

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/235272899>

# The personal curation of digital objects: A lifecycle approach

Article in *Aslib Proceedings* · July 2009

DOI: 10.1108/00012530910973767

---

CITATIONS

32

READS

432

3 authors, including:



Peter Williams

University College London

121 PUBLICATIONS 2,439 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Using QR codes to aid accessibility in a museum [View project](#)

Project

British Library Personal Archives project [View project](#)



AP  
61,4

# The personal curation of digital objects

## A lifecycle approach

Peter Williams

*CIBER, Department of Information Studies, University College, London, UK*

Jeremy Leighton John

*Department of Western Manuscripts, Directorate of Scholarship & Collections,  
The British Library, London, UK*

Ian Rowland

*CIBER, Department of Information Studies, University College, London, UK*

340

Received 3 November 2008

Revised 4 December 2008

Accepted 1 January 2009

### Abstract

**Purpose** – This paper aims to set out a coherent intellectual framework to help to better understand how people create, organise, manage, use and dispose of their personal digital archives. The context for this is the increasing volume and diversity of digital information objects being captured and stored by individuals in their personal capacities and the need to find ways to preserve this material for posterity.

**Design/methodology/approach** – The research presented here is based on literature analysis, the questions having been informed by an earlier series of in-depth interviews. The approach taken is to synthesise key concepts from the computer science, information management, and archives and records management literatures. Key concepts from the existing literature in computer science, information management, and archives and records management were elicited and synthesised to create a coherent document lifecycle narrative.

**Findings** – Individuals exhibit great diversity in terms of personal information management and digital archiving practice at just about every point in the digital information cycle: much more so than is the case in formal repositories. Practices exhibited are not always conducive to efficient document management. This represents a very keen challenge for professional curatorial practice.

**Practical implications** – Little is known about how individuals manage digital information resources in their personal capacity, outside of their corporate or institutional employment. Yet both individuals on their own and professional curators on behalf of repositories are increasingly being faced with the challenge of how to deal with digital media. It is hoped that this paper will contribute to a growing debate in this area.

**Originality/value** – Personal information management from the perspective of personal digital archives is a surprisingly under-researched area and the proposed model adopts an archival information lifecycle approach. It seeks to apply and promote an archivally-oriented personal information management.

**Keywords** Information facilities, Digital storage, Information exchange, Collections management, Individual psychology

**Paper type** Research paper



---

## Introduction

Individuals are capturing and storing an ever-increasing amount of digital information about or for themselves, including documents, articles, portfolios of work, digital images, and audio and video recordings. People can now correspond by e-mail, have personal web pages, blogs, and electronic diaries. Computer processing, storage, and software tools are increasing in power year on year. Many issues arise from this increasingly empowered landscape of personal collection, dissemination, and digital memory, which will have major future impacts on librarianship and archival practice. Not only the media and formats but also the contents of works created by individuals are changing. Consider the example, in the history of science, of the fate of the field notebook as research councils move towards supporting more standardised forms of recording eScience.

Personal information collections and management have been the focus of attention of researchers for many years, particularly in the field of Human-Computer Interaction (e.g. Malone, 1983). However, the rapid rise in digital applications and storage capacity has both stimulated interest in this field and also opened up a new research area within it highlighted by Beagrie (2005). In particular, the way individuals use their personal computers at home (or, more accurately, away from work and the corporate or institutional environment) is ripe for detailed investigation.

We adopt the term “personal digital archives” to refer to these informal, diverse, and expanding memory collections created or acquired and accumulated and maintained by individuals in the course of their personal lives, and belonging to them, rather than to their institutions or other places of work. These digital collections are essentially the digital equivalent of the “personal papers” manifested in contemporary historical archives and manuscript collections and the individual items within can be referred to as “personal digital objects” or as “eManuscripts”, a term used by the British Library (BL) since 2000. Archivists working with personal papers tend to eschew publications but even these will be embraced if there are annotations or if there is a valuable compilation of obscure publications or of newspaper cuttings; and in any case a list of the publications held by an individual is invariably welcome. A little historical perspective may help clarify terminology: just as a collection of “books” (which might include journals and magazines) held by an individual is referred to as a personal library, so the “papers” held by an individual are referred to as a personal archive, and its development involves many informational aspects of a person’s life from the passive receiving of letters, the selective retaining and discarding of notes, and the creating of diaries, essays and photographs. Conceptually, it is useful to contrast the formal, carefully edited publication with the personal paper which may be unfinished, informal, candid, and in some sense closer to the individual. In the digital era, the distinction is becoming increasingly blurred – though not entirely absent.

At the British Library, interest in personal digital archives was stimulated by the arrival of the W.D. Hamilton Archive in 2000. It contained a wide range of computers and computer media from five-hole paper tapes and punched cards through to optical disks and hard drives requiring the research and development of new techniques and methodologies for their curation (Summers and John, 2001; Grafen, 2004; John, 2005, 2008). In 2003 the library created for the first time a post for a Curator of eManuscripts.

---

Similarly, the Paradigm (Personal ARchives Accessible in DIGital Media) project involving the research libraries of the University of Oxford and University of Manchester set out “to survey and accession sample hybrid personal archives (i.e. personal archives with digital and physical components)” (Paradigm, 2008, p. 16) from a representative sample of United Kingdom politicians (Thomas and Martin, 2006; Paradigm, 2008). This was undertaken to explore the issues involved in preserving digital private papers. Project archivists first investigated the structure of politicians’ personal archives in an attempt to characterise hybrid personal archives. The project also considered the impact that current technologies might have on the content and structure of the average personal archive.

Currently, the personal digital archive is the focus of a major study, the Digital Lives Research Project, which is funded by the Arts and Humanities Research Council and led by the British Library. It focuses on such collections and their relationship with research repositories. It brings together expert curators and practitioners in digital manuscripts, digital preservation, modern literary manuscripts, web-archiving, history of science, and oral history from the BL with researchers in the School of Library, Archive and Information Studies at University College London, and The Centre for Information Technology and Law at the University of Bristol. Interim results from the study have been published elsewhere (Williams *et al.*, 2008).

The Digital Lives project aims to explore how individuals use, organise and manage their digital collections. This will help to inform not only curators and archivists who will be entrusted to manage important personal archives for posterity but also people generally who may want (as seems increasingly so) to look after their personal digital archives for themselves and for family members. We are seeking to:

- clarify our understanding of an enormously complex and changing environment;
- identify the key issues; and
- start to evaluate radical new practices and tools that could advance curatorial processes in the future.

This paper focuses on the first stage of the digital archive process – individuals’ own digital behaviour and their build-up of a digital archive, and concentrates on the literature carried out to inform our fieldwork. It reviews some of the key literature on Personal Information Management, with particular emphasis on digital collections, and – following fieldwork reported elsewhere (Williams *et al.*, 2008) – looks at the management of information from the perspective of a “lifecycle”. This entails examining the stages of the life histories of personal digital objects as a chronological narrative, beginning with their creation, acquisition and manipulation, continuing with decisions about their retention, disposal and backing-up, and organisation into, for example, a folder structure or directory – and, finally, their long-term storage or accommodation. All of these stages mirror the formal activities of archiving and records management – except that in this case the emphasis is on the individual rather than the institution, and on personal “lay” practices rather than professional ones. The investigators believe that this chronological lifecycle approach illuminates personal digital information practice and informs the wider project to an extent that perhaps a thematic one might not do.

Research into Personal Information Management has often been motivated by the desire to design successful software and hardware that meets specific functional

---

objectives; and in many cases is commercially driven. The personal archive approach explored in the present paper is modelled on the archival lifecycle and takes a holistic view of the entire collection of personal digital objects within the control of an individual. Moreover, an archivally-oriented PIM has objectives that are 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 uncover the sequence of events and thoughts that led to a discovery or insight and a curator of scientific manuscripts takes this into account. The desire from the archival perspective is for information to be interpretable and maximally useful for future as well as current generations.

Much PIM research has sought to improve or understand functionality based on present conditions; but an archivally-oriented PIM is concerned not only with the use of currently active digital objects but also with the accurate retrieval of past events through digital objects that were created, acquired, amended or organised in the past. However, in a similar vein, usability research increasingly has employed aspects of past behaviour and activity (i.e. elements of personal information) as a means of improving current performance (e.g. Balazinska *et al.*, 2007; Bellotti *et al.*, 2008). A dynamically accessible personal digital archive might be seen as the ultimate resource for identifying a suitable personal profile that matches a device to an individual in order to optimise its effective use.

There are innumerable references to “lifecycles” and “cycles” of information in the computer science, information science, management and archival literatures but in relation to personal digital objects and the activities of individuals during their personal lives there has been much less prominence given to a personal digital information lifecycle that gives the appearance of emulating a prescribed archival or curatorial process.

Roes (1994) describes the “information cycle” in terms of the publication process with regard to scholarly works:

[...] in which the actors are scholars, both producing and consuming information; primary and secondary publishers which organise the editing and refereeing process ... and libraries which select titles ... and supply the information on demand.

The “information cycle” proposed in the present paper relates to personal documents as managed in an individual’s own collection. Clearly, the addition of the adjective “digital” makes the cycle applicable specifically to electronic documents – which have a different cycle in that they can be replicated exactly with multiple copies existing in diverse locations, can be disseminated as replicates and kept at the same time, and can be converted into more or less equivalent entities that exist in different digital formats.

The huge 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 formal more uniform “cycle” or “system” takes over. With regard to digital artefacts, one such system is The Open Archival Information System (OAIS) or model, “a reference model that has been widely accepted by the digital preservation community as a key standard for digital repositories” (Paradigm, 2008). This form of document cycle is beyond the scope of the current review.

The present authors view a digital object cycle as one that embraces the document or other electronic artefact from gestation or acquisition to long-term (personal)

---

retention or archival storage. Between these poles there is the “short-term” document manipulation and management stage, when artefacts are said to be “hot” or “warm” (Sellen and Harper, 2002) rather than “cold”, as they are once set aside or completed.

This exploratory review categorises the digital document or eManuscript cycle into three phases:

- (1) obtaining information;
- (2) short-term information management; and
- (3) long-term personal archiving and preservation.

The literature on each of these is reviewed briefly, following an introduction to Personal Information Management (PIM) and then personal digital archives.

### **Personal digital information management**

#### *Personal information management*

Bruce *et al.* (2004) contends that “individuals create a personalised subset of the *information world* that they can use when they are faced with information needs. This . . . is a personal information collection”. A personal information collection is defined as the space we turn to first when we need information to do a task or pursue an interest. It is a collection of information sources and channels that we as individuals have acquired, cultivated, and organised over time and in response to a range of stimuli. Such information may consist of files, e-mails, web bookmarks, papers, sound collections and so on. Jones *et al.* (2005, p. 457) prefers to use the term “personal space of information”, describing this as “all the information items that are . . . under an individual’s control”. Finally, for Bergman *et al.* (2004) a collection of personal information is a self-contained set of items, usually sharing a particular technological format and accessed through a particular application. Maintaining this collection, in terms of filing, indexing and access has been described as “the management of information by an individual in support of his/her roles and tasks” (Bergman *et al.*, 2004, p. 1598). It is commonly called Personal Information Management (PIM).

PIM is now a topic of much interest because of the size and variety of individuals’ personal collections.

[PIM] refers to both the practice and study of the activities a person performs in order to acquire or create, store, organise, maintain, retrieve, use and distribute information needed to complete tasks (work-related or not) and fulfil various roles and responsibilities (for example, as parent, employee . . .) (The) . . . information items . . . are stored for later . . . and repeated re-use (Jones, 2007, p. 453).

Other definitions of PIM include that of Lansdale (1988, p. 55), being “the methods and procedures by which we handle, categorise and retrieve information on a day-to-day basis”; and Barreau and Nardi (1995, p. 327) for whom PIM was the somewhat narrower “system developed by or created for an individual for personal use in a work environment”.

According to Jones (2007) (albeit citing Lansdale, 1988), the phrase “Personal Information Management” was “apparently” first used in the 1980s, along with a number of appropriate tools to help with, for example, appointments and scheduling, to-do lists, phone numbers, and addresses.

---

As Jones (2004) points out, “a person may maintain several separate, . . . inevitably inconsistent, organisational schemes for electronic documents, paper documents, e-mail messages and web references”. Furthermore, “the number of organisational schemes may increase if a person has several e-mail accounts, uses separate computers for home and work, uses a PDA or a smart phone”.

As Jones implies, PIM has become predominantly concerned with the management of electronic information, to which this review now turns.

### *Personal digital archives*

Bergman *et al.* (2004) contends that PIM is a fundamental aspect of computer-based activity - millions of computer users manage personal information (e.g. files, e-mail, contacts, bookmarks, reminders) every day to support their work and leisure needs. Beagrie (2005) used the phrase “personal digital collections” to refer to “the informal, diverse, and expanding digital memory collections accumulated and maintained by individuals”. Although this is essentially the digital equivalent to “personal papers”, or “personal digital archives”, a closely related concept found in archival and curatorial literature, it can be used to emphasise the inclusion of digital publications held by the individual (a convergence of personal library and personal archive). As these terms focus on individuals, they exclude, for example, information on them that may be held in government sources such as census records or reviews of an individual’s work created and maintained by third parties.

Jones *et al.* (2005) contend that: “it is very important to note that the personal information collection also includes information sources and channels that have been *left*” (original emphasis) – in other words, bookmarks and the electronic pages to which they lead should be considered part of a digital collection. This view is given strength by the literature showing that many people do not “download” electronically discovered information, as they know they will be able to find it *in situ* when needed (Bruce *et al.*, 2004; Jones *et al.*, 2005). Jones (2007) suggests that a personal collection (termed by him as a “Personal Space of Information”, or PSI, which can include hardcopy) does not include the web pages visited but does include copies made (or that are cached on a person’s computer) and the bookmarks created to those pages. Jones further states that, in addition, a PSI “also includes applications, tools (such as a desktop search facility) and constructs (e.g. associated properties, folders, . . .) that support the acquisition, storage, retrieval and use of the information” (Jones, 2007, p. 463). Metadata – system generated data about a document’s creation, location, size, etc. may also be considered part of a digital collection, as it may be required in retrieval, especially where a vast amount of documentation is being generated – an extreme example being efforts to capture every moment of one’s life (Gemmell *et al.*, 2006a, b).

The inclusion of links to external information sources – and those sources themselves – and of system-generated metadata, leads to the question of what digital artefacts constitute and do not constitute a digital collection. Even with hardcopy, there is evidence that individual collections contain much material that is only kept for future reference and may never even be consulted (Whittaker and Hirschberg, 2001). With the ever-increasing storage capacity of digital media, deciding what constitutes a meaningful collection can be philosophically somewhat challenging. Clearly, this is of some relevance for curators taking individual archives into the holdings of long-term

repositories; but in general if it is in the possession of an individual it is deemed part of the personal archive (or personal library).

Deken (1999) contrasts digital collections with hardcopy. She points out that electronic records “are quite different from . . . paper-based records in a number of significant ways”. Although they are not “human-eye-readable” they do have many advantages. They are, for instance, “much more compact, . . . easier to create, alter, and transmit; and . . . in some ways, easier to store”. Another difference is, of course, that electronic documents include the system-generated data already mentioned, and which in part be perceived as the computer’s own “finding and retrieval aids”.

One interesting area within the topic of PIM is e-mail. As Bellotti *et al.* (2003, p. 345) state, “An increasing body of literature points to the importance of e-mail as a task management resource”. The authors cite Mackay (1988) who, as far back as the late 1980s, was describing how e-mail supports “a variety of time and task management activities”. They also mention Whittaker and Sidner (1996), who show that the e-mail inbox is “a repository of ‘to-dos’, ‘to-reads’, items of ‘indeterminate status’ and ‘ongoing correspondence’” (Bellotti *et al.*, 2003, p. 345). Finally, the authors’ own work is cited (Ducheneaut and Bellotti, 2001), in which e-mail is “transforming into a ‘habitat’, the central place from which work is received, managed, and delegated in organisations” (Bellotti *et al.*, 2003, p. 345).

There appears to be four contributory factors in the growth of individuals’ digital collections. These are:

- (1) the growth of digital storage capacity;
- (2) the availability of content-creating tools;
- (3) the ever burgeoning availability of information accessible via the world wide web; and
- (4) people’s natural creative instincts.

**The growth of digital storage capacity**

According to Bell and Gemmell (2007), “today, a \$600 [£400] hard drive can hold . . . Enough [data] to store everything you read . . . all the music you purchase, eight hours of speech and ten pictures a day for the next 60 years.” Beagrie (2005) describes the continuing increase in computer power as “exponential”, invoking “Moore’s Law” (from a seminal article by Gordon Moore which first flagged-up the implications of the rise in computer power) (Moore, 1965) to argue his case. In a recent study on information collection, Sweeney (2001) showed the extent to which disk storage per person was increasing by calculating the amount of rigid disk drive space sold in a year divided by the adult world population (see Table I).

Two white papers attempt to quantify the digital universe: the amount of information created, captured and replicated (Gantz, 2007, 2008). It is estimated that in 2007 all of “the empty or usable space on hard drives, tapes, CDs, DVDs, and memory

**Table I.**

	1983	1996	2000
Increase in disk storage space per person	90	160,623	2,829,288
Global disk storage per person (MB/person)	0.02	28	472



---

(volatile and non-volatile) in the market equaled 264 exabytes". At the same time it was suggested that the digital universe in 2007 was 281 exabytes or 281 billion gigabytes. Thus the created information exceeded, for the first time, the available storage, which emphasises that not all information that is created and transmitted is stored. Nonetheless, Gantz (2008) predicted that 1,082 exabytes of storage would ship between 2007 and 2010; and noted that the segment of the storage consumption that IDC had previously underestimated is that for personal data protection:

As consumers generate more and more of the world's digital content, they are finally coming to understand the need to preserve their information heirlooms.

### The availability of content-creating tools

As computer power has increased, so have the number of information and communications technology (ICT) applications and variety of hardware (mobile phones, iPods, digital cameras, etc.) that can be used to make content. There are also a myriad of organisations facilitating and/or encouraging people to be digitally creative. Trendwatching.com (2007) offers several examples:

- *Hewlett-Packard*: "spending US\$300 million on a campaign telling consumers it's all about 'You', and 'You' should be taking pictures, and sharing them and forwarding and printing them, AND posting photographic essays on a HP web site".
- *Blogger*: "offering . . . instant communication power by letting you post your thoughts to the web whenever the urge strikes".
- *Xingtone.com*: "letting consumers compose their own ringtones".
- *Xlibris*: which specialises in offering the opportunity to budding writers to self-publish, providing "all of the tools you need to publish your book quickly, easily and affordably . . . (including) Design, Production and Publishing Services, as well as online distribution . . . and Marketing Products" (Xlibris, 2007).

Perhaps not surprisingly, Microsoft is also promoting digital creativity, not least in currently exhorting its customers to view a software suite called Expression Studio, a "powerful suite of professional design tools that enable you to take your creativity to a whole new level" (Microsoft, 2007), and Apple has been very successful with its iLife software.

### The availability of information

Statistics on the growth of the internet invariably suggest continuing rapid growth. One site, for example, the well known "Hobbes Internet Timeline" (Zakon, 2006), charts the number of hosts (a computer system with registered IP, or Internet Protocol address), and shows an increase from 16,729,000 to 439,286,364 in the ten years from July 1996 to July 2006 – a growth of 2,600 per cent. Similarly, the online magazine Netcraft (2008) report a growth in internet sites of 3.1 million, between March and April 2008. The ubiquity of internet search engines, and perhaps in particular, the popularity of Google – whose dominance of the search market is attested by its 75 per cent market dominance in the UK (Wakefield, 2007) – together with this growth of digital information have made it much easier for individuals to accumulate their own digital store.

---

### People's natural creative instincts

The take up of creative digital tools demonstrates that there is a strong appetite for people to be creative online. According to the Pew Internet and American Life Project (Lenhart *et al.*, 2004), 44 per cent of adult internet users based in the USA have used the internet to “publish their thoughts”, post pictures, share files and “otherwise contribute to the explosion of content available online”, and 21 per cent of internet users have posted photographs to web sites. A total of 13 per cent of internet users maintain their own web sites, and “around 7 per cent” have web cams operating to allow other internet users to see live pictures of them and their surroundings.

American research (iProspect, 2007) has shown that social networking services that allow user-generated content such as comments, reviews, feedback, ratings or own dedicated pages were attracting steadily growing traffic. Several are visited at least monthly by roughly a quarter of (US) internet users, including: MySpace (28 per cent), and YouTube (20 per cent), although the attraction often appears to be to view the work of others – it is not yet common practice among the vast majority of the online user population to post their own material. Nevertheless, such sites do demonstrate enormous levels of user input. Flickr (a photo-sharing site) logs uploads, and consistently records over 2,000 per minute[1]. It is outside the scope of this paper to discuss the value of such user content, but worth noting that not everyone greets the innovation with open arms. Keen (2007), for example, sees the rise of amateur content as a threat to traditional culture, diminishing the quality of civil public discourse, encouraging plagiarism and stifling genuine creativity.

Gantz (2007) predicted that 70 per cent of the digital universe will be created by individuals in 2010. Gantz (2008) subsequently observed that less than half of that component of the digital universe existing in 2007 and created by individuals is explained by user activities; the remainder constitutes a “digital shadow” comprising web search histories, financial transactions, surveillance images and so on. EMC Corporation provides a tool for estimating how fast an individual's personal digital universe is expanding. The Digital Footprint Calculator is available at: [www.emc.com/digital\\_universe/downloads/web/personal-ticker.htm](http://www.emc.com/digital_universe/downloads/web/personal-ticker.htm)

### The digital information cycle

#### *Obtaining information*

There are three ways in which individuals obtain and build up information:

- (1) actively seeking or acquiring it;
- (2) passively seeking or receiving it; and
- (3) actively creating it themselves.

(A fourth process might be referred to as the “Passive creation of information’ in that computer operating systems and application software routinely and automatically create digital objects and metadata in association with the other three activities: it includes the computer's propensity to cache files for example, as already mentioned. The present paper will not discuss this process further but the resulting files and content represent a very important element of a personal digital archive.)

There is a huge body of literature concerned with information seeking behaviour, a topic that is somewhat peripheral to this review, which is more concerned with information once it is in a collection, and so this part of the current review points only

---

to some key texts. By contrast, there is a lack of literature on the process of creating information, in terms of its manipulation during the creation process. More can be found about how documents and other digital artefacts are stored and backed-up during their longer-term development – a topic discussed under the heading of “Short-term information management”.

*Actively sought information behaviour.* Case (2006) provides a comprehensive overview of information seeking, while Wilson (1997) shows how the subject draws from and influences other disciplines. The extensive body of work on the subject includes literature on: “everyday” information seeking (e.g. Savolainen, 2005); scholarly information seeking in various subjects (e.g. Bichteler and Ward, 1989; Voorbij, 1999; Siegfried *et al.*, 1993; Dalton and Charnigo, 2004); and information-seeking by the public on general subjects such as health (e.g. Gollup, 1997), or education (e.g. Nicholas and Marden, 1998).

*Passive reception of information.* As Marshall and Jones (2006, p. 66) remark:

[...] not all of the information that comes into our purview is actively sought to meet a clearly defined need. . . . it may have potential merit as a reminder, for its evocative qualities, for its educational value, for the ideas it spurs, for its potential utility as a reference, or as something to share.

Ross (1999, p. 784) notes that generally in information seeking or information behaviour research, individuals are studied who “experience a ‘problem situation’ and then formally initiate the search process by querying one of our systems: a reference service, an online catalogue, a database, a collection of books”. Noting the “emphasis on goal-directed, problem-solving information”, Ross (1999, p. 785) goes on to add: “We know, in fact, that in the course of everyday living people constantly . . . encounter, and use textual information without ever posing a formal question.”

Williamson (1998) calls the latter “incidental information acquisition”, to be distinguished from “purposeful information seeking”. Williamson studied elderly people and their acquisition and use of information. She concluded that while respondents purposefully sought information in response to perceived needs, they also monitored their world, and acquired information that they did not seek – some of which simply “cropped up” whilst going about their daily activities such as reading a newspaper or talking to friends.

One interesting view of encountered information comes from the world of systems analysis. Peter Checkland (e.g. Checkland and Holwell, 2006; Checkland and Winter, 2000) has coined the term “capta”. Checkland (1999, p. 54) explains that “we select certain items of data out of the mass potentially available. They are selected as being relevant to our concerns.” It is the “data” that we select which are “capta” (from the Latin *capere*, to take). Bawden (2001, p. 95) adds that *capta* is “the transformation from data to information as involving the addition of context, and hence meaning”. Here, one could argue that unread or unevaluated documents represent “data” (in that they only acquire meaning, and hence become information, when read), and that resultant retention/discarding choices render documents retained as “capta”, to be acted upon later.

The issue of deciding what to do with encountered information, assuming it is in a storable form (i.e. and not just something heard on the radio, etc.) represents a key challenge for the field of personal information management (PIM) (Marshall and Jones, 2006). This is the topic of the later section on “Short-term information management”.

*Information creation.* The process of creating information does not appear to have been studied in terms of the creator’s information behaviour at the point of creation (for

example, preferences for hard copy or digital media and the integration of the two, tracking progress in electronic documents, over-writing or renaming files in the course of a project undertaken over a number of sessions). Trace (2007, p. 142) takes a different view. Noting that information creation “has been largely overlooked in the LIS [Library and Information Science] literature”, Trace suggests that research into information creation “focuses on how and why people are socialised to create information in various contexts”. In this regard, he found that understanding how to create and manage documents constituted “a fundamental part of student affiliation; enabling students [to become] competent members of a school community” (Trace, 2007, p. 142). They did this in the sense that:

[...] students learned that school documents ... were entities that controlled, reflected, and organised their environment .. [e.g. timetables, rules, etc.]; that documents could serve as evaluative instruments [student documents were assessed, and report cards etc. recorded evaluations of students] that documents could hold students accountable [progress charts etc.]; ... and that documents had a role to play in managing social relationships within the classroom context (Trace, 2007, p. 150).

The latter refers to collaborative working, sharing documents and creating, modifying and storing them in a way that gained approval (or opprobrium) from peers and teachers.

Another way of looking at information creation is given by Ikujiro Nonaka and colleagues (Nonaka *et al.*, 1996, p. 205) who were concerned not so much with “documents” but with passing “tacit” and “explicit” knowledge between colleagues throughout companies. Tacit knowledge consists of “intuitions, unarticulated mental models, or embodied technical skills”, and “explicit knowledge” of “meaningful set of information articulated in clear language including numbers or diagrams” skills. Four patterns of interactions exist that help create and transfer information:

- (1) “socialisation (from individual tacit knowledge to group tacit knowledge);
- (2) externalisation (from tacit knowledge to explicit knowledge);
- (3) combination (from separate explicit knowledge to systemic explicit knowledge);
- and
- (4) internalisation (from explicit knowledge to tacit knowledge)”.

The authors conclude that “only human beings can take the central role in knowledge creation and that computers are merely tools, however great their information-processing capabilities are” (Nonaka *et al.*, 1996, p. 217).

This review now turns to the point in the cycle at which information – once created or acquired – is managed.

### **Short-term information management**

#### *Immediate retention decisions*

Immediately following the acquisition or creation of information the first choice has to be made – that of retention (for the short or long-term) or immediate rejection.

Bruce *et al.* (2004) suggests that:

[...] whenever an individual locates or encounters ... information ... s/he faces an essential choice, whether consciously exercised or not...: to keep the information, to leave the information where it is (it is useful but I can find it when I need it) or to ignore the information.

---

Thus, there are three possible actions:

- (1) *Keeping*, where “[An individual] identifies information that is useful and then engages in an act that relates to collecting, representing, indexing, cataloguing, classifying, and storing [it].”
- (2) *Leaving*: consisting of “information sources and channels that have been left (in situ). In this case, the individual . . . concedes that it is in a . . . place that can easily be located again.”
- (3) *Ignoring*: Where the individual judges that the information has no use or value to them, “either now or in the future”. Examples (though not supplied by Bruce) might be football results or the weather, acquired for instant information but not needed for future reference.

Two points may be made here. First, although not mentioned by Bruce, similar (if not equal) choices also apply to information created by the individual him or herself. For example, a person writing a business letter may decide to keep it (e.g. for future reference, or for later amending for a different recipient), or to delete it (or to delete earlier drafts if not the final copy). Second, the “delete” choice also applies to information acquired. “Leaving” suggests the material is remote from the individual – a web site, for example. Unsolicited information received – an article sent via e-mail by a colleague, for example – may be “left” (e.g. as an attachment until the message itself is actioned in some way), or may be actively deleted.

### Document and task management

Bondarenko and Janssen (2005) argue that document management is strongly related to task management. Common patterns of document use were found for participants across professions, differences in document management were noted particularly according to activity type. Specifically, “administrative” and “research” activities were analysed. Much is made of Sellen and Harper’s (2002) three types of documents:

- (1) those that are currently being used and/or written (“hot”);
- (2) those that are still referred to and may be updated (“warm”); and
- (3) those that are no longer used (“cold”).

Administrative activity entails documents coming in and out rapidly: forms, reports, letters, minutes, etc.

The transitions of a document from “cold” to “hot” and vice versa happen very often – some documents are only active until a short phone call is made and then can be filed away. In contrast, a “research” activity corresponds to a small variety of documents (articles and reports of different kinds are the most common types), but these documents stay “warm” for a long time (Sellen and Harper, 2002, p. 124).

Printing helps people to keep documents from various sources (e-mail, digital file folders, and paper copies) together. The transition from digital to paper is far easier than from paper to digital.

Bruce and colleagues (Bruce *et al.*, 2004) found a large number of methods for keeping encountered information. Each of the following was used by at least one of the participants of this study:

- (1) send an e-mail to oneself, with a URL referencing a web page;
- (2) send to others an e-mail that contains a web page reference;
- (3) print out the web page;
- (4) save the web page as a file;
- (5) paste the URL for a web page into a document;
- (6) add a hyperlink to a personal web site;
- (7) bookmark the web page;
- (8) write down notes on paper containing the URL and actions to be taken;
- (9) copy to a links toolbar so that the web address is always in view in the browser;
- (10) note URL in a task management or calendaring system which contains the URL and can be associated with a date.

“Refinding” is also a topic of interest in personal information management. E-mail messages (Whittaker and Sidner, 1996) and placing documents on the computer desktop (Barreau and Nardi, 1995) have been shown to be popular methods of finding information already acquired or created. By contrast, over 90 per cent of respondents interviewed in an “informal survey” by Jones (2007) indicated that they used Google Desktop Search (about which more later) only as a “last resort” after other methods of return had failed.

With regard to naming and grouping documents, Kwasnik (1989) maintains that the use to which something is put is often the basic level at which it will be classified. For instance, although a person may distinguish between books on various topics, books acquired at different times, and books of different formats and sizes, all of them may be physically as well as cognitively grouped as:

[...] “books used in teaching Anthro 101”. . . . A document’s intended use or purpose is often the first classificatory rule invoked. Documents may be further divided and organised, but the first cut is frequently determined by use or purpose, or use in combination with another dimension (Kwasnik, 1989, p. 210).

It is worth examining, in this section, attempts to create systems that integrate different document types and applications, for easier task and document management. According to Jones (2007) people frequently wish to group together information relating to particular ongoing or completed tasks, or simply view different types of items (e.g. e-mails, web pages and Word documents) together. Several problems become apparent in attempting these two actions. It may not be possible, for example, to put Word documents and e-mails into one folder readily with little effort. Even viewing different types of document (and corresponding application) can be problematic – a computer display may be filled with windows, often obscuring each other and each competing for our attention (Jones, 2007).

This leads to a consideration of how documents and applications can be linked or integrated, to facilitate more efficient usage and to obviate “fragmentation” problems that arise from using and storing different kinds of documents that relate to the same overall task or project.

Jones outlines three ways in which tasks and applications may be integrated:

- (1) integration through e-mail;

- 
- (2) integration through search; and
  - (3) integration through projects.

Regarding the first of these, e-mail has been shown to be used for:

- task reminding (by leaving information about current tasks there (Jones *et al.*, 2001). . . . or even by sending oneself an e-mail as a reminder);
- personal archiving;
- linking to useful information;
- storing information about completed tasks (e.g. Bellotti *et al.*, 2005); and
- accumulating contact information (e.g. Whittaker *et al.*, 2004).

Two information access problems arise from performing PIM functions in e-mail – fragmentation and lack of direct support for PIM functions (Whittaker *et al.*, 2006, p. 70). Fragmentation can occur where attachments or information are left in an inbox or other folder, and not transferred to a more appropriate directory; or where attachments are removed, so that the sender and the document are no longer linked. The fragmentation is to a significant extent a manifestation of the lack of PIM support. E-mail was not designed for these tasks, and support for organising messages and attachments is limited or rarely learned or used. Whittaker *et al.* (2006) discuss possible solutions to the problem, including automatic extraction of signature files into address books and better search facilities – the latter being considered a tool for PIM and the integration of documents generally.

The integration of powerful searching facilities may be gaining momentum through the success of the “Google” search engine, which offers a desktop computer search facility. In fact, Google is only one of several such “desktop search” applications (“Spotlight” of Apple computers was a pioneer) and fill a need in the market expressed neatly by Boutin (2004) who observed, “I can find anything online in under a minute, but it takes me days to find an e-mail address on my PC.” Desktop searches usually index Word, Excel, and PowerPoint files and e-mail, and “most of them search music, image, and video files, as well as web bookmarks” (Boutin, 2004).

The existence of this facility suggests that finding documents may be easier via a search engine than by organising directories. However, as already indicated, research by Jones (2007) suggests that this option is not popular. Indeed, Jones was involved in a previous study looking at this issue. In a study looking at the organisation of folders and folder hierarchies, researchers offered 14 participants a “simple search tool”: all but one declined. Reasons given for continuing to want folders fell into one of three categories:

- (1) *Trust*: there was an unwillingness to trust the efficacy of the search engine.
- (2) *Control*: wanting files to be in a place of one’s own choosing.
- (3) *Visibility/understandability*: being able to see the relationship between things, or being reminded of tasks.

With regard to integration through project management, both the highly prevalent Outlook of Microsoft and Entourage ([www.microsoft.com/mac/products/entourage2008/default.mspx#/interacting\\_entourage/](http://www.microsoft.com/mac/products/entourage2008/default.mspx#/interacting_entourage/)), a Microsoft initiative for

---

Apple computers, provide various functionalities together, namely: e-mail, calendar, address book, task list, note list, and a facility akin to a general project manager. Other initiatives are continuously emerging that aim to facilitate file organisation. One is the “Universal Labeler” (Jones *et al.*, 2005) which helps in the organisation of documents, e-mail messages and web pages via a system of labelling different documents and, in effect, creating a kind of “super shortcut maker” (Jones, 2007). There are several other integrated systems available that attempt to integrate and index disparate files and applications. These include:

- *AirSet* ([www.airset.com/](http://www.airset.com/)). Which is, according to its producers, a virtual computer: “a computer that lives in the internet and is accessible from any device with a web browser”. AirSet includes a PIM system / calendar / blogging / wiki / messaging / calendar /
- *Knowledge Workshop* ([www.lmsweb.com/main/index fla.shtm](http://www.lmsweb.com/main/index fla.shtm)). A freeware personal information manager which enables web pages, documents, local files, tasks, e-mails, notes, etc. to be managed from one interface.
- *EssentialPIM* ([www.essentialpim.com/](http://www.essentialpim.com/)). Which includes calendar, contact management, to do list and notes. It also includes synchronisation with Outlook, Windows Mobile devices, Palm, iPod, and Google Calendar.

Online tools also include, of course, the popular web services such as:

- [www.rememberthemilk.com](http://www.rememberthemilk.com) which facilitates task management; and
- [www.zoho.com](http://www.zoho.com) which provides multiple applications and makes it possible to have everything in one place from documents to presentations.

Another emerging development worth noting as an example of the kind of solution to be expected is that of Geambasu *et al.* (2007): HomeViews is a peer-to-peer system for building personal data management applications involving:

- views for organising files into dynamic collections;
- protected sharing of views across the internet; and
- integration into a user’s local environment.

#### *Long-term personal archiving, preservation and disposal*

The stage beyond the day-to-day use and manipulation of personal documents is that of archiving for the longer term. For Barreau and Nardi, 1995), “archived information has a shelf life of months or years, but is only indirectly relevant to the user’s current work”. According to Sue McKemmish (1996):

The functionality of a personal archive, its capacity to witness to a life, is dependent on how systematically we go about the business of creating our records as documents, capturing them as records (i.e. ordering them in relation to each other and “placing” them in the context of related activities), and keeping and discarding them over time (i.e. organising them to function as long-term memory of significant activities and relationships).

These decisions correspond to the activities of professional archivists.

A special consideration in the context of personal digital objects is the need to migrate digital objects to more current formats to guard against obsolescence (see, for example, John, 2008; Paradigm, 2008). This important topic will not be discussed



---

further in the present paper. This review will instead begin by considering the literature that relates to why people create and maintain a personal archive over the long-term.

Kaye *et al.* (2006) studied 48 academics' personal files, and concluded from examining contents and interviewing participants, that there were five main reasons for maintaining an archive and storing its contents, which the researchers labelled:

- (1) *Finding it later*: "Not surprisingly, one reason why people archive materials is so that they can easily find them later, a need which currently drives many of the design strategies for personal archiving tools" (Kaye *et al.*, 2006, p. 276). A distinction can be made between an active document or digital object, which may be in the process of being created or amended (having been retrieved from the personal store or recently acquired online, for example) and an inactive one that is residing in the store.
- (2) *Building a legacy*: Kaye *et al.* (2006, p. 277) found that some records were not kept for future use, but as a kind of "legacy" archive. "Essentially, these [records] were testament to the subject's 'life work', a self constructed permanent record of the achievements and movements of the archiver".
- (3) *Sharing resources*: Kaye *et al.* (2006, p. 277) "identified sharing as an archival goal common to many of the personal archives we visited which were accessed often and by many individuals".
- (4) *Fears of loss*: "a sense of anxiety or fear of a specific catastrophe involving the potential loss of irreplaceable information fuelled the drive for preservation. . . . this fear influenced the archive's physical structure, from regular personal backups to specialised folders" Kaye *et al.* (2006, p. 278).
- (5) *Identity construction*: the personal archive is also, according to Kaye *et al.* (2006, p. 279) "a kind of 'identity kit': materials reflect and describe the owner . . . Items in the archive serve as 'tokens', indicating who the archiver is and what they have achieved".

This research, as mentioned, reported on academics' archive creation. As Barreau and Nardi (1995) note, however, for most forms of employment, information that is not current has less relevance and utility than it might have for academics or researchers. "This [is] true for the managers, administrative assistants (etc.)." Their study found, contrary to that of Kaye *et al.* (2006), that their sample of professionals did not bother to archive for the long-term, as the information was unlikely to be useful in the future. Considerations uncovered by Kaye *et al.* (e.g. having an archive as a legacy or as an expression of personal identity) did not manifest themselves, but this might be because the researchers in this case considered only electronic documents. In part this mirrors the somewhat divergent philosophies and attitudes of, on the one hand, records managers of institutions, and, on the other hand, archivists of historical and literary documents derived from individuals such as distinguished scientists and writers.

As Barreau and Nardi (1995) note, not all documents end up being archived beyond the time when they might be considered "useful" or "warm". This leads to the issue of why people discard (as in actually "delete") electronic archives. There does not appear to be any work associated specifically with the issue, although there is some literature advising people about managing files, which includes decisions on deleting (e.g. Lanza,

---

2006). This is particularly true in the case of e-mail (e.g. Taylor, 2007), partly due to legal considerations; and Hookham (2007), for example, discusses the tension between keeping personal e-mail indefinitely to avoid falling foul of the Financial Services and Market Act and deleting it as soon as possible to comply with the Data Protection Act.

With regard to personal, individual and autonomous decisions regarding whether to delete documents, this can largely be inferred from work on document retention (or what Bruce *et al.* (2004) and Marshall and Jones (2006) term “keeping”)[2]. In this context, documents may be deleted if they:

- are not needed (e.g. there may be no need to find them later) (Kaye *et al.*, 2006);
- are not required by others (Kaye *et al.*, 2006);
- do not contribute to one’s “legacy” or one’s “identity construction” (Kaye *et al.*, 2006, p. 279); and
- can be easily found again elsewhere (a variation on Jones, 2004, who discussed the leaving of documents *in situ* when they were first found, rather than capturing and storing them locally and to be subsequently removed).

As noted earlier (e.g. Jones, 2007) it is now possible to build up a digital archive without needing to delete anything, so it should not be assumed that documents that are not needed (e.g. because they are not required for any purpose) will actually be deleted. Even with paper personal archives, some people, “hoarders”, are inclined to keep nearly everything: at any rate far more than might be ostensibly needed or useful, but a powerful motivation can be the desire to store “just in case” the objects are needed.

With that caveat, it may be worth concluding this section with reference to work that has been undertaken on hardcopy documents. Whittaker and Hirschberg (2001, p. 156) studied why hardcopy files of individual office-staff members were retained or discarded during the course of an office move. They found that people tended to discard 22 per cent of their paper files. Two main reasons emerged:

- (1) The data were obsolete, even if these were useful in the past, and of unlikely future benefit.
- (2) The data were unread: nearly a quarter of discarded data were “non-urgent” documents workers had simply never had time to read. As the researchers put it: “Rather than discarding once-valuable information that is now of little utility, much of what people discard is unprocessed information they have never properly evaluated”.

The latter comment has much resonance with Checkland’s (1999) concept of “capta”. Unread and therefore unprocessed data never acquires “capta” status and so is (after a long period) discarded. However, some information of this class is frequently retained because it cannot be easily separated from information that is recognised as valuable, and this applies to both paper and digital archives.

The concluding phase of the archival lifecycle is the retrieval of information from a long-term store for use by researchers and other consumers. Hoang *et al.* (2006) have designed a Virtual Query System of SemanticLIFE that seeks to direct emerging semantic web technologies towards the task of creating a personal information management system for an individual’s lifetime data. In the slightly different context of usability and ubiquitous computing, Crabtree *et al.* (2006) are developing procedures

for the exploitation of digital records, e.g. tools for replay that enable “faithful representations” of temporal sequences.

Clearly, the better and more careful the original capture of the information, with contextual information, the more effective and profound will be any improvements in reuse potential.

### **Conclusion**

This review has adopted the lifecycle approach to personal archives of professional archivists and curators, and has transferred it to the context of personal information management; and thus it looks at the personal archives of individuals away from any repository and draws a parallel with the activities within a repository.

At first glance this may seem a little contrived – after all, the personal archive of a living person is a dynamic entity. New objects are created; others are amended; some are discarded. Yet the parallel is in fact an apt one as well as a useful one: for increasingly, professional curators of contemporary personal archives are engaged in creative activities that supplement the original contents of an archive. Oral history programmes have long sought to produce new information that complements the contents of any existing archive of an individual. More recently, and using the same kind of modern technologies of digital capture available to people generally - digital video and audio and digital photography – curators at the British Library have started developing a series of Enhanced Curation activities. A recent example was the creation of a high resolution 360 degree panoramic view of the study where the late Poet Laureate Ted Hughes wrote some of his poetry; this image of a creative landscape is in a sense as much a part of Ted Hughes’ archive as the collection of his personal papers that joined the British Library’s rich literary holdings in 2008. There is the ongoing process of enabling and improving the potential for reuse.

The theoretical perspectives developed in this paper around personal information management and the digital information cycle are intended to stimulate debate, reflection and further research. Many archivists and curators of personal archives and contemporary historical papers and eManuscripts are already encountering and engaging with the challenges presented by digital media. We hope that this opening review will stimulate further reflection among these professionals on the practical implications of ideas and observations made as they negotiate their way with hybrid personal archives, and as they contemplate plans and actions for the day when personal archives almost entirely comprise digital objects. The paper should also provide much room for thought for those actively constructing their own digital identities and planning for the long-term preservation of their personal “digital memories”.

This is clearly a rich field for research as there are still major gaps in our knowledge, especially in regard to the strategies that individuals use to organise and manage their personal digital archives, their motivations for doing so, and how they subsequently locate materials in those archives? There is also a critical lack of empirical evidence concerning individual attitudes and behaviour towards data security, their experiences of catastrophic data loss, or the arrangements they may or may not have in hand in the case of death or sudden incapacity. In a more speculative vein, is there the possibility that different styles or underlying archetypes of personal information management and archiving practice exist and, if so, what form might such underlying behaviours take? These are all highly practical questions that need answers

if we are to plan effectively for the future. For this reason, the Digital Lives team has conducted two large-scale online surveys of members of the general public and of academics from a wide range of disciplines, building directly on the issues and frameworks discussed in this paper. The findings will be reported both in the literature and through the Digital Lives project web site ([www.bl.uk/digital-lives/](http://www.bl.uk/digital-lives/)).

### Notes

1. 2,607 photos had apparently been uploaded during the minute before the present author accessed the page at 14.36 on 03.10.07.
2. This contrasts with a large volume of work on document retention and destruction in curatorial literature (e.g. Bearman, 2006; Nakajima, 2006) where the issue is discussed in terms of institutional repositories.

### References

- Balazinska, M., Kwon, Y., Kuchta, N. and Lee, D. (2007), "Moirae: history-enhanced monitoring", *Proceedings of the Third Biennial Conference on Innovative Data Systems Research (CIDR), 7-10 January 2007, Asilomar, California*, available at: <http://db.cs.washington.edu/moirae/moirae-cidr2007.pdf> (accessed 14 April 2009).
- Barreau, D. and Nardi, B.A. (1995), "Finding and reminding: file organization from the desktop", *ACM SIGCHI Bulletin*, Vol. 27 No. 3, available at: [www.sigchi.org/bulletin/1995.3/barreau.html](http://www.sigchi.org/bulletin/1995.3/barreau.html) (accessed 23 March 2009).
- Bawden, D. (2001), "The shifting terminologies of information", *Aslib Proceedings*, Vol. 53 No. 3, pp. 93-8.
- Beagrie, N. (2005), "Plenty of room at the bottom? Personal digital libraries and collections", *D-Lib Magazine*, Vol. 11 No. 6, available at: [www.dlib.org/dlib/june05/beagrie/06beagrie.html](http://www.dlib.org/dlib/june05/beagrie/06beagrie.html) (accessed 15 August 2007).
- Bearman, D. (2006), "Moments of risk: identifying threats to electronic records", *Archivaria*, Vol. 62, Fall, pp. 15-46.
- Bell, G. and Gemmell, J. (2007), "A digital life", *Scientific American*, March 2007, available at: [www.sciam.com/article.cfm?articleID=CC50D7BF-E7F2-99DF-34DA5FF0B0A22B50&chanID=sa006](http://www.sciam.com/article.cfm?articleID=CC50D7BF-E7F2-99DF-34DA5FF0B0A22B50&chanID=sa006) (accessed 13 October 2007).
- Bellotti, V., Ducheneaut, N., Howard, M.A. and Smith, I.E. (2003), "Taking e-mail to task: the design and evaluation of a task management centered e-mail tool", in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003) April 5-10, ACM, Fort Lauderdale, FL*, pp. 345-352.
- Bellotti, V., Ducheneaut, N., Howard, M., Smith, I. and Grinter, R. (2005), "Quality vs quantity: e-mail-centric task-management and its relationship with overload", *Human-Computer Interaction*, Vol. 20 Nos 1-2, pp. 89-138.
- Bellotti, V., Begole, J., Chi, E.H., Ducheneaut, N., Fang, J., Isaacs, E., King, T.H., Newman, M. and Partridge, K. (2008), "Activity-based serendipitous recommendations with the Magitti mobile leisure guide", *Proceedings of the 26th Annual ACM Conference on Human Factors in Computing Systems (CHI 2008), 5-10 April 2008, Florence, Italy, ACM, New York, NY*, pp. 1157-1166.
- Bergman, O., Boardman, R., Gwizdka, J. and Jones, W.P. (2004), "Personal information management", *Proceedings of the CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ACM Press, Vienna pp. 1598-1599.

- 
- Bichteler, J. and Ward, D. (1989), "Information-seeking behavior of geoscientists", *Special Libraries*, Vol. 80 No. 3, pp. 169-78.
- Bondarenko, O. and Janssen, R. (2005), "Documents at hand: learning from paper to improve digital technologies", *Proceedings of the CHI '05 SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 121-130.
- Boutin, P. (2004), "Keeper finders: five new programs that let you search your hard drive without having a seizure", *Slate*, available at: [www.slate.com/id/2111643/](http://www.slate.com/id/2111643/) (accessed 27 October 2008).
- Bruce, H., Jones, W. and Dumais, S. (2004), "Information behaviour that keeps found things found", *Information Research*, Vol. 10 No. 1, available at: <http://InformationR.net/ir/10-1/paper207.html> (accessed 19 October 2007).
- Case, D. (2006), *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, 2nd ed., Academic Press, San Diego, CA.
- Checkland, P. (1999), "Systems", in Currie, W. and Galliers, B. (Eds), *Rethinking Management Information Systems: An Interdisciplinary Perspective*, OUP, Oxford, pp. 45-56.
- Checkland, P.B. and Holwell, S.E. (1999), "Data, capta, information and knowledge", in Hinton, M. (Ed.), *Introducing Information Management: The Business Approach*, Elsevier, London, pp. 47-55.
- Checkland, P.B. and Winter, M.C. (2000), "The relevance of soft systems thinking", *Human Resource Development International*, Vol. 3 No. 3, pp. 411-7.
- Crabtree, A., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M. and Brown, B. (2006), "Supporting ethnographic studies of ubiquitous computing in the wild", *Proceedings of the 2006 ACM Symposium on Designing Interactive Systems*, ACM, University Park, Pennsylvania, PA.
- Dalton, M.S. and Charnigo, L. (2004), "Historians and their information sources", *College & Research Libraries*, Vol. 65 No. 5, pp. 400-25.
- Deken, J.M. (1999), "Electronic record keeping: an introduction", paper presented at the US Department of Energy Records Management Conference, Orange, CA, 17-20 May, available at: [www.slac.stanford.edu/~jmdeken/papers/SLAC-PUB8152.html](http://www.slac.stanford.edu/~jmdeken/papers/SLAC-PUB8152.html) (accessed 1 November 2007).
- Ducheneaut, N. and Bellotti, V.E.-m. (2001), "mail as habitat: an exploration of embedded personal information management", *Interactions*, Vol. 8 No. 5, pp. 30-8.
- Gantz, J.F. (2007), *The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010*, IDC (International Data Corporation) White Paper, available at: [www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf](http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf)
- Gantz, J.F. (2008), *The Diverse and Exploding Digital Universe. An Updated Forecast of Worldwide Information Growth through 2011*, IDC (International Data Corporation) White Paper, available at: [www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf](http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf)
- Geambasu, R., Balazinska, M., Gribble, S.D. and Levy, H.M. (2007), "HomeViews: peer-to-peer middleware for personal data sharing applications", *Proceedings of the SIGMOD International Conference on Management of Data, 11-14 June 2007, Beijing, China*, pp. 235-246.
- Gemmell, J., Bell, G. and Lueder, R. (2006a), "MyLifeBits: a personal database for everything", *Communications of the ACM*, Vol. 49 No. 1, pp. 88-95.

- 
- Gollup, C.J. (1997), "Health information-seeking behavior and older African American women", *Bulletin of the Medical Library Association*, Vol. 85, pp. 141-6.
- Grafen, A. (2004), "Willam Donald Hamilton", *Biographical Memoirs of Fellows of the Royal Society*, Vol. 50, pp. 109-32.
- Hoang, H., Tjoa, A. and 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, Studia Informatica Universalis*, pp. 39-45.
- Hookham, J. (2007), "Regulatory compliance: e-mail -delete it all or keep it all?", *Records Management Bulletin*, Vol. 136, pp. 8-11.
- iProspect (2007), "Social networking user behaviour study", available at: [www.iprospect.com/about/researchstudy\\_2007\\_socialnetworkingbehavior.htm](http://www.iprospect.com/about/researchstudy_2007_socialnetworkingbehavior.htm) (accessed 3 October 2007).
- John, J.L. (2005), "Because topics often fade: letters, essays, notes, digital manuscripts and other unpublished works", in Ridley, M. (Ed.), *Narrow Roads of Gene Land, Vol. 3. Last Words*, Oxford University Press, Oxford, pp. 399-422.
- John, J.L. (2008), "Adapting existing technologies for digitally archiving personal lives. Digital forensics, ancestral computing, and evolutionary perspectives", *Proceedings of The Fifth International Conference on Preservation of Digital Objects Joined Up and Working: Tools and Methods for Digital Preservation, 29-30 September 2008*, The British Library, London, available at: [www.bl.uk/ipres2008/ipres2008-proceedings.pdf](http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf) (accessed 15 March 2009).
- Jones, W. (2004), "Finders, keepers? The present and future perfect in support of personal information management", *First Monday*, available at: [www.firstmonday.dk/issues/issue9\\_3/jones/index.html](http://www.firstmonday.dk/issues/issue9_3/jones/index.html) (accessed 27 October 2008).
- Jones, W. (2007), "Personal information management", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Information Today, Medford, NJ, pp. 453-504.
- Jones, W., Bruce, H. and Dumais, S. (2001), "Keeping found things found on the web", *Proceedings of the 10th International Conference on Information and Knowledge Management, Atlanta, GA, November 5-10*, ACM Press, New York, NY, pp. 119-126.
- Jones, W., Munat, C. and Bruce, H. (2005), "The Universal Labeler: plan the project and let your information follow", *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology*, American Society for Information Science & Technology, Charlotte, NC, available at: [http://kftf.ischool.washington.edu/docs/UL\\_ASIST05.pdf](http://kftf.ischool.washington.edu/docs/UL_ASIST05.pdf) (accessed 14 April 2009).
- Kaye, J., Vetis, J., Avery, A., Dafoe, A., David, S., Onaga, L., Rosero, I. and Pinch, T. (2006), "To have and to hold: exploring the personal archive", *CHI Proceedings, Personal Information Management, April 22-27 2006, Montreal*, pp. 275-285.
- Keen, A. (2007), *The Cult of the Amateur*, Nicholas Brealey, London.
- Kwasnik, B.H. (1989), "How a personal document's intended use or purpose affects its classification in an office", paper presented at the 12th Annual ACM SIGIR Conference 25-28 June, available at: <http://delivery.acm.org/10.1145/80000/75356/p207-kwasnik.pdf?key1=75356&key2=3717634911&coll=GUIDE&dl=GUIDE&CFID=5417112&CFTOKEN=33985921> (accessed 12 February 2009).
- Lansdale, M. (1988), "The psychology of personal information management", *Applied Ergonomics*, Vol. 19 No. 1, pp. 55-66.

- Lanza, S.R. (2006), "Arrange some, delete some: TaskArrange and AM-Deadlink", *Searcher*, Vol. 14 No. 5, pp. 21-3.
- Lenhart, A., Horrigan, J. and Fallows, D. (2004), *Content Creation Online Pew Internet and American Life Project Report*, available at: [www.pewinternet.org/pdfs/PIP\\_Content\\_Creation\\_Report.pdf](http://www.pewinternet.org/pdfs/PIP_Content_Creation_Report.pdf) (accessed 3 October 2007).
- McKemmish, S. (1996), "Evidence of me . . .", *Archives and Manuscripts*, Vol. 24 No. 1, pp. 28-45, available at: [www.mybestdocs.com/mckemmish-s-evidofme-ch10.htm#\\_edn1](http://www.mybestdocs.com/mckemmish-s-evidofme-ch10.htm#_edn1) (accessed 27 October 2008).
- Mackay, W.E. (1988), "More than just a communication system: diversity in the use of electronic mail", *Proceedings of CSCW'88, Conference on Computer-Supported Cooperative Work*, ACM, New York, NY, pp. 26-28.
- Malone, T.W. (1983), "How do people organize their desks? Implications for the design of office information systems", *ACM Transactions on Office Information Systems*, Vol. 1 No. 1, pp. 99-112.
- Marshall, C. and Jones, W. (2006), "Keeping encountered information", *Communication of the ACM*, Vol. 4 No. 1, pp. 66-7.
- Microsoft (2007), *Expression Studio: Better Designer Tools for Better End-User Experiences*, available at: [www.microsoft.com/uk/expression/launch/nosilverlight/default.html?WT.mc\\_id=074LUK0004](http://www.microsoft.com/uk/expression/launch/nosilverlight/default.html?WT.mc_id=074LUK0004) (accessed 3 October 2007).
- Moore, G.E. (1965), "Cramming more components onto integrated circuits", *Electronics*, Vol. 38 No. 8, pp. 114-7.
- 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)*, Vol. 56 No. 12, pp. 554-8.
- Netcraft (2008), *April 2008 Web Server Survey*, available at: [http://news.netcraft.com/archives/web\\_server\\_survey.html](http://news.netcraft.com/archives/web_server_survey.html) (accessed 29 April 2008).
- Nicholas, D. and Marden, M. (1998), "Parents and their information needs: a case study: parents of children under the age of five", *Journal of Librarianship and Information Science*, Vol. 30 No. 1, pp. 35-48.
- Nonaka, I., Umemoto, K. and Senoo, D. (1996), "From information processing to knowledge creation: a paradigm shift in business management", *Technology in Society*, Vol. 18 No. 2, pp. 203-18.
- Paradigm (2008), *Workbook on Digital Private Papers, 2005-7: The Personal Archives Accessible in Digital Media (paradigm) Project*, available at: [www.paradigm.ac.uk/workbook/index.html](http://www.paradigm.ac.uk/workbook/index.html) (accessed 20 March 2008).
- Roes, H. (1994), "Electronic journals: transforming the information cycle?", paper presented at the Internet Conference, Prague, June, available at: [www.hroes.de/articles/ej\\_praag.htm](http://www.hroes.de/articles/ej_praag.htm) (accessed 20 March 2008).
- Ross, C.S. (1999), "Finding without seeking: the information encounter in the context of reading for pleasure", *Information Processing & Management*, Vol. 35 No. 6, pp. 783-99.
- Savolainen, R. (2005), "Everyday life information seeking", in Fisher, K., Erdelez, S. and McKechnie, L. (Eds), *Theories of Information Behavior*, Information Today, Medford, NJ, pp. 143-8.
- Sellen, A.J. and Harper, R. (2002), *The Myth of the Paperless Office*, MIT Press, Cambridge, MA.

- Siegfried, S., Bates, M.J. and Wilde, D.N. (1993), "A profile of end-user searching behaviour by humanities scholars: the Getty Online Searching Project report 2", *Journal of the American Society for Information Science*, Vol. 44 No. 5, pp. 273-91.
- Summers, A., John, J.L. and The, W.D. (2001), "Hamilton Archive at the British Library", *Ethology, Ecology & Evolution*, Vol. 13 No. 4, pp. 373-84.
- Sweeney, L. (2001), "Information explosion", in Doyle, P., Lane, J.I., Theeuwes, J.J.M and Zayatz, L.V. (Eds), *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, Elsevier Science, Amsterdam, pp. 43-74.
- Taylor, S. (2007), "E-mail: best practice advice", *Library and Information Update*, Vol. 6 No. 5, pp. 34-5.
- Thomas, S. and 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*, Vol. 27 No. 1, pp. 29-56.
- Trace, C.B. (2007), "Information creation and the notion of membership", *Journal of Documentation*, Vol. 63 No. 1, pp. 142-64.
- Trendwatching.com (2007), *Generation C*, available at: [www.trendwatching.com/trends/GENERATION\\_C.htm](http://www.trendwatching.com/trends/GENERATION_C.htm) (accessed 3 October 2007).
- Voorbij, H.J. (1999), "Searching scientific information on the Internet: a Dutch academic user survey", *Journal of the American Society for Information Science*, Vol. 50 No. 7, pp. 598-615.
- Wakefield, J. (2007), "Trying to break the Google habit", *BBC News Online*, available at: <http://news.bbc.co.uk/1/hi/technology/6546501.stm> (accessed 29 April 2008).
- Whittaker, S. and Hirschberg, J. (2001), "The character, value, and management of personal paper archives", *ACM Transactions on Computer-Human Interaction*, Vol. 8 No. 2, pp. 150-70.
- Whittaker, S. and Sidner, C. (1996), "E-mail overload: exploring personal information management of e-mail", paper presented at the Conference on Human Factors in Computing Systems (CHI 1996), Vancouver, BC.
- Whittaker, S., Bellotti, V. and Gwizdka, J. (2006), "E-mail in personal information management", *Communications of The ACM*, Vol. 49 No. 1, pp. 69-75.
- Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E. and Hainsworth, J. (2004), "ContactMap: organizing communication in a social desktop", *ACM Transactions on Computer Human Interaction*, Vol. 11 No. 4, pp. 445-71.
- Williams, P., Dean, K., Rowlands, I. and John, J.L. (2008), "Digital lives: report of interviews with the creators of personal digital collections", *Ariadne*, Vol. 55, available at: [www.ariadne.ac.uk/issue55/williams-et-al/](http://www.ariadne.ac.uk/issue55/williams-et-al/) (accessed 25 April 2008).
- Williamson, K. (1998), "Discovered by chance: the role of incidental information acquisition in an ecological model of information use", *Library and Information Science Research*, Vol. 20 No. 1, pp. 23-40.
- Wilson, T. (1997), "Information behaviour: an inter-disciplinary approach", *Information Processing and Management*, Vol. 33 No. 4, pp. 551-72.
- Xlibris (2007), *Welcome to Xlibris*, available at: [www2.xlibris.com/](http://www2.xlibris.com/) (accessed 3 October 2007).
- Zakon, R.H. (2006), *Hobbes Internet Timeline*, available at: [www.zakon.org/robert/internet/timeline/#2000s](http://www.zakon.org/robert/internet/timeline/#2000s) (accessed 29 April 2008).



### Further reading

- Barreau, D.K. (1995), "Context as a factor in personal information management systems", *Journal of the American Society for Information Science*, Vol. 46 No. 5, pp. 327-39.
- Gemmell, J., Bell, G. and Lueder, R. (2006b), "MyLifeBits: a personal database for everything (expanded version)", available at: <ftp://ftp.research.microsoft.com/pub/tr/TR-2006-23.doc> (accessed 2 October 2007).
- Gemmell, J., Williams, L., Wood, K., Lueder, R. and Bell, R. (2004), "Passive capture and ensuing issues for a personal lifetime store", *Proceedings of the CARPE (Continuous Archival and Retrieval of Personal Experiences) 15 October, 2004, New York, NY*, pp. 48-55.
- Tauscher, L.M. and Greenberg, S. (1997), "How people revisit web pages: empirical findings and implications for the design of history systems", *International Journal of Human-Computer Studies*, Vol. 47 No. 1, pp. 97-137.

### Corresponding author

Peter Williams can be contacted at: [peter.williams@ucl.ac.uk](mailto:peter.williams@ucl.ac.uk)

---

To purchase reprints of this article please e-mail: [reprints@emeraldinsight.com](mailto:reprints@emeraldinsight.com)  
Or visit our web site for further details: [www.emeraldinsight.com/reprints](http://www.emeraldinsight.com/reprints)