with DROID that obtains file signatures from Pronom there is XCDL from Planets itself. The approach can be complemented and partly cross checked with the information yielded by forensic systems and - in particular - the use of hash libraries.

Extract metadata and properties

>> Extract core metadata such as times and dates. Much of the metadata can simply be exported from the forensic system along with unique hash values (digital fingerprints), and imported into a cataloguing system

>> This can be done with forensic and other software and services, eg Pronom, DROID, JHOVE & JHOVE2 and National Library of New Zealand Metadata Extractor through Planets. Furthermore, there is the powerful XCL newly minted by the Planets project: software uses XCEL definitions to extract XCDL characteristics

---

Box: XC*L or XCL

The eXtensible Characterisation Languages of XCEL and XCDL (known together as XCL) represent a major outcome of the Planets project. A key distinction from other XML applications in digital preservation is that it does not store purely preservation metadata; nor does it seek to migrate a digital object in its entirety to XML. Instead XCL transforms the object into an abstract unified structure.

The XCEL extraction language is designed to be used to create XCEL documents for diverse kinds of files; these documents identify and outline the information that can be extracted from files of a particular kind. An XCEL document for TIFF, for example, can be used by XCEL processing software known as 'Extractor' to process a particular TIFF file: extracting the extractable information and expressing the information in the file in XCDL definition language as an XCDL document for that specific TIFF file. This XCDL document can in turn be processed by software that serves as an XCDL interpreter. The XCDL document essentially expresses the

[287] In view of the personal and archival nature of the eMSS, archival repositories may wish to install the Planets system locally within the repository - for the immediate future

informational content of the original digital object in a way that is independent of the format type.

The beauty of the system is that it is not necessary to write a separate extractor for every file type. The XCL Extractor is designed to process any file type based on the pertinent XCEL document; and although these documents need to be written for each file format (a good number have already been completed) this writing is done in the universal and open language of XCEL.

STEP 4b: Digital facsimiles

Create the digital facsimiles

>> Convert the digital replicate into an interoperable form, the digital facsimile (ie copy and convert, retaining the digital replicate too). Batch naming and conversion. Document conversion process for the object

>> It is anticipated that Planets will offer a convenient framework for integrating various conversion tools. If obsolete file format is not represented, the digital preservation community can be informed and new functionality recommended. In the meantime, seek other tools if necessary; eg from ancestral computing community. Batch naming and conversion (eg from Word document to archival PDF/A using JodConverter and Open Office or Adobe Acrobat Pro)[288].

Validate digital facsimiles

>> Validate the conversion and its longterm sustainability

>> Through Planets framework, as more XCEL documents are prepared. Otherwise file format validation through JHOVE directly

---

Box: Validation

One of the main functions of XCL is to enable the evaluation of migration quality and provide for the automatic validation of conversions. For example if an ODF document is converted into a PDF/A document, the XCDL document for the original ODF document can be compared with the XCDL document for the emerging PDF/A document using software known as the XCL Comparator: this software looks at and compares the values of properties based on certain

---

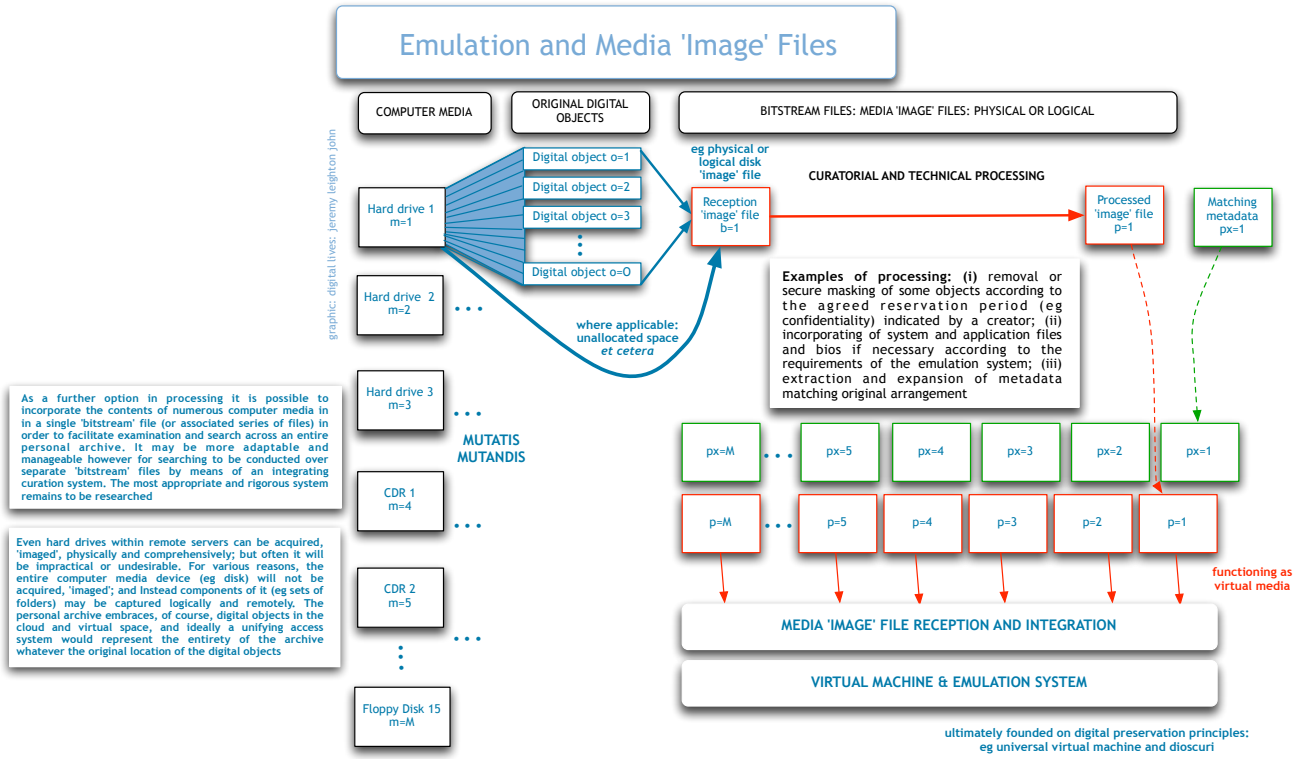
[288] As might be anticipated, there is no universally agreed list of recommended file formats for longterm preservation but there are some open lists currently available: The Florida Digital Archive, http://fclaweb.fcla.edu/fda_format_landing_page; Harvard University, http://hul.harvard.edu/ois/digpres/guidance.html; Library of Congress, http://www.digitalpreservation.gov/formats/intro/intro.html; State Archives of Antwerp, http://www.antwerpen.be/david; University of Texas, http://repositories.lib.utexas.edu/recommended_file_formats; Virginai Tech, http://etd.vt.edu/howto/accept.html. If nothing else these sites demonstrate the consistency and diversity of opinion. There is also a helpful presentation (available on the Planets website) by Manfred Thaller, Universität zu Köln, File formats and significant properties, 24 March 2009; it largely but not entirely follows the Florida Digital Archive listing

metrics. There is also the possibility for users to amend the weighting and selection of properties and metrics used by the Comparator in making comparisons, according to the goals of the user, eg repository.

STEP 4c: Produce access copies (of the replicates and facsimiles) and describe process

>> Characterise, validate and outline the duplication or conversion (eg lower resolution photo for remote delivery). This is important in order that the scholar or scientist knows exactly what is being delivered for examination

>> Through Planets framework



STEP 4d: Option for making available media ‘image’ file

>> Access can also be made possible through the media ‘image’ file directly, without exporting the digital replicates and creating digital facsimiles. Indeed this may be the most effective form of access. A copy of the primary media ‘image’ files can be modified or selectively encrypted in order that entities that require access restriction can be made unavailable. This processing may need to be carried out prior to cataloguing

>> Media ‘image’ editing or selective encrypting (eg in case of AFF) software, compatible with chosen emulation and virtual engine systems. Various

combinations are feasible: eg Xen and Qemu are open source. It may be anticipated that sustainable universal virtual machine technologies will be increasingly available and comprehensive: ie UVC

Box: Original and Access Virtual Paths

An important requirement is for curators to ensure that details of the original computing system used by the creator are captured and documented. Much of this information (hardware profiles, software employed, user preferences and past histories) is captured automatically when disks are forensically imaged (physically or logically), and hash libraries allow the software to be closely identified. It is still necessary to be able to succinctly document the way a creator used the computer for reference purposes.

A helpful concept that has arisen from IBM Netherlands is the Virtual Path (attributed to R. J. van Diessen)[289], which maps the combination of software and hardware that produces a specific rendering of the digital object. This procedure could be used to identify the Original Virtual Path: in reality there may be more than one original virtual path due to uncertainty, with confidence in the various options being quantified and qualified according to the archivists' determinations. Correspondingly, a particular digital object might have a set of possible Access Virtual Paths of varying authenticity, fidelity, utility and cost for the viewing experience: one virtual path might almost exactly represent the authentic original one but others may be more practical in some circumstances. Formal and detailed access virtual paths could provide scholars and scientists with a way to reference their viewing experience of a personal digital object in a consistent and repeatable way.

The concept of canonicalisation might be applied in order to consistently reference a definitive (even if not necessarily a high fidelity) viewing experience (extending the concept beyond the context of file format)[290]. A complication is that it is anticipated that curators will in some circumstances capture a personal archive successively (eg each year the archive is captured entirely, and not incrementally), which means that it is conceivable that identical digital objects (eg a Word document of an essay or play), captured on separate occasions but yielding the same hash value, might be associated with different original virtual paths (presumably scholars will be most - though not exclusively - interested in the original virtual path that was associated with the creation of the essay).

STEP 5: Curatorial examination and cataloguing

>> Archive examined for privacy and copyright issues. Cataloguing of the digital objects, compiled with extracted metadata

>> Curatorial examination computer, securely isolated and protected. In addition to the extracted

[289] D. von Suchodoletz and J. van der Hoeven (2009) Emulation: from digital artefact to remotely rendered environments, International Journal of Digital Curation 4(3): 146-155, also citing R. J. van Diessen (2002) Preservation requirements in a deposit system, Technical Report, IBM/KB Long-term Preservation Study

[290] Canonicalisation has been elaborated in the context of archived digital objects, C. Lynch (1999) Canonicalzation: a fundamental tool to facilitate preservation and management of digital information, D-Lib Magazine 5, September 1999, http://www.dlib.org/dlib/september99/09lynch.html, but the approach might equally be modified for use in referencing the rendering experience

metadata there are other metadata to be prepared, assessed and checked, partly by hand at this time[291]. Archivists' Toolkit provides an advancing open source cataloguing tool. EAD[292], EAC and forthcoming EAF together form a comprehensive metadata system for archives; Dublin Core is widely used.

STEP 6: Digital archival storage system

>> Pass digital objects and metadata to digital store

>> Many repositories are creating their own system of storage alone or in partnership. With continued and significant advancement, flexibility and authentication processes Hoopla might be an emerging exemplar for small repositories or individuals. A common standard for interoperability among memory institutions is the Protocol for Metadata Harvesting, OAI-PMH

STEP 7: Access including use of emulators

>> Digital objects and catalogue information made available according to rights constraints: from reading rooms to online

>> Open source systems range from more informal Drupal and Joomla to repository-dedicated DSpace, Fedora (now merged as DuraSpace[293]) and Greenstone with many of these products being founded on the open source MySQL, PostgreSQL; others will no doubt emerge before long. Emulators such as Dioscuri from Planets allow high fidelity and, where applicable, dynamic interactive access using digital replicates. Other tools will be emerging from KEEP project, and from ancestral computer enthusiasts, many being integrated by Planets

---

Box: Remote Emulation and Existing Emulators

As part of the Planets and KEEP projects, a remote emulation service is being developed founded on a demonstrator known as GRATE (Global Remote Access to Emulation services), which will automate many of the processes involved, making emulation accessible to many people and institutions *via* a web browser. One key advantage of making the service available remotely is that the expertise and effort required in developing and maintaining the

---

[291] It will be important to make clear to researchers the precise origin and nature of the extracted metadata

[292] C. Smith (2006) Encoded archival context: authority control for archives. A review of the literature, Info 633, Tech Processes in Libraries, submitted 15 May 2006; B. Stockting and F. Queyroux (2005) Encoding across frontiers: an introduction, Journal of Archival Organization 3(2/3): 1-8

[293] http://duraspace.org/about.php

emulation capability - which is highly demanding in resources - can be centralised with specially trained personnel[294].

Remote access raises an important issue for personal digital archives - in that digital objects that are to be rendered would generally be uploaded over the internet which raises security and privacy issues - but in principle it should be possible for a repository to use the system locally with support from the central service in those cases where the security of the internet is a concern, or to adopt in the future secure encryption on the fly. Ultimately this emulation system will be based on the Universal Virtual Computer and modular emulators akin to Dioscuri as well as a judicious combination of enhancements to existing emulators.

The emulators work typically by accessing media 'images', especially floppy disk 'images' (rather than physical disks or tapes), in an approach that is similar to that adopted for virtual forensic computing that commonly involves hard disk 'images'. The use of disk 'images' in emulation has been motivated largely by a need to provide the emulator with suitable software and information for it to be able to run effectively and to render accurately a digital object of interest (eg a Wordstar document), whereas the use of disk 'images' in virtual archival computing is strongly motivated by an additional desire to allow *a set of personal digital objects to be perceived together* in a way that was experienced by the originator of the archive. The large sizes of hard drives found in a personal archive further suggests benefits of local access to any rendering system along with any remote access to preservation and curation services over the internet. Nonetheless, there is clearly an effective and pleasing consistency and compatibility in the two emerging aspects of digital curation.

PARALLEL STEPS FOR ANALOGUE OBJECTS: Inventory on reception *et cetera*

>> Inventory listing and fast digitisation of analogue objects at 'file-level'

>> Digital camera connected to workstation with digital asset management software; photographic lighting; desktop photographic platform and background

ONGOING 1: Digital preservation

>> Systematic management and control with audits and chains of custody

>> Plato and Drambora provide tools for planing and audit. Life3 promises practical cost analysis. The foundation standard is the OAIS Reference Model[295]. Well established technical metadata are METS and PREMIS

---

[294] D. von Suchodoletz and J. van der Hoeven (2009) Emulation: from digital artefact to remotely rendered environments, International Journal of Digital Curation 4(3): 146-155. See also R. Verdegem, P. Bright, P. Wheatley, J. Schröder, J. van der Hoeven and D. von Suchodoletz (2008) Report describing results case studies, Deliverable PA/5-D3, Planets Project, 30 May 2008, 33 pp; R. Welte, D. von Suchodoletz, K. Rechert, R. Verdegem, J. van der Hoeven, M. van den Dobbelsteen (2006) First version of GRATE, Deliverable PA/5-D7, Planets Project

[295] For a clear explanation of OAIS from the perspective of personal archives, see S. Thomas and J. Martin (2006), Using the papers of contemporary British politicians as a testbed for the preservation of personal archives, Journal of the Society of Archivists, 27(1): 29-56

ONGOING 2: Enhanced curation

Digital life capture

>> Add value to collection items with contextual information onsite and online: panoramic photography, graphical 3D imagery, video tours, life stories, histories of personal manuscripts and other archival objects; personal library composition; life inventory and 2D & 3D imagery of personal artifacts and possessions; portable photographic lighting

>> Digital video camera and editing software; panoramic tripod head and digital camera with stitching and modeling software; portable bar code readers for recording personal library; handheld laser rangefinder allowing single curator to measure dimensions of rooms and external spaces; turntable for imagery of 3D artifacts

Community participation

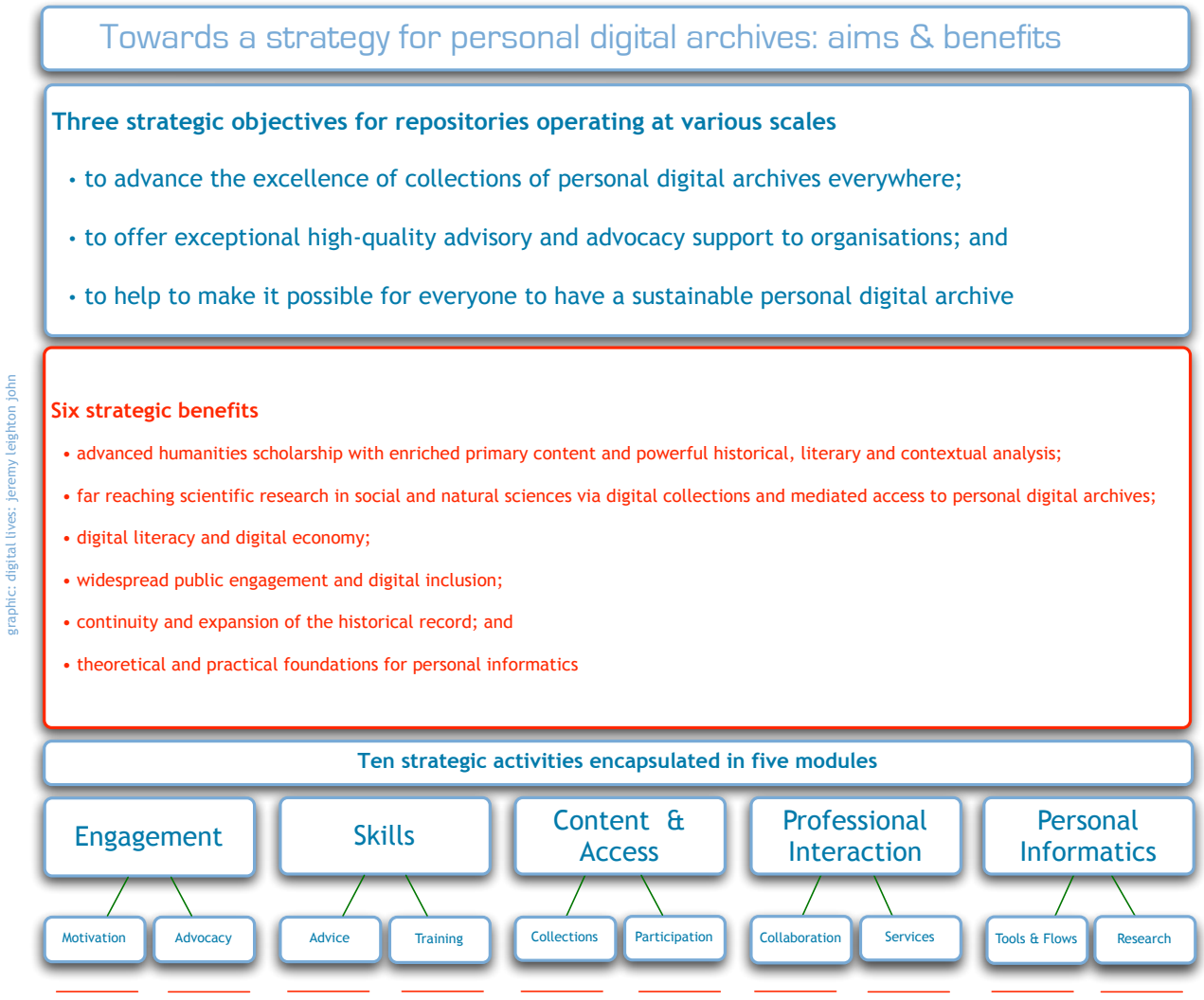
>> Make available to community of creators, curators and experts, and encourage provision of additional contextual information

>> HubZero offers an open source route for online collaboration, to be released in 2010. WikiGenes allows contributions that are attributable to individual authors

Adaptive curatorship

>> In the face of continuing and even intensifying change due to emerging technologies, business models and social and research expectations, it is necessary to develop an adaptive curatorship and stewardship. There is also a requirement for records and monitoring of curatorial processes and activities

>> Technology watch reports. Agile and evolutionary development with incremental iterative approaches. Ongoing and close involvement in research. Forensic software offers a model of an accountable system for recording curatorial actions



graphic: digital lives: jeremy leighton john

Chapter 12: Vision, in Brief

If there is a single message it is that there is, within our generation's grasp, an unprecedented opportunity for advancing humanity.

Personal archives can help us prevent diseases and find cures, they can help in the race to control climate change effectively, they can contribute to the battle against social deprivation and poverty, they will help future scholars and researchers to better understand creativity and how best to nuture it. Personal archives will help us to sustain and be sustained by literary and artistic achievement, they will help us nourish innovation and find solutions to the management of organisations and the attainment of community aspirations; they will help us deepen and broaden the historical record.

All of these things cannot be done without understanding people and their relationship with their cultural and natural environment, and it is personal archives that provide the avenue, the ultimate resource, for this purpose. The opportunity presented by personal digital archives and their successful curation depends on advocacy, research and pragmatic demonstration of their exponentially increasing usefulness for society, for research institutions and for individuals.

The potential benefit is not first and foremost in memory although that is fundamental and the challenge does not lie first and foremost in the technologies of preservation and capture although these are crucial. These are means to an end, and will be fruitless without the motivation of individuals and the will of decision makers. It is about utility, about the use of personal digital objects.

It is anticipated that the beneficial use of personal history, of life information, will become increasingly obvious, that archival memory is so valuable, so essential and so worthwhile and crucial for our future that it would be neglected at very great cost.

The root challenge for repositories stems from the fact that collecting and even simply handling and processing these digital objects will require extraordinarily sensitive and well judged policies and processes. It relates, of course, to ethics, privacy and rights of the individual. The possibility is being contemplated that in time and under suitable circumstances, numerous individuals might sustain and make available for mediated access by researchers portions of their personal archives, their digital lives.

Moreover, it is a challenge not only for public and not-for-profit archival repositories but for any commercial organisations that hope to play a role. These objects have to be stored somewhere and if there is a comprehensive shift to the cloud and away from relying purely on the local computer at home, then these objects will need to be stored in some kind of distant repository (commercial or otherwise) anyway.

In any event it is not simply in the context of repositories proffering research access to personal archives that the issue of personal information arises. Society is going to be confronted by this issue more and more as the power of digital technology increases and becomes (count on it) ever more essential.

Would ensuring that individuals have a high level of control and ownership of their personal information, determining extent and nature of access to their archives help to counter and

ameliorate the potential for technological misuse, officious surveillance and inappropriate data fusion? Perhaps. It remains to be seen.

In considering undeniable and even daunting hurdles we should not lose sight of the potential power for scientific and creative discovery provided by this resource of unfathomed value. Already personal information has been used to elucidate novel and unexpected connexions between genes within the newly determined genome databases, network and longitudinal analyses have yielded revealing social patterns that have informed policy, literary digital scholarship has captured for posterity influential dynamic works of art, and medical relationships have been uncovered between diseases and lifestyles.

Indeed, it may be - somewhat counterintuitively - that through researching and understanding personal informatics that the healthy balance between individual freedom and community priorities and shared ethics can be found.

Chapter 13: A Selected Bibliography - *Under Construction*

This bibliography is not intended to be comprehensive[296]. While it does include some definitive articles and papers, it does not focus on them. Instead its aim is to evoke - in an informal way - the field of personal informatics in its widest sense. The list is structured loosely according to the ten research topics outlined at the end of §10.

Note that there are many more references in the footnotes in this current version of the Synthesis. Later versions of this Digital Lives Synthesis will supply a more complete selection, suitable as as a starting place for reading lists and a syllabus.

13.1 Continuity and Change

Prehistory

d'Errico F (1998) Palaeolithic origins of artificial memory systems: an evolutionary perspective. In *Cognition and material culture: the archaeology of symbolic storage*, pp. 19-50 [C Renfrew and C Scarre, editors]. Cambridge: McDonald Institute Monographs.

d'Errico F, Henshilwood C, Lawson G, Vanhaeren M, Tillier A-M, Soressi M, Bresson F, Maureille B, Nowell A, Lakarra J, Backwell L & Julien M (2003) Archaeological evidence for the emergence of language, symbolism, and music - an alternative multidisciplinary perspective. *Journal of World Prehistory* **17**, 1-70.

Henshilwood CS (2007) Fully symbolic sapiens behaviour: innovation in the Middle Age at Blombos Cave, South Africa. In *Rethinking the human revolution: new behavioural and biological perspectives on the origin and dispersal of modern humans*, pp. 123-132 [P Mellars, K Boyle, O Bar-Yosef and C Stringer, editors]: McDonald Institute for Archaeological Research.

Jacobs Z & Roberts RG (2009) Human history written in stone and blood. In *American Scientist*, July-August 2009, pp. 302-309.

Gestures and the hand

Barbieri F, Buonocore A, Dalla Volta R & Gentilucci M (2009) How symbolic gestures and words interact with each other. *Brain & Language* **110**, 1-11.

Corballis MC (2009) Language as gesture. *Human Movement Science* **28**, 556-565.

Gentilucci M & Dalla Volta R (2008) When the hands speak. *Journal of Physiology - Paris* **102**, 21-30.

Goldin-Meadow S (2003) *Hearing gesture. How our hands help us think*. Cambridge, Massachusetts: The Belknap Press of the Harvard University Press.

Kellogg RT (1994) *The psychology of writing*. Oxford: Oxford University Press.

[296] It was compiled by Jeremy Leighton John with the exception of the section on Archival PIM which is based on a literature review by Peter Williams with Ian Rowlands and Jeremy Leighton John

Saffer D (2009) *Designing gestural interfaces*. Sebastopol, California: O'Reilly Media.

Wilson FR (1999) *The hand. How its use shapes the brain, language, and human culture*, Paperback ed. New York: Vintage Books.

Scribal history, and history in the margins

Beal P (1998, reprinted 2004) *In praise of scribes. Manuscripts and their makers in seventeenth-century England. The Lyell Lectures, Oxford 1995-1996*. Oxford: Clarendon Press.

Brown MP (1993) *A guide to western historical scripts from antiquity to 1600*. London: The British Library.

Camille M (1992) *Image on the edge. The margins of medieval art*. London: Reaktion Books.

Christin A-M (2002) *A history of writing from hieroglyph to multimedia (originally published in French, 2001)*. Paris: Flammarion.

Fairbank A (1970) *The story of handwriting. Origins and development*. London: Faber and Faber.

Whalley JI (1975) *Writing implements and accessories. From the Roman stylus to the typewriter*. Newton Abbot: David & Charles (Publishers).

History of computer and information technology

Anonymous (2001) *Aural history. Essays on recorded sound*. London: The British Library.

Bauer FL (2004) *Informatik. Führer durch die Ausstellung*. München: Deutsches Museum.

Berners-Lee T (2000) *Weaving the web. The original design and ultimate destiny of the world wide web*. New York: HarperCollins Publishers.

Burnet MM & Supnik RM (1996) Preserving computing's past: restoration and simulation. *Digital Technical Journal* **8**, 23-38.

Ceruzzi PE (2000) *A history of modern computing (Original publication 1998)*. Cambridge: The MIT Press.

Cortada JW (2000) *Before the computer. IBM, NCR, Burroughs, & Remington Rand & the industry they created 1865-1956 (originally published 1993)*. Princeton: Princeton University Press.

Finn CA (2002) *Artifacts. An archaeologist's year in Silicon Valley*. Cambridge, Massachusetts: The MIT Press.

Fischer CS (1992) *America calling. A social history of the telephone to 1940*. Berkeley: University of California Press.

Freiberger P & Swaine M (2000) *Fire in the valley. The making of the personal computer*, Second ed. New York: McGraw-Hill.

Hafner K & Lyon M (1998) *Where wizards stay up late. The origins of the internet (Originally published in 1996)*. New York: Touchstone.

Headrick DR (2000) *When information came of age. Technologies of knowledge in the age of reason and revolution, 1700-1850*. Oxford: Oxford University Press.

Nadeau M (2002) *Collectible microcomputers. A Schiffer Book for Collectors with Price Guide*. Atglen: Schiffer Publishing.

Naughton J (2000) *A brief history of the future. The origins of the internet (Originally published 1999)*. London: Phoenix.

Norman JM (2004) Collecting the history of computing, networking and telecommunications. In *The origins of cyberspace. Wednesday 23 February 2005. Christie's, New York*, pp. 2-8. New York: Christie's.

Digital life

Bell G & Gemmell J (2009) *Total Recall: How the e-Memory Revolution will Change Everything*. New York, Dutton.

Boellstorff T (2008) *Coming of age in Second Life. An anthropologist explores the virtually human*. Princeton: Princeton University Press.

Coleman AD (1998) *The digital evolution. Visual communication in the electronic age: essays, lectures and interviews 1967-1998*. Tucson: Nazraeli Press.

Keen A (2007) *The cult of the amateur*. London: Nicholas Brealey.

Leadbeater C (2008) *We-think: the power of mass creativity*: Profile Books.

Meadows MS (2008) *I, Avatar. The culture and consequences of having a Second Life*. Berkeley, California: New Riders.

Palfrey J & Gasser U. *2008. Born digital. Understanding the first generation of digital natives*. New York: Basic Books.

Rheingold H (2002) *Smart mobs: the next social revolution*. Cambridge, Massachusetts: Perseus.

Shirky C (2009) *Here comes everybody. How change happens when people come together*. London: Penguin Books.

Stille A (2002) *The future of the past. The loss of knowledge in the age of information*. London: Picador, Pan Macmillan.

Stross, R (2008) *Planet Google. One company's audacious plan to organize everything we know.* New York: Free Press

Tapscott D & Williams AD (2008) *Wikinomics. How mass collaboration changes everything*, Revised ed. London: Atlantic Books.

Tuomi I (2006) *Networks of innovation. Change and meaning in the age of the internet (Originally published 2002)*. Oxford: Oxford University Press.

Vogelstein F (2009) The great wall of Facebook. The social network wants to dominate a new, friendlier internet - and keep Google out. There's going to be war. *Wired (American Edition)*, July 2009, pp. 96-101, 120.

Zittrain J (2008) *The future of the internet: and how to stop it*. New Haven: Yale University Press.

Emerging technologies

Bentley PJ (1999) *Evolutionary design by computers*. San Francisco: Morgan Kaufman Publishers.

Bentley PJ & Corne, DW (2002) *Creative evolutionary systems*. London: Academic Press.

Frank SA (1997) The design of adaptive systems: optimal parameters for variation and selection in learning and development. *Journal of Theoretical Biology* **184**, 31-39.

Gershenfeld N (1999) *When things start to think*. London: Hodder and Stoughton.

Gershenfeld N (2005) *FAB. The coming revolution on your desktop - from personal computers to personal fabrication*. New York: Basic Books.

Hancock, PJB & Frowd, CD (2002) Evolutionary generation of faces. In *Creative evolutionary systems*, pp. 409-423 [PJ Bentley and DW Corne, editors]. London: Academic Press.

Heikkinen J, Rantala J, Olsson T, Raisamo R, Lylykangas J, Raisamo J, Surakka V & Ahmaniemi T (2009) Enhancing personal communication with spatial haptics: two scenario-based experiments on gestural interaction. *Journal of Visual Languages and Computing* **20**, 287-304.

Kampis G & Gulyas L (2006) Emergence out of interaction: developing evolutionary technology for design innovation. *Advanced Engineering Informatics* **20**, 313-320.

Laursen L (2009) A memorable device. *Science* **323**, 1422-1423.

Malone E & Lipson H (2007) Fab@Home: the personal desktop fabricator kit. *Rapid Prototyping Journal* **13**, 245-255.

O'Hara K, Morris R, Shadbolt N, Hitch GJ, Hall W & Beagrie N (2006) Memories for life: a review of the science and technology. *Journal of the Royal Society Interface* **3**, 351-365.

Paterson M (2007) *The senses of touch. Haptics, affects and technologies*. Oxford: Berg.

Vilbrandt T, Malone E, Lipson H & Pasko A (2008) Universal desktop fabrication. In *HOMA, LNCS 4889*, pp. 259-284 [A Pasko, V Adzhiev and P Comninos, editors]. Berlin: Springer-Verlag.

Volino P, Davy P, Bonanni U, Luible C, Magnenat-Thalmann N, Mkinen M & Meinander H (2007) From measured physical parameters to the haptic feeling of fabric. *The Visual Computer: International Journal of Computer Graphics* **23**, 133-142.

Personal history, biography and life-writing

Abate CS (2003) *Privacy, domesticity, and women in early modern England*. Aldershot: Ashgate Publishing.

Beale P (2005) *England's mail. Two millenia of letter writing*. Stroud: Tempus Publishing.

Bowman AK (2003) *Life and letters on the Roman frontier. Vindolanda and its people*. London: The British Museum Press.

Castor H (2004) *Blood & Roses. The Paston family in the fifteenth century*. London: Faber and Faber.

Crain C (2006) Surveillance society. The mass-observation movement and the meaning of everyday life. *The New Yorker.* 11 September 2006.

Etherton J (2006) The role of archives in the perception of self. *Journal of the Society of Archivists* **27**, 227-246.

Figes, O (2007) *The whisperers. Private life in Stalin's Russia*. London: Penguin Books.

Garton Ash T (1997) *The file. A personal history*. New York: Vintage Books.

Hareven TK (1973) The history of the family as an interdisciplinary field. In *The family in history: interdisciplinary essays* [TK Rabb and RI Rotberg, editors]. New York: Harper Torchbook.

Harris F (2004) *Transformations of love. The friendship of John Evelyn & Margaret Godolphin (first published 2002)*. Oxford: Oxford University Press.

Hubble N (2005) *Mass observation and everyday life. Culture, history, theory*. London: Palgrave Macmillan.

Hunt C (2001) Recovery and healing in life writing. In *Encyclopaedia of Life Writing*, pp. 737-739 [M Jolly, editor]. London: Fitzroy Dearborn.

Jolly M (2008) Twenty-First Century Epistolarity and the Truth about Email. In *Escrever A Vida: Verdade e Ficcao*, pp. 115-138 [P. Morao, editor]. Lisbon: Campo Das Letras Press.

Lee H (2005) *Body parts. Essays on life-writing*. London: Pimlico.

Lee R (2005) *Unquiet country. Voices of the rural poor, 1820-1880*. Macclesfield: Windgather Press.

Lüdtke A (1995) *The history of everyday life. Reconstructing historical experiences and ways of life (originally published in German 1989)*, English Translation ed. Princeton: Princeton University Press.

Nozawa S (2007) The meaning of life: regimes of textuality and memory in Japanese personal historiography. *Language & Communication* **27**, 153-177.

Oates SB (1991) *Biography as history*. Waco, Texas: Baylor University Press.

Perks R & Thompson A, editors. (2006) The Oral History Reader. London: Routledge.

Sheridan D (1991) *The Mass-Observation diaries. An introduction*. The Mass-Observation Archive of the University of Sussex Library, and the Centre for Continuing Education, University of Sussex.

Literature, stories, art

Boyd B (2009) *On the origin of stories: evolution, cognition, and fiction*. Cambridge, Massachusetts: Harvard University Press.

Eco U (2002) Literary game of drafts, Guardian Online, http://www.guardian.co.uk/books/2002/mar/30/scienceandnature.philosophy, 30 March 2002.

McGann JJ (2002) Literary scholarship in the digital future. *The Chronicle of HIgher Education* **49,** B7, 13 December 2002.

Sopčák P (2007) 'Creations from nothing': a foregrounding study of James Joyce's drafts for Ulysses. *Language and Literature* **16**, 183-196.

Southam BC (2001) Jane Austen's literary manuscripts. A study of the novelist's development through the surviving papers. London: The Athlone Press. Revised ed.

Sumi K (2009) Interactive storytelling system using recycle-based story knowledge. In *Interactive Storytelling, Second Joint International Conference on Interactive Digital Storytelling (ICIDS 2009), LNCS 5915*, Guimarães, Portugal, 9-11 December 2009, pp. 74-85 [I. A. Iurgel, N. Zagalo and P. Petta, editors]. Berlin: Springer.

iScience

Chippendale P, Zanin M & Andreatta C (2009) Re-photography and environment monitoring using a social sensor network. In *5th International Conference on Image Analysis and Processing (ICIAP 2009), LNCS 6716*, Vietri sul Mare, Italy, 8-11 September 2009, pp. 34-42. Berlin: Springer.

Fowler JH, Dawes CT & Christakis NA (2009) Model of genetic variation in human social networks. *Proceedings of the National Academy of Sciences USA* **106**, 1720-1724.

Gonzalez MC, Hidalgo CA & Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* **453**, 779-782.

Kossinets G & Watts DJ (2006) Empirical analysis of an evolving social network. *Science* **311**, 88-90.

Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D & Van Alstyne M (2009) Computational social science. *Science* **323**, 721-723.

Lewis K, Kaufman J, Gonzalez M, Wimmer A & Christakis N (2008) Tastes, ties, and time: a new social network dataset using Facebook.com. *Social Networks* **30**, 330-342.

Nelson B (2009) Empty archives. *Nature* **461**, 160-163.

Steed A & Milton R (2008) Using tracked mobile sensors to make maps of environmental effects. *Personal Ubiquitous Computing* **12**, 331-342.

Todoroki S-i, Konishi T & Inoue S (2006) Blog-based research notebook: personal informatics workbench for high-throughput experimentation. *Applied Surface Science* **252**.

Watts DJ (2007) A twenty-first century science. *Nature* **445**, 489.

Wolf G (2009) Know thyself. The personal metrics movement goes way beyond diet and exercise. It's about tracking every fact of life, from sleep to mood to pain, 24/7/365. In *Wired*, July 2009, pp. 92-95.

13.2 Specific Research Topics

Forensics and authenticity

Aitken CGG & Stoney DA (1991) *The use of statistics in forensic science*. Chichester: Ellis Horwood.

Carrier B (2005) *File system forensic analysis*. Upper Saddle River: Pearson Education, Addison-Wesley.

Copeland P (2001) Forensic evidence in historical sound-recordings. In *Aural history. Essays on recorded sound*, pp. 107-115. London: The British Library.

Erzinçlioglu Z (2000) *Every contact leaves a trace. Scientific detection in the twentieth century*. London: Carlton Books.

Fraser J & Williams R (2009) *Handbook of forensic science*. Cullompton, Devon: Willan Publishing.

Garfinkel S & Cox D (2009) Finding and archiving the internet footprint. In *Digital Lives Research Conference*. The British Library, London. http://simson.net/clips/academic/2009.BL.InternetFootprint.pdf.

Hagan WE (1894) *A treatise on disputed handwriting and the determination of geunine from forged sigatures. The character and composition of inks, and their determination by chemical tests. The effect of age as manifested in the appearance of written instruments and documents*. New York: Banks & Brothers.

Hilton O (1993) *Scientific examination of questioned documents*, Revised ed. Boca Raton: CRC Press.

Kirschenbaum MG (2008) *Mechanisms. New media and the forensic imagination*. Cambridge, Massachusetts: The MIT Press.

Koppenhaver KM (2007) *Forensic document examination: principles and practice*. Totowa, New Jersey: Humana Press.

Love H (2002) *Attributing authorship: an introduction*. Cambridge: Cambridge University Press.

Nickell J (2005) *Detecting forgery. Forensic investigation of documents*, Paperback ed. Lexington: University Press of Kentucky.

Ordway H (1982) *Scientific examination of questioned documents*, Revised ed. New York: Elsevier.

Osborn AS (1910) *Questioned documents. A study of questioned documents with an outline of methods by which the facts may be discovered and shown*. Rochester, New York: The Lawyers' Co-operative Publishing Company.

Robertson B & Vignaux GA (1995) *Interpreting evidence. Evaluating forensic science in the courtroom*. Chichester: John Wiley & Sons.

Yule GU (1939) On sentence length as a statistical characteristic of style in prose. *Biometrika* **30**, 363.

Archival personal information management

Barreau DK (1995) Context as a factor in personal information management systems. *Journal of the American Society for Information Science* **46**, 327-339.

Bellotti V, Ducheneaut N, Howard M, Smith I & Grinter R (2005) Quality vs quantity: email-centric task-management and its relationship with overload. *Human-Computer Interaction* **20**, 89-138.

Bondarenko O & Janssen R (2005) Documents at hand: learning from paper to improve digital technologies. ACM *Conference on Human Factors in Computing Systems (CHI 2005)*, pp. 121-130.

Checkland PB & Holwell SE (2006) Data, capta, information and knowledge. In *Introducing information management. The business approach*, pp. 47-55 [M Hinton, editor]. London: Elsevier.

Ducheneaut N & Bellotti V (2001) Email as habitat: an exploration of embedded personal information management. *Interactions* **8**, 30-38.

Gemmell J, Bell G & Lueder R (2006) MyLifeBits: a personal database for everything. *Communications of the ACM* **49**, 88-95.

Jones W (2007) Personal information management. In *Annual review of information science and technology*, pp. 453-504 [B Cronin, editor]. Medford, New Jersey: Information Today.

Lanza SR (2006) Arrange some, delete some: TaskArrange and AM-Deadlink. *Searcher* **14**, 21-23.

Marshall C & Jones W (2006) Keeping encountered information. *Communications of the ACM* **49**, 66-67.

McKemmish S (1996) Evidence of me.... *Archives and Manuscripts* **24**, 28-45.

Nakajima Y (2006) What piece of information and how should we select for archival preservation or destruction? A view from the perspectives of archival appraisal theories. *Journal of Information Science and Technology Association (Joho no Kagaku to Gijutsu)* **56**, 554-558.

Savolainen R (2005) Everyday life information seeking. In *Theories of Information Behavior*, pp. 143-148 [K Fisher, S Erdelez and L McKechnie, editors]. Melford, New Jersey: Information Today.

Sellen AJ & Harper R (2002) *The myth of the paperless office*. Cambridge, Massachusetts: MIT Press.

Siegfried S, Bates MJ & Wilde DN (1993) A profile of end-user searching behavior by humanities scholars. The Getty Online Searching Project Report 2. *Journal of the American Society for Information Science* **44**, 273-291.

Tauscher LM & Greenberg S (1997) How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies* **47**, 97-137.

Trace CB (2007) Information creation and the notion of membership. *Journal of Documentation* **63**, 142-164.

Whittaker S, Bellotti V & Gwizdka J (2006) E-mail in personal information management. *Communications of the ACM* **49**, 69-75.

Whittaker S & Hirschberg J (2001) The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction* **8**, 150-170.

Whittaker S, Jones Q, Nardi B, Creech M, Terveen L, Isaacs E & Hainsworth J (2004) ContactMap: organizing communication in a social desktop. *ACM Transactions on Computer-Human Interaction* **11**, 445-471.

Williamson K (1998) Discovered by chance: the role of incidental information acquisition in an ecological model of information use. *Library and Information Science Research* **20**, 23-40.

Usability

Bellotti V, Begole B, Chi EH, Ducheneaut N, Fang J, Isaacs E, King T, Newman MW, Partridge K, Price B, Rasmussen P, Roberts M, Schiano DJ & Walendowski A (2008) Activity-based serendipitous recommendations with the Magitti mobile leisure guide. *ACM Conference on Human Factors in Computing Systems (CHI 2008)*, 1157-1166.

Cooper A, Reimann R & Cronin D (2007) *About face. The essentials of interaction design*. Indianapolis: Wiley Publishing.

Dix A, Finlay J, Abowd GD & Beale R (2003) *Human-computer interaction*, Third ed. London: Prentice Hall.

Moggridge B (2007) *Designing interactions*. Cambridge, Massachusetts: The MIT Press.

Uchyigit G & Ma MY (2008) *Personalization techniques and recommender systems*. Singapore: World Scientific.

Wallace M, Angelides MC & Mylonas P (2008) *Advances in semantic media adaptation and personalization*. Berlin: Springer.

Weir CS, Douglas G, Carruthers M & Jack M (2009) User perceptions of security, convenience and usability for ebanking authentication tokens. *Computers & Security* **28**, 47-62.

Zaphiris P & Ang CS (2009) *Cross-disciplinary advances in human computer interaction: user modeling, social computing, and adaptive interfaces*. Information Science Reference.

Evolutionary dynamics, complex systems and information ecology

Bedau MA (2003) Artificial life: organization, adaptation and complexity from the bottom up. *Trends in Cognitive Science* **7**, 505-512.

Bentley RA, Lipo CP, Herzog HA & Hahn MW (2007) Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* **28**, 151-158.

Blackmore S (1999) *The meme machine*. Oxford: Oxford University Press.

Bryden KM, Ashlock DA & McCorkle D (2004) An application of graph based evolutionary algorithms for diversity preservation. *IEEE Xplore*, 419-426.

Cisne JL (2005) How science survived: medieval manuscripts' 'demography' and classic texts' extinction. *Science* **307**, 1305-1307.

Emmeche C (1994) *The garden in the machine. The emerging science of artificial life (Published in Danish, 1991)*. Princeton: Princeton University Press.

Fagerberg J & Verspagen B (2002) Technology-gaps, innovation-diffusion and transformation: an evolutionary interpretation. *Research Policy* **31**, 1291-1304.

Jablonka E & Lamb MJ (2005) *Evolution in four dimensions. Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, Massachusetts: The MIT Press.

Mesoudi A (2007) Biological and cultural evolution: similar but different. *Biological theory* **2**, 119-123.

Mesoudi A, White S & Dunbar R (2006) A bias for social information in human cultural transmission. *British Journal of Psychology* **97**, 405-423.

Wilke CO & Adami C (2002) The biology of digital organisms. *Trends in Ecology and Evolution* **17**, 528-532.

Phylogenetics and stemmatics

Bandelt H-J, Forster P & Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37-48.

Barbrook AC, Howe CJ, Blake N & Robinson P (1998) The phylogeny of *The Canterbury Tales*. *Nature* **394**, 839-840.

Bryant D, Filiman F & Gray RD (2005) Untangling our past: languages, trees, splits and networks. In *The evolution of cultural diversity. A phylogenetic approach*, pp. 67-83 [R Mace, CJ Holden and S Shennan, editors]. London: UCL Press.

Carrera E & Erdelyi G (2004) Digital genome mapping - advanced binary malware analysis. *Virus Bulletin Conference*, 187-197.

Goldberg LA, Goldberg PW, Phillips CA & Sorkin GB (1998) Constructing computer virus phylogenies. *Journal of Algorithms* **26**, 188-208.

Howe CJ, Barbrook AC, Mooney LR & Robinson P (2004) Parallels between stemmatology and phylogenetics. In *Studies in Stemmatology II*, pp. 3-11 [P van Reenen, A den Hollander and M van Mulken, editors]. Amsterdam: John Benjamins Publishing.

Howe CJ, Barbrook AC, Spencer M, Robinson P, Bordalejo B & Mooney LR (2001) Manuscript evolution. *Trends in Genetics* **17**, 147-152.

Mace R, Holden CJ & Shennan S (2005) *The evolution of cultural diversity. A phylogenetic approach*. London: UCL Press.

Pagel M (2009) Human language as a culturally transmitted replicator. *Nature Reviews Genetics* **10**, 405-415.

Reeves PA & Richards CM (2007) Distinguishing terminal monophyletic groups from reticulate taxa: performance of phenetic, tree-based, and network procedures. *Systematic Biology* **56**, 302-320.

Reynolds LD & Wilson NG (1991) *Scribes and scholars. A guide to the transmission of Greek and Latin literature*, Third ed. Oxford: Clarendon Press.

Spencer M, Bordalejo B, Wang L-S, Barbrook AC, Mooney LR, Robinson P, Warnow T & Howe CJ (2003) Analyzing the order of items in manuscripts of *The Canterbury Tales*. *Computers and the Humanities* **37**, 97-109.

Spencer M, Davidsion EA, Barbrook AC & Howe CJ (2004) Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology* **227**, 503-511.

Infoethics and value

Butler D (2007) Data sharing threatens privacy. *Nature* **449**, 644-645.

Capurro R (2000) Ethical challenges of the information society in the 21st century. *International Information and Library Review* **32**, 257-276.

Donath J (2007) Virtually trustworthy. *Science* **317**, 53-54.

Dyson, E (2003) Online registries: the DNS and beyond. *Release 1.0. Esther Dyson's Monthly Report*, **21**(8), pp. 1-33, 16 September 2003. New York: EDventure Holdings.

Enserink M (2009) Are you ready to become a number? *Science* **323**, 1662-1664.

Golbeck J (2008) Trust on the world wide web: a survey. *Foundations and Trends in Web Science* **1**, 131-197.

Goldberg I & Atallah MJ (2009) *Privacy enhancing technologies. Proceedings, 9th International Symposium PETS 2009, August 2009, Seattle, Washington*. Berlin: Springer.

Lehikoinen JT, Lehikoinen J & Huuskonen P (2008) Understanding privacy regulation in ubicomp interactions. *Personal Ubiquitous Computing* **12**, 543-553.

Méndez-Durón R & García CE (2009) Returns from social capital in open source software networks. *Journal of Evolutionary Economics* **19**, 277-295.

O'Hara K & Shadbolt N (2008) *The spy in the coffee machine. The end of privacy as we know it*. Oxford: Oneworld.

Schull J (2007) Predicting the evolution of digital rights, digital objects, and digital rights management languages. In *Digital rights management: an introduction* [D Satish, editor]. Andhra Pradesh, India: ICFAI Books.

Solove DJ (2004) *The digital person: technology and privacy in the information age*. New York: New York University Press.

Solove DJ (2007) *The future of reputation. Gossip, rumor, and privacy on the internet*. New Haven: Yale University Press.

Solove DJ (2008) *Understanding privacy*. Cambridge, Massachusetts: Harvard University Press.

Zerhouni E & Nabel EG (2008) Protecting aggregate genomic data. *Science* **322**, 44.

Advanced cataloguing: icons, semantics, contexts

Al-Khalifa HS (2007) Automatic document-level semantic metadata annotation using folksonomies and domain ontologies. PhD, University of Southampton.

Atkinson-Abutridy J, Mellish C & Aitken S (2003) A semantically guided and domain-independent evolutionary model for knowledge discovery from texts. *IEEE Transactions on Evolutionary Computation* **7**, 546-560.

Cook DL, Farley JF & Tapscott SJ (2001) A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biology* **2**.

Feigenbaum L, Herman I, Hongsermeier T, Neumann E & Stephens S (2007) The semantic web in action. *Scientific American* **297**, 64-71.

Hoang H, Tjoa A & Nguyen M (2006) Ontology-based virtual query system for the SemanticLIFE Digital Memory Project: concepts, designs and implementations. *Addendum Contributions of the 4th IEEE International Conference on Computer Sciences*, 39-45.

Kolhoff P, Preuß J & Loviscach J (2008) Computer-based icons for music files. *Computers & Graphics* **32**, 550-560.

Neamatullah I, Douglass MM, Lehman L-wH, Reisner A, Villarroel M, Long WJ, Szolovits P, Moody GB, Mark RG & Clifford GD (2008) Automated de-identification of free-text medical records. In *BMC Medical Informatics and Decision Making* **8**, issue 32.

Pauwels EJ, de Zeeuw PM & Ranguelova EB (2009) Computer-assisted tree taxonomy by automated image recognition. *Engineering Applications of Artificial Intelligence* **22**, 26-31.

Suh B & Bederson BB (2007) Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. *Interacting with Computers* **19**, 524-544.

Watson AT, O'Neill. MA & Kitching IA (2003) Automated identification of live moths (Macrolepidoptera) using Digital Automated Identification SYstem (DAISY). *Systematics and Biodiversity* **1**, 287-300.

New research techniques, visualisation and metrics

Berners-Lee T, Hall W, Hendler JA, O'Hara K, Shadbolt N & Weitzner DJ (2006) A framework for web science. *Foundations and Trends in Web Science* **1**, 1-130.

Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, Rodriguez MA & Balakireva L (2009) Clickstream data yields high-resolution maps of science. *PLoS ONE* **4**, e4803.

Bollen J, Van de Sompel H, Hagberg A & Chute R (2009) A principal component analysis of 39 scientific impact measures. *PLoS ONE* **4**, e6022.

Butler D (2009) Web usage data outline map of knowledge. *Nature* **458**, 135.

Christakis NA & Fowler JA (2009) Social network visualization in epidemiology. *Norsk Epidemiologi* **19**, 5-16.

Danis CM, Viegas FB, Wattenberg M, Kriss J. (2008) Your place or mine? Visualization as a community component. In *26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, Florence, Italy, pp. 275-284.

Donath J & boyd d (2004) Public displays of connection. *BT Technology Journal* **22**, 71-82.

Eckmann J-P, Moses E & Sergi D (2004) Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences USA* **101**, 14333-14337.

Forster P, Toth A & Bandelt H-J (1998) Evolutionary networks of word lists: visualising the relationships between Alpine Romance languages. *Journal of Quantitative Linguistics* **5**, 174-187.

Hendler J, Shadbolt N, Hall W, Berners-Lee T & Weitzner D (2009) Web science: an interdisciplinary approach to understanding the web. *Communications of the ACM* **51**, 60-69.

Snee H (2008) Web 2.0 as a social science research tool. In *British Library Social Science Collections and Research*, pp. 1-34. London: The British Library.

Turchin P (2008) Arise 'cliodynamics'. *Nature* **454**, 34-35.

Viégas FB, Wattenberg M & Dave K (2004) Studying cooperation and conflict between authors with *history flow* visualizations. *CHI 2004,* Vienna, Austria. **pp.** 575-582.

Adaptive curatorial systems and digital archiving

Beagrie N (2005) Plenty of room at the bottom? Personal digital libraries and collections. *D-Lib Magazine* **11**.

Churchill E & Ubois J (2008) Designing for digital archives. *Interactions*, March & April 2008, 10-13.

Cunningham A (1994) The archival management of personal records in electronic form: some suggestions. *Archives and Manuscripts* **22**, 94-105.

Feeney M (1999) *Digital culture: maximising the nation's investment*. London: The National Preservation Office, The British Library.

Group NFW (2007) A framework of guidance for building good digital collections, pp. 1-95. Baltimore, Maryland: National Information Standards Organization.

Larman C & Basili VR (2003) Iterative and incremental development: a brief history. *Computer*, 47-56.

Marshall C, Bly S & Brun-Cottan F (2006) The long term fate of our digital belongings: toward a service model for personal archives. *IS&T Archiving 2006*, 25-30.

Thomas S & Martin J (2006) Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives. *Journal of the Society of Archivists* **27**, 29-56.

Digital preservation, digital media and storage

Abrams SL (2004) The role of format in digital preservation. *VINE: The Journal of Information and Knowledge Management Systems* **34**, 49-55.

Aitken B, Helwig P, Jackson A, Lindley A, Nicchiarelli E, Ross S. 2008. The Planets testbed: science for digital preservation. *The Code{4}Lib Journal*, issue 3, 2008-06-23, http://journal.code4lib.org/articles/83.

Becker C, Kulovits H & Rauber (2010) A trustworthy preservation planning with Plato. *European Research Consortium for Informatics and Mathematics, ERCIM News* **80**, 24-25.

Bergeron B (2002) *Dark Ages II. When the digital data die*. Upper Saddle River: Prentice Hall.

Born G (1995) *The file formats handbook*. London: International Thomson Computer Press.

Farquhar A & Hockx-Yu H (2008) Planets: integrated services for digital preservation. *Serials* **21**, 140-145.

Ferreira, M, Baptista AA & Ramalho JC (2007) An intelligent decision support system for digital preservation. *International Journal on Digital Libraries* **6**, 295-304.

Hedstrom M & Lee CA (2002) Significant properties of digital objects: definitions, applications, implications. *DLM-Forum: Parallel Session 3*, 218-223.

Heydegger V (2009) Just one bit in a million: on the effects of data corruption in files, *European Conference on Digital Libraries (ECDL 2009)*, LNCS 5714, pp 315-326.

Hockx-Yu H & Knight G (2008) What to preserve? Significant properties of digital objects. *International Journal of Digital Curation* **3**, 141-153.

Holdsworth D & Wheatley P (2001) Emulation, preservation, and abstraction. *RLG DigiNews 5*, 15 August 2001.

Jones M & Beagrie N (2001) *Preservation management of digital materials. A handbook*. London: The British Library.

Ross S (2007) Digital preservtion, archival science and methodological foundations for digital libraries, *European Conference on Digital Libraries (ECDL 2007)*, Budapest, 11 September 2007.

Todd M (2009) File formats for preservation. *DPC Technology Watch Series Report 09-02*, Digital Preservation Coalition, October 2009.

van der Hoeven JR (2007) Dioscuri: emulator for digital preservation. *D-Lib Magazine* 12.

van der Hoeven JR, van Diessen RJ & van der Meer K (2005) Development of a Universal Virtual Computer for long-term preservation of digital objects. *Journal of Information Science* **31**, 196-208.

Watkinson J (2001) *The art of digital audio*. Oxford: Focal Press, Elsevier.

Chapter 14: The First Digital Lives Research Conference

14.1 Programme

Day 1 Monday 9 February 2009
Digital Lifelines: Practicalities, Professionalities and Potentialities

Welcoming Address

Ronald Milne, British Library
Welcome and Introduction

Opening Keynote Lecture

Cathy Marshall, Microsoft Research
Benign Neglect in a Digital World: a Pragmatic Look at Personal Digital Archiving

Session 1: Aspects of Digital Curation
Digital Lives Invited Presentations

Cal Lee, University of North Carolina
Pondering the Professionalities: Curating the Personal

Naomi Nelson, Emory University
Collecting Digital Lives: A Tale of Two Donors

Michael G. Olson, Stanford University Libraries
Born Digital Collection Materials at Stanford

Ludmila Pollock, Cold Spring Harbor Laboratory
Cold Spring Harbor Laboratory's Oral History Project: Yesterday, Today and Tomorrow

Session 2: On the Monetary Value of Personal Digital Objects
Valuing Digital Manuscripts Panel Discussion
with Jamie Andrews, British Library

Gabriel Heaton, Sotheby's
Julian Rota, Bertram Rota
Joan Winterkorn, Bernard Quaritch

Tours and Demonstrations

A *Digital Scriptorium for eMANUSCRIPTS at the British Library*

Audio Technologies in the Conservation Centre at the British Library

Keynote Lecture on Digital Economy and Philosophy

Annamaria Carusi, University of Oxford
Value, Reality and Objects: What is Digital about a Digital Economy?

Session 3: Archiving the Moving Image
Digital Lives Invited Presentation

Luke McKernan, British Library
From Home Movies to Lifecasting: Archiving Personal Lives on Film

Session 4: Digital Preservation
Digital Lives Invited Presentations

Peter Bright, British Library
British Library Digital Preservation Team: Introduction and Activities

Will Prentice, British Library
Preserving Sound Recordings

Juan-José Boté, University of Barcelona
Preserving Private Collections

Session 5: Practical Experiences
Digital Lives Invited Presentations

John Blythe, University of North Carolina
Digital Dixie: Processing Born-digital Materials in the Southern Historical Collection

Erika Farr, Emory University
When Papers Aren't Just Paper: Hybrid Archives at Emory

Gabriela Redwine, Harry Ransom Center, University of Texas, Austin
Born-digital Materials at the Harry Ransom Center

William Snow, Stanford Univeristy Libraries
Self-Archiving Legacy Toolkit: Computer-assisted Semantic Annotation of Scientific Life Works

Session 6: Professional Matters Arising, Options for the Future and Resolutions

Brief concluding discussion with questions and answers

Last words

Day 2 Tuesday 10 February 2009
Personal Information Lifecycles: Creator, Curator, Consumer

Inaugurating Keynote Address

Dame Lynne Brindley DBE, Chief Executive, British Library
The First Digital Lives Research Conference at the British Library

Session 1: Digital Lives
Digital Lives Presentation

Jeremy Leighton John, British Library
Brief Introduction to Digital Lives

The First Digital Lives Conference Keynote Lecture

Gordon Bell, Microsoft Research
Preserving the 20th & 21st Century Digital Lives: MyLifeBits, a 'digital life' experience

Session 2: Personal Information Management and Usability
Digital Lives Presentation

Ian Rowlands and Peter Williams, University College London
Personal Information Management and Usability

Usability Keynote Lecture

Victoria Bellotti, Xerox PARC, California
Inventing the Future

Session 3: Forensics, Authenticity, Security and Digital Capture
Digital Lives Presentation

Jeremy Leighton John, British Library
Digital Forensics for Personal Archives

Forensics & Digital Capture Keynote Lecture

Simson Garfinkel, Naval Postgraduate School, California
Finding and Archiving the Internet Footprint

Session 4: Historical Research and Private Lives
Digital Lives Presentation

Rob Perks, British Library
Historians, Curators and Personal Data in a Digital World

Historical Keynote Lecture

Orlando Figes, Birkbeck College London
Working with the Families of Victims of Repression in Stalinist Russia

Session 5: Scientific Research with People

Developmental Psychology Keynote Lecture

Charles Fernyhough, University of Durham
Constant Observers: Children in the Camcorder's Eye

Session 6: Towards Digital Biography

Literary Biography Keynote Lecture

Andrew Lycett, London
The Dog in the Nighttime

Scientific Biography Keynote Lecture

Georgina Ferry, Oxford
Molecules and the Meaning of Life

Session 7: Legal and Ethical Issues

Digital Lives Presentation

Andrew Charlesworth, University of Bristol
We Can Remember It for You Wholesale: Law and Ethics in Digital Repositories

Privacy Keynote Lecture

Kieron O'Hara, University of Southampton
Privacy Rights and Privacy Responsibilities

Session 8: Creators' Experiences, Anticipations and Thoughts

Digital Lives Presentation

Jamie Andrews, British Library
Hedda Gabler 2.0: *Literary Production in a Digital Age*

Writers in Conversation

with Dame Wendy Hall DBE

Rt Hon Anthony Wedgewood Benn PC
Dame Antonia Byatt DBE
Wendy Cope

Closing Remarks

Day 3 Wednesday 11 February 2009
Living Online and Digital Archives in the Wild

Archival Keynote Lecture

Dorothy Sheridan MBE, Mass Observation Archive
Is Blogging the new Mass Observation? Diary-Writing and Documentary in the Digital Era

Computer Science Keynote Lecture

Dame Wendy Hall DBE, University of Southampton
Memories for Life

Introduction

Jeremy Leighton John, British Library
Digital Archives in the Wild and Living Online

Session 1: iSCIENCE?
Science Online Keynote Lecture

Timo Hannay, Nature Publishing Group
Scientific Communication in the Digital Age

Digital Lives Invited Presentation

Peter Shepherd, Centre for Longitudinal Research, Institute of Education
Following 50,000 Lives: the Centre for Longitudinal Studies and the British Cohort Studies

Session 2: Mobile Learning
Digital Lives Invited Presentation

Eileen Scanlon, Open University
Personal Inquiry and the Use of Mobile Technology in Learning in Formal and Informal Settings

Session 3: Web 2.0 & Cloud Computing
Digital Lives Invited Presentation

Simone Brunozzi, Amazon Web Services
Cloud Computing with Amazon Web Services

Session 4: iLITERATURE
Digital Lives Invited Presentation

K. Faith Lawrence, Royal Irish Academy
Subversion, Devotion and Community: Online Fiction in the Digital World

Session 5: Emerging Technologies
Social Networks Keynote Lecture

Jon Crowcroft, University of Cambridge
Online Social Networks and Real Life

Virtual Research and Digital Pens Keynote Lecture

Mark Baker, University of Reading
Virtual Research Environments and the Use of Digital Devices

Session 6: Visualisation, Future Access and iART
Evolutionary Technology Keynote Lecture

Peter J. Bentley, University College London
The Day Before Darwin's 200th Birthday: Evolution and Computers

Digital Lives Invited Presentation

Stefanie Posavec, Penguin Books, formerly Central Saint Martin College of Art and Design
Writing without Words: Visualising Writing Styles in Literature

Session 7: Internet Law
Digital Lives Invited Presentation

Lilian Edwards, University of Sheffield
Walled Gardens or Open Worlds? Virtual Lives, Real World Laws, and Issues of Control

Session 8: Virtual Worlds

Digital Lives Invited Presentations

Ren Reynolds, the Virtual Policy Network
The Social Life of Virtual Worlds

Dave Taylor, Imperial College London
Second Health: Second Life and Healthcare

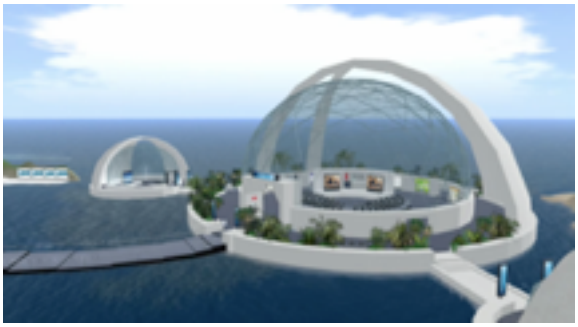
Ian Hughes, IBM
Metaverses and Leading the World

Jerome McDonough, University of Illinois
Preserving Virtual Worlds

Session 9: Digital Life in Remote Places

Expedition Keynote Lecture

Ben Saunders, Polar Athlete and Explorer and Expedition Guide
Digital Life at the Extremes



Session 10: Digital Conversazione

Digital Lives Virtual Presentation
from the Elucian Islands Archipelago in Second Life
Jean-Claude Bradley, Drexel University
Open Notebook for the Laboratory

Digital Lives Virtual Presentation
from the Elucian Islands Archipelago in Second Life
George Oates, formerly Flickr
On Online Sustainability

Shutdown

14.2 Conference Session Chairs

>> Jamie Andrews, British Library
>> Maxine Clarke, Nature Publishing Group
>> Tim Gollins, The National Archives
>> Dame Wendy Hall, University of Southampton
>> Peter Hirtle, Cornell University
>> Jeremy Leighton John, British Library
>> Clifford Lynch, CNI
>> Cathy Marshall, Microsoft Research
>> Mags McGeever, University of Edinburgh
>> David Nicholas, University College London
>> Richard Ranft, British Library
>> Anne Sebba, PEN
>> Dorothy Sheridan, Mass Observation Archive, University of Sussex
>> Susan Thomas, Bodleian Library, Oxford University
>> John Tuck, Royal Holloway College London
>> Mark van Harmelen, University of Manchester

Unable to attend at last minute due to unforeseen circumstances, and greatly missed: Liz Lyon (University of Bath) and Pelle Snickars (Swedish National Library)

14.3 Some of the Comments on the Conference

To encourage readers of this synthesis to come to the next Digital Lives Research Conference:

I enjoyed the event a great deal – I thought you did a splendid job with a diverse but fascinating set of speakers

I realise it's been quite a while now since the conference, but I just wanted to say a huge thank you for organising it. The whole thing was incredibly useful, and I learnt so much. By the end of the third day I felt like my brain would explode if I learnt anything more!

Thanks for inviting me! It was an exceptionally interesting conference. I was particularly interested to learn about all of the interesting pockets of research on people's daily lives going on in the UK, like the Mass Observation research. Fun stuff.

First, I can't thank you enough for all your efforts at hosting the Digital Lives conference. It was an amazing three days—I learned so much and met people I am very happy to have met.

BL colleague

Congratulations. That was brilliant, and I have to say among the 3 most inspiring days I've had since joining the BL.

There's so much for us to pick up on in terms of future collaborations...

Thank you again for hosting such a wonderful conference. I'm still reeling from it.

It was a fabulous, well organized conference with such an interesting group of presenters. I came away very inspired!

Chapter 15: Digital Lives Publications and Presentations

15.1 Publications

Published, in press or submitted

(1) Williams, Peter; Katrina Dean; Ian Rowlands; and Jeremy Leighton John. **Digital Lives: Report of interviews with the creators of personal digital collections.** *Ariadne* 55, April 2008. http://www.ariadne.ac.uk/issue55/williams-et-al/. Published online (2008).

(2) John, Jeremy Leighton. **Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives and tools.** *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects*. The British Library Conference Centre, London. http://www.bl.uk/ipres2008/programme.html. Published online, peer-reviewed (2008).

(3) Williams, Peter; Jeremy Leighton John; and Ian Rowlands. **The personal curation of digital objects: a lifecycle approach.** *Aslib Proceedings: New Information Perspectives, 61 (4): 340-363*. Published, peer-reviewed (2009).

(4) John, Jeremy Leighton. **The future of saving the past.** *Nature*, 459: 775-776, 11 June 2009. Published (2009).

(5) Andrews, Jamie. **Save, get, delete.** *Times Literary Supplement*. 13 March 2009, page 15. Published (2009).

(6) John, Jeremy Leighton. **Digital forensics, ancestral computing, evolutionary tools and perspectives.** Submitted, peer-reviewed.

(7) Andrews, Jamie; with Jeremy Leighton John. **Save the ephemeral.** *The Author. Journal of the Society of Authors*. Summer 2009 issue, pages 70-71. Published (2009).

(8) Andrews, Jamie. **'Laid aside'? Collecting contemporary literary archives and manuscripts.** *Archives (British Records Association)*. In press, peer-reviewed.

(9) Charlesworth, Andrew. **Digital Lives >> Legal and Ethical Issues.** Digital Lives Research Paper, 14 October 2009. Published online (2009).

(10) John, Jeremy Leighton; Ian Rowlands; Peter Williams; and Katrina Dean. **Digital Lives >> An Initial Synthesis.** Digital Lives Research Paper, 22 February 2010, Beta version 0.1A. Published online (2010), present document.

In preparation

(11) John, Jeremy Leighton (and others). **iCURATION: from I to i**, 40+ pp, to be submitted shortly, peer-reviewed.

(12) Dean, Katrina; Jeremy Leighton John; and Peter Williams. **Scientists and their personal digital archives** (provisional title, authors and author order). Peer-reviewed.

(13) John, Jeremy Leighton; and others. **Personal digital manuscripts and the individual** (provisional title and authorship), an online publication. Peer-reviewed.

(14) John, Jeremy Leighton. **On phylogeny and history** (provisional title). Peer-reviewed.

Planned

(15) Rowlands, Ian; Peter Williams; and Jeremy Leighton John. **Personal information management for digital lives** (provisional title and author order). Peer-reviewed.

(16) Rowlands, Ian; Jeremy Leighton John; and Peter Williams. **Digital lives for all: diversity and usability** (provisional title and author order). Peer-reviewed.

(17) John, Jeremy Leighton; Jamie Andrews; Peter Williams; Ian Rowlands; and others. **Personal curation: context and creativity** (provisional title, authors and author order). Journal article. Peer-reviewed.

15.2 Presentations

(1) **Digital Preservation: Setting the Course for a Decade of Change.** Neil Beagrie, 19 November 2007, Brussels. *L'Europe face au défi de la conservation des documents numériques à long terme*, 60th Anniversary Conference, Association Belge de Documentation, Brussels.

(2) **Digital Manuscripts, Digital Lives. The Recycling of Freestyle and Personal Information.** Jeremy Leighton John, 23 April 2008, Stockholm. *Future Proof IV. International Conference on Scientific Archives.* Royal Swedish Academy of Sciences, Stockholm, Sweden.

(3) **Digital Lives: an AHRC-funded Research Project.** John Tuck, 3 June 2008, Philadelphia. *Creating the Future*, a meeting session, RLG Programs 2008 Annual Partners Meeting, Philadelphia, USA.

(4) **Digital Lives Reseach Project.** Ian Rowlands, 11 July 2008, Belfast. *The JISC/CNI Meeting: Transforming the User Experience,* Seventh International JISC/CNI Conference, Belfast, Northern Ireland.

(5) **The UK Digital Lives project: funded by the Arts and Humanities Research Council.** John Tuck, 8 August 2008, Perth. *Archives: Discovery and Exploration.* Australian Society of Archivists Annual Conference, Perth, Australia.

(6) **An Introduction to Digital Lives: an AHRC-funded Research Project.** John Tuck, 28 August 2008, York. *Spanning the Spectrum: Confronting Recordkeeping Challenges*, Society of Archivists Conference, University of York.

(7) **Digital Lives: a Personal Information Management Perspective.** Ian Rowlands, 28 August 2008, York. *Spanning the Spectrum: Confronting Recordkeeping Challenges*, Society of Archivists Conference, University of York.

(8) **Digital Lives, Digital Manuscripts. The Recycling of Personal and Freestyle Information.** Jeremy Leighton John, 28 August 2008, York. *Spanning the Spectrum: Confronting Recordkeeping Challenges*, Society of Archivists Conference, University of York.

(9) **Adapting Existing Technologies for Digitally Archiving Personal Lives. Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools.** Jeremy Leighton John, 29 September 2008, London. *iPRES 2008: The Fifth International Conference on Preservation of Digital Objects*, British Library Conference Centre.

(10) **Digital Archives in the Wild. The Recycling of Life Information.** Jeremy Leighton John, 11 November 2008, Lewes. *Epistemology of the Archive. Archiving and Reusing Qualitative Data: Theory, Methods and Ethics across Disciplines.* Workshop organised by the Centre for Research on Socio-cultural Change (CRESC), University of Manchester, with University of Sussex.

(11) **Archives for Digital Lives. Writers and their eMANUSCRIPTS.** Jeremy Leighton John (speaker) with Jamie Andrews, 15 November 2008, Austin, Texas. *Creating a Usable Past: Writers, Archives, and Institutions. Flair Symposium at the* Harry Ransom Center, University of Texas at Austin, Texas, USA.

(12) **Personal Archives, Digital Lives. Writers and their eMANUSCRIPTS.** Jeremy Leighton John, 2 December 2008, London. *Between the Lines: Perspectives on Literary Archives*, Annual Conference 2008 of the British Records Association, British Library Conference Centre.

(13) **Operation Manuscript: Collecting the Papers of Living Writers in Britain.** Jamie Andrews, 27 January 2009, Cardiff. *Centre for Editorial & Intertextual Research, Seminar,* Cardiff University, Wales.

(14) **From eMANUSCRIPTS to Digital Lives: Introduction to the Digital Lives Research Project.** Jeremy Leighton John, 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(15) **Personal Information Management and Usability.** Ian Rowlands and Peter Williams (joint speakers), 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(16) **Digital Forensics for Personal Archives.** Jeremy Leighton John, 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(17) **Historians, Curators and Personal Data in a Digital World.** Rob Perks, 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(18) **We Can Remember It for You Wholesale. Law and Ethics in Digital Repositories.** Andrew Charlesworth, 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(19) ***Hedda Gabler 2.0:* Literary Production in a Digital Age.** Jamie Andrews, 10 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, British Library Conference Centre.

(20) **Digital Archives in the Wild and Living Online.** Jeremy Leighton John, 11 February 2009, London. *Personal Digital Archives for the 21st Century*, First Digital Lives Research Conference, Conservation Centre at the British Library.

(21) **Digital Lives. Usability for Personal Digital Archives for Usability.**
Jeremy Leighton John (speaker) with Ian Rowlands and Peter Williams, 25 March 2009 Edinburgh. *The Influence and Impact of Web 2.0 on eResearch Infrastructure, Applications and Users eScience Conference* at the National eScience Institute, Edinburgh, Scotland.

(22) **Digital Lives: eMANUSCRIPTS and Personal Digital Archives.** Jeremy Leighton John, 3 April 2009, Chapel Hill. *Digital Curation: Practice, Promise & Prospects. DigCCurr 2009.* University of North Carolina at Chapel Hill, North Carolina, USA.

(23) **Digital Lives, iCURATION, and Archives in the Wild.** Jeremy Leighton John, 19 May 2009, Oxford. *The OeRC Seminar* organised by the Oxford eResearch Centre (OeRC), University of Oxford, with live video link to other eResearch Centres including University of Southampton, University of Reading and the Science and Technology Facilities Council eScience Research Centre at the Rutherford Appleton Laboratory and the Daresbury Laboratory.

(24) **Digital Lives: How People Create, Manipulate and Store their Personal Digital Archives.** Peter Williams (speaker) with Ian Rowlands and Jeremy Leighton John, 24 June 2009, College Park. *Digital Humanities 2009 Conference*, Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland, College Park, USA.

ACKNOWLEDGEMENTS

We are very grateful indeed to the people who kindly agreed to be interviewed for the project: >> Joshua Benn >> Tony Benn >> James Brown >> Simon Coles >> David Gurteen >> Richard Henderson >> Rolfe Kentish >> M. J. Long >> Patrick Marber >> Daniel Meadows >> Martin Siegert >> and the even greater number of anonymous interviewees.

Participants at the workshops played a crucial role in directing the project: >> Fran Baker >> Guy Baxter >> Tilly Blyth >> Gary Brannan >> Gareth Burfoot >> Martin Campbell-Kelly >> Else Churchill >> Maxine Clarke >> Ifor ap Dafydd >> Sue Donnelly >> Annette Faux >> Maggie Ferguson >> Stella Halkyard >> Frances Harris >> Lorna Hughes >> Oliver Urquhart Irvine >> Arwel Jones >> Jack Latimer >> Hannah Little >> Anna Mayer >> Luke McKernan >> Kathleen O'Riordan >> Alysoun Sanders >> Anne Sebba >> David Shaw >> Dorothy Sheridan >> Helene Snee >> Boni Sones >> Bill Stockting >> Tilli Tansey >> Susan Thomas >> Dave Thompson >> Natalie Walter >> John Wells >> Lynn Young >>

From speakers and session chairs through to the people who conducted the tours of audio technology, many people made the conference possible: >> Christine Adams >> Jamie Andrews >> Rob Ainsley >> Adrian Arthur >> Mark Baker >> Kate Bates >> Gordon Bell >> Victoria Bellotti >> Rt Hon Anthony Wedgewood Benn >> Josh Benn >> Peter J. Bentley >> John Blythe >> Juan-José Boté >> Jean-Claude Bradley >> Peter Bright >> Dame Lynne Brindley >> Simone Brunozzi >> Mirjam Brusius >> Gary Burden >> Dame Antonia Byatt >> Annamaria Carusi >> Matt Casswell >> Andrew Charlesworth >> Maxine Clarke >> Wendy Cope >> Jon Crowcroft >> Colin Day >> Suhkvinder Dhaliwal >> Alf Eaton >> Lilian Edwards >> Kelvin Eli >> Alison Faraday >> Erika Farr >> Keren Flavell >> Charles Fernyhough >> Georgina Ferry >> Orlando Figes >> Lois Froud >> Simson Garfinkel >> Jim Gemmell >> Tim Gollins >> Dame Wendy Hall >> Timo Hannay >> Gabriel Heaton >> Peter Hirtle >> Ian Hughes >> Sav Ioannou >> Jacob Lant >> K. Faith Lawrence >> Cal Lee >> Kissley Leonor >> Andrew Lycett >> Clifford Lynch >> Liz Lyon >> Steve Mahar >> Cathy Marshall >> Andrew MacCalman >> Jerome McDonough >> Mags McGeever >> Luke McKernan >> Ronald Milne >> Savan Modha >> Tom Moulton >> Naomi Nelson >> George Oates >> Kieron O'Hara >> Michael G. Olson >> Charlotte Orrell-Jones >> John Overeem >> Jennie Patrice >> Rob Perks >> Ludmila Pollock >> Stefanie Posavec >> Will Prentice >> Richard Ranft >> Gabriela Redwine >> Ren Reynolds >> Julian Rota >> Ian Rowlands >> Ben Sanderson >> Jane Sartin >> Ben Saunders >> Eileen Scanlon >> Joanna Scott >> Anne Sebba >> Peter Shepherd >> Dorothy Sheridan >> Pelle Snickars >> William Snow >> Mark Stewart >> Dave Taylor >> Susan Thomas >> John Tuck >> Adrian Turner >> Mark van Harmelen >> Andy Ward >> Colin Wight >> Peter Williams >> Joan Winterkorn >> Louise Woodley >>

Experts from a variety of disciplines supported the work of the project and conference with assistance, suggestions or invitations to speak: >> Mark Baker >> Grace Baynes >> Sacha Brostoff >> Alexandrina Buchanan >> Declan Butler >> Tom Cramer >> Joy Davidson >> Alexandra Eveleigh >> Kristen French >> Marina Jirotka >> Margaretta Jolly >> Nicola Jones >> Matthew Kirschenbaum >> John Lester >> Frank McCown >> Matthew Mascord >> Niamh Moore >> Heather Needham >> Susan Nicholls >> Jim Purbrick >> Chris Rusbridge >> Chris Russell >> Deepak Singh >> Richard Waller >> Sarah Walton >> Claire Warwick >>

Similarly two conferences in particular played an influential role, and thanks are extended to the organisers: iPres 2008 >> Adam Farquhar >> Jane Humphreys >> Rui Miao; DigCCurr 2009 >> Rachel Clemens >> Carolyn Hank >> Cal Lee >> Helen Tibbo >>

The following people at the British Library helped the project in important and diverse ways: Clive Billenness >> Richard Boulderstone >> Lynne Brindley >> Stephen Bury >> Jude England >> Adam Farquhar >> Jill Finney >> Chris Grohmann >> Anna Grundy >> Frances Harris >> Helen Hockx-Yu >> Simon Hughes >> Oliver Urquhart Irvine >> Eileen Kinghan >> Jacob Lant >> Lee-Ann Coleman >> Miki Lentin >> Sean Martin >> Scot McKendrick >> Luke McKernan >> Philip Michel >> Ronald Milne >> Joanna Newman >> Barbara O'Connor >> Richard Ranft >> Martin Reagan >> John Rhatigan >> Elfrida Roberts >> Ben Sanderson >> Jane Sartin >> Matthew Shaw >> Helen Shenton >> Allan Sudlow >> Phil Spence >> Suvi Kankainen >> Pauline Thomson >> Sophie Villiers >> Richard Wakeford >> David Way >> Amanda Wyburn. The first and last pages of the synthesis are adapted (by JLJ, with apologies) from the design for the conference programme created by John Overeem of the British Library.

Carl Wilson and Paul Wheatley are thanked specifically for helpfully making available a local and early instance of the Planets system.

Many thanks to immediate manuscripts colleagues who kindly helped in picking up additional enquiries, audit work, manuscript inspection and rotas: >> Justin Clegg >> Kathleen Doyle >> Rachel Foss >> Elliot Fountain >> William Frame >> Juan Garces >> Frances Harris >> Julian Harrison >> Arnold Hunt >> Michael St John-McAlister >> Rachel Stockdale. The Access for eMSS team within the internal Digital Manuscripts Project recently played a specific practical role: Jamie Andrews >> Helen Broderick >> Katrina Dean >> William Frame >> Tim Hadlow >> Frances Lill >> Bill Stockting >>

The project is especially grateful to the following people for their support, ideas or enthusiastic interest: >> Victoria Bellotti >> John Blythe >> Peter Bright >> Maxine Clarke >> Adrian Cunningham >> Joy Davidson >> Adam Farquhar >> Erika Farr >> William Bradley Glisson >> Simson Garfinkel >> Timo Hannay >> Wendy Hall >> Frances Harris >> Peter Hirtle >> Matthew Kirschenbaum >> Cal Lee >> Clifford Lynch >> Cathy Marshall >> Kieron O'Hara >> Naomi Nelson >> Michael Olson >> David Pearson >> Mila Pollock >> Richard Ranft >> Andreas Rauber >> Gabriela Redwine >> Elfrida Roberts >> Helen Shenton >> Will Snow >> Susan Thomas >> Dave Thompson >> Jeffrey van der Hoeven >> Raymond J. van Diessen >> Natalie Walters >>

Professors Tony Sammes and Brian Jenkinson of the Defence Academy of the United Kingdom and Cranfield University are owed special thanks for sharing their experience and expertise in pertinent aspects of advanced computer forensics.

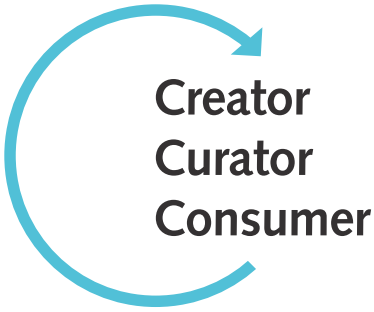
It was a distinct pleasure and great privilege to meet and interview Gordon Bell in person and to obtain *via* Skype and webcam the thoughts and advice of >> Chris DiBona of Google >> Brewster Kahle of the Internet Archive >> Jimmy Wales of Wikipedia.

The authors thank all the team members for their participation and help: Jamie Andrews >> Andrew Charlesworth >> Alison Hill >> Kristian Jensen >> Rory McLeod >> David Nicholas >> Rob Perks >> John Tuck >> Paul Wheatley >> Lynn Young. A very special thank you to Jennie Patrice, Alison Faraday and John Overeem for their hard work during the Digital Lives conference.

Finally the authors would like to pay tribute to Neil Beagrie who initiated the project and did so much to get it off the ground.

A special thank you too to the creators and their families who make their personal archives available to repositories for scholarly and scientific research.

The W. D. Hamilton Archive: some of its computer media



Arts & Humanities Research Council

Grant Number BLRC 8669