A PIM Perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records[1]

JORDAN BASS

RÉSUMÉ Cet article se penche sur l'environnement numérique personnel – trop souvent simplifié à l'extrême – afin de faire ressortir les nombreuses nuances qui existent dans le contexte de création des documents et de leur utilisation par des individus à l'ère du numérique. Il explore spécifiquement les stratégies de gestion de documents numériques personnels, les décisions d'évaluation et les désignations de valeur, ainsi que les pratiques de conservation numérique, du point de vue des études en gestionnaires d'informations personnelles (« *Personal Information Management* »), à partir d'une recension des écrits publiés hors des revues et monographies archivistiques traditionnelles. En examinant comment les gens créent, rassemblent, classent, conservent et (ré-)accèdent à l'information numérique, la recherche en gestionnaires d'informations personnelles sert de complément à nos connaissances actuelles sur les documents numériques personnels et révèle de nouvelles informations au sujet de ce matériel qui n'ont pas encore paru dans les écrits en archivistique. Ce texte suggère qu'une vraie compréhension des processus de médiation des documents d'archives, qui s'effectue dans l'environnement des archives numériques personnelles bien avant le versement à un centre d'archives, fait partie de la découverte et de l'exploitation de l'information nécessaire au sujet de sa provenance.

ABSTRACT This paper investigates the often oversimplified personal digital archiving environment to expose the many nuances in the context of the creation and use of records by individuals in the digital era. It specifically examines personal digital recordkeeping strategies, appraisal decisions, and identifications of value, as well as digital preservation practices from the perspective of Personal Information Management (PIM) studies through a review of pertinent literature published outside traditional archival journals and monographs. Through explorations of how people create, collect, organize, maintain, and (re)access digital information, PIM research complements our existing knowledge about personal digital records and reveals addi-

1 This article was awarded the first Gordon Dodds Prize, which recognizes superior research and writing on an archival topic by a student enrolled in a master's level archival studies program at a Canadian university. Instituted in 2011, the award honours Gordon Dodds (1941–2010), who was the first president of the ACA and *Archivaria*'s longest-serving general editor.

ARCHIVARIA 75 (Spring 2013): 49–76

tional information about these materials heretofore undisclosed by archival scholarship. This paper suggests that a genuine understanding of the processes of records mediation occurring in the precustodial environment of personal digital archives is integral to the discovery and exploitation of their requisite provenancial information.

> *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.*[2]

> *Inscribers and pre-archival custodians of records document some things and not others (that is an appraisal decision of sorts) and they choose to destroy certain records, without knowledge of archives, or offer only certain records to archives, holding back others for other times.*[3]

Introduction

The phrase "archives in the wild," as defined by British archival curator Jeremy Leighton John, refers to "the personal digital archives that exist outside an official long-term repository," including the personal archives of academics, literary figures, and politicians, as well as the digital collections of ordinary, everyday people.[4] These archives in the wild are created and preserved by individuals with diverse recordkeeping behaviours, and comprise dynamic documentary forms dispersed throughout multiple online and offline digital landscapes. Within this wilderness, digital archives are usually created, accumulated, and maintained instinctively and expediently as opposed to systematically and routinely, which is the case in institutional environments. Moreover, this digital wilderness is the hinterland for memory institutions (archives, libraries, and museums), whose mandate is the acquisition, preservation, and provision of access to collections of personal records. Yet micro-level analyses of the context of creation and use of records in private digital environments remains a much-neglected area of study and is in many respects still frontier research in professional archival scholarship. Moreover, there is a consensus among archivists that "only with an accurate and comprehensive perception of how electronic records are organized can archivists have a good understanding of what they are going to deal with at the time that records are transferred to archival systems."[5]

2 Jeremy Leighton John, Ian Rowlands, Peter Williams, and Katrina Dean, "Digital Lives: Personal Digital Archives for the 21st Century – An Initial Synthesis, Beta Version 0.2" (March 2010), 9, http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf (accessed 14 June 2010). Jeremy Leighton John is the principal investigator of the Digital Lives Research Project and curator of eMANUSCRIPTS at the British Library.

3 Tom Nesmith, "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (Fall 2005): 263–64.

4 Leighton John et al., "Digital Lives: Personal Digital Archives for the 21st Century," 5.

5 Jane Zhang, "The Principle of Original Order & the Organization and Representation of

This paper examines personal digital recordkeeping strategies, appraisal decisions, and designations of value, as well as digital preservation practices from the perspective of Personal Information Management (PIM) studies, through a review of pertinent literature published outside traditional archival journals and monographs. It seeks to outline the context of creation and use of personal digital records before they are acquired by archival institutions in order to discover why, how, and where individuals create and preserve documentary forms in the digital era. This paper argues that a genuine understanding of the processes of records mediation[6] occurring in the precustodial environment of personal digital archives is integral to the discovery and exploitation of their requisite provenancial information.

Personal Information Management (PIM)

> We all keep information in our work and domestic lives. It may be books, notes, diaries, personal records, files or whatever. This is personal information not necessarily in the sense that it is private, but that we have it for our own use. We own it, and would feel deprived if it were taken away.[7]

Personal information management, or PIM, is described as "both the practice and the study of the activities people perform to acquire, organize, maintain, retrieve, use and control the distribution of information items such as documents (paper-based and digital), Web pages, and email messages for everyday use to complete tasks (work-related and not) and to fulfill a person's various roles (as parent, employee, friend, member of community, etc.)."[8] As an academic area of study, PIM "draws upon the best work from a range of disciplines including cognitive psychology, human-computer interaction, database management, information retrieval, and information science."[9] Personal information management studies are also used to discover innovative ways to

Digital Archives" (PhD diss., Simmons College Graduate School of Library and Information Science, Boston, 2010), 189.

6 Discussing the complexity and nuance of archival provenance, archival educator and theorist Tom Nesmith suggests that "a record is an evolving mediation of understanding about some phenomena – a mediation created by social and technical processes of inscription, transmission, and contextualization." Tom Nesmith, "Still Fuzzy, But More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory," *Archivaria* 47 (Spring 1999): 145. In this article, the term "mediation" refers to any action or series of actions taken on the meaning, form, or function of personal digital information.

7 M.W. Lansdale, "The Psychology of Personal Information Management," *Applied Ergonomics* 19, no. 1 (March 1988): 55. This article is commonly cited as the first expression of the term "personal information management" as a practice and area of study.

8 William Jones and Jaime Teevan, "Introduction," in *Personal Information Management*, ed. William Jones and Jaime Teevan (Seattle, 2007), 3.

9 William Jones, "Personal Information Management," in *Annual Review of Information Science and Technology*, ed. Blaise Cronin (Medford, NJ, 2007), 454.

assist individuals in managing excessive volumes of digital information more efficiently through the design of successful software and hardware to meet specific objectives in what may be referred to as PIM tools or applications. The development of PIM applications is in most cases commercially driven to assist individuals in the short-term management of a particular technological format such as email, text, image and music files, or Web bookmarks. For example, ubiquitous PIM applications include email client software (Microsoft Office Outlook and Mozilla Thunderbird), file manager applications (Windows Explorer and Mac Finder), as well as organizational calendars (Google Calendar) and music file managers (iTunes).[10]

While they continue to evolve over time, all PIM technologies invariably involve three primary functions: to create, arrange, and (re)access information in personal digital collections. However, archivally oriented PIM adds a fourth function, which may be referred to as the long-term preservation of personal digital information. Archival PIM considers the factors involved in maintaining personal digital information throughout its entire 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."[11] In these endeavours to design better information management technology, PIM research is obligated to examine and evaluate user behaviours and strategies involved in individual recordkeeping, appraisal, and preservation activities, and as such provides unique insights and an alternative perspective on personal digital archives. To fully articulate the significance of PIM research to the archival profession, however, requires that it first be situated within current archival theory and methodology.

The Precustodial Environment

Archivists principally adhere to one of two conceptual models, or derivatives thereof, when discussing the creation and management of documentary forms: the records life cycle and the records continuum. The life cycle concept portrays records traversing two phases with eight particular stages: a records management phase consisting of stages relating to creation, classification, maintenance and use, and disposition, succeeded by an archival phase with stages of selection and acquisition, description, preservation, and

10 Emerging and more robust PIM applications include note-taking software suites (Microsoft OneNote and Evernote), web-based file managing services (DropBox and iCloud), as well as reference/personal bibliographic management software (Zotero and Mendeley). PIM applications may be proprietary or open source, used online or offline, and synchronized between desktop (Mac OSX) and mobile (iOS) operating systems.

11 Leighton John et al., "Digital Lives: Personal Digital Archives for the 21st Century," x.

use.[12] Records continuum thinking, on the other hand, posits the passing of records through four integrated time-space dimensions involving creation, capture, organization, and pluralization. While there are similarities and differences between the records life cycle and the records continuum, both models concede a temporal and spatial period of documentary activity that may be referred to as the precustodial environment preceding the archival mediation of records.[13] Once the archives takes custody of the records, archivists ascertain what they can about this precustodial environment, such as the circumstances surrounding the creation and use of personal records and the relationships between them, in their endeavours to determine provenance in retrospect. In personal archives, these extrapolations are in essence based on the documentary fragments the archivist possesses at the point of acquisition, which, depending on the creator or donor of the records, may be quite limited in quantity and quality. PIM studies, on the other hand, document extant personal recordkeeping systems in situ.

The traditional records life cycle model hinges on institutional transactions and is therefore inclined to express the trajectory of documentary forms through fixed-linear stages of records management. Moreover, this model is often applied, either by implication or by default, to personal archives in both digital and non-digital domains. As a result, precustodial creation, recordkeeping, appraisal, and preservation of personal documentary forms are compartmentalized as a single epoch (creation) in a finite series of temporal and spatial businesslike progressions (maintenance, scheduling, and disposition) leading to the record's eventual mediation by archivists. These progressions through time and space inform the basic provenance of records, which is in turn either confirmed or denied by archival mediation. In personal private settings, it may be more appropriate to reconceptualize the records life cycle as occurrences of "social and technical processes of inscription, transmission, contextualization, and interpretation" streaming through both precustodial and archival sites of mediation and culminating in the construction of prov-

12 Jay Atherton, "From Life Cycle to Continuum: Some Thoughts on the Records Management–Archives Relationship," *Archivaria* 21 (Winter 1985–86): 43–51. Atherton amended the traditional life cycle model from eight stages to a single-phase, four-stage model consisting of creation, classification, scheduling, and maintenance and use. In doing so, Atherton bridged the gap between records management and archival management.

13 Barbara Reed, "Reading the Records Continuum: Interpretations and Explorations," www.records.com.au/pdf/Reading_the_Records_Continuum.pdf (accessed 11 June 2011); originally published in *Archives & Manuscripts* 33, no. 1 (May 2005): 18–43. In the Australian records continuum, this documentary activity occurs within the first (create) and second (capture) dimensions, or, as Reed explains, "the first 'create' dimension of the records continuum represents the locus of all action … including representations of actions in documents … where characteristics from the second dimension, records, now attest to evidence of action."

enancial attributes.[14] The attestation of this rich provenancial information, however, has become overtly speculative with the profoundly ephemeral and fragile nature of personal digital records. Thus, the (re)encounter of archival provenance in the digital era makes necessary the timely documentation of records creation, use, and transmission within the precustodial environment, a task to which the PIM discipline is aptly suited because of its temporo-spatial orientation within the mediation stream of personal records.

Personal Records Creation and Recordkeeping

Throughout the course of a lifetime, people naturally document their private and public activities during encounters with everyday phenomena. At times, this documentation is an imperative for financial reasons, required to demonstrate accountability or essential in the performance of ongoing occupational duties. In addition, people compose electronic missives and narrative, take digital photos, and record audio and video without preparation or for tentative purposes. Day after day people continually and consciously document their existence as part of the innate need to communicate with others and to engage in life's pursuits. These inaugural acts of documentation may be viewed as the first horizons of personal recordkeeping, whereby digital information is created but has yet to undergo processes of organization and management.

When personal documentation takes digital form, it is at once designated as a specific file type. For instance, when an individual is creating a Microsoft Word document, data are allocated to random access memory (RAM) until an auto-save or manual save command is executed, at which time the data are encoded to a unique formatting algorithm, assigned a suffix with a three- or four-character extension (.doc or .docx), and stored on the hard disk drive (HDD).[15] Operating systems (Windows based) use these file extensions to locate and execute the associated application software (Microsoft Word) required to render the data in a specific format to make it human readable.[16] While it is extremely difficult to calculate an exact number of file formats existing in personal computing environments, it is, however, possible to

14 Nesmith, "Still Fuzzy, But More Accurate," 145.

15 Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA, 2008), 50–53. In his detailed grammatology of the hard drive, Kirschenbaum explains that "word processors and other productivity software routinely include an auto-save function that writes a snapshot of an open file to the disk at set intervals."

16 Windows operating system XP used the .doc extension for MS Word documents. This has since changed with the Vista OS, with the suffix .docx now used. As there are thousands of proprietary and open software applications available to the public, technical registries such as FileInfo.com and PRONOM aim to provide definitive information on the thousands of file formats and extension suffixes associated with those software applications.

categorize them broadly in two classes: files created offline, such as those analogous to analog forms, and online files used exclusively in the construction of web-based content. Both classes of files rely on associated software applications, while online files introduce the additional requirement of host servers to publish and manipulate content on the Web. In short, the file is the rudimentary element of personal digital records.

At one point or another, many of these file classes are created and/or manipulated in a common domain – the personal computer. For example, files created in a digital camera are typically edited by software such as Adobe Photoshop installed on the personal computer, where the files subsequently remain until further action is taken. Similarly, HTML, Cascading Style Sheets, and JavaScript files are created using text editors installed on the personal computer before FTP and hosting servers render and make available the content on the Web. Likewise, content uploaded to file-sharing platforms such as YouTube and Flickr or to social-networking services like Facebook emanates primarily from files created or edited on the personal computer.[17] Regardless of its operating system (Mac OSX, Linux, Microsoft Windows, or Unix), the personal computer has for some time acted as a central hub for the creation of digital files as well as their continued use and management in what may be referred to as series of recordkeeping actions:

> [F]iling and finding are such basic aspects of working with computers that while we scarcely notice their existence – hence the lack of research – every computer user spends time and effort in filing and finding every time the computer is used. As designers we should be concerned with optimizing finding and filing.[18]

All personal computers provide the ability to create folders via file manager applications (Windows Explorer or Mac Finder) for organizing text, audio, video, spreadsheets, and dynamic presentations, among other file classes. These folders are in turn nested hierarchically to create either deep or shallow structures containing hundreds, if not thousands, of aggregated digital items. As personal computers are often the centre of creation and organization, file management is a core component of personal recordkeeping. A second important hub of personal recordkeeping is email. Considered the digital

17 As with all content published on the Web, social-networking and file-sharing services involve server-side augmentation, which client-side entities have little or no control over in terms of page layout or functionality. Smartphone technology (iPhone and BlackBerry) also supports mobile uploading of content to file sharing and social-networking sites, but mobile data, such as captured images, are still kept in program or application directories of the personal computer when it is synchronized with the smartphone or exported using extraction software.

18 Deborah Barreau and Bonnie A. Nardi, "Finding and Reminding: File Organization from the Desktop," *ACM SIGCHI Bulletin* 27, no. 3 (July 1995): 39.

counterpart of handwritten and typewritten correspondence, email is one of the most widely adopted applications of computer-mediated communication thanks to its almost universal availability and the negligible effort required to learn to use it.[19] Like written correspondence, email is a recorded manifestation of asynchronous communication between two or more people and is in many respects the major means of non-face-to-face communication in the early twenty-first century.[20] But far from being a simple accumulation of messages, email also serves as a primary information conduit through which individuals manage daily tasks and exchange a number of different file classes through attachments or embedding.[21] As a PIM application, email has grown to encompass the functions of personal recordkeeping to such a degree that many people "tend to live in email" as demonstrated by the sheer amount of time spent using it.[22] The next section of this paper discusses personal recordkeeping within the context of file and email management.

In her 2009 study, Sarah Henderson profiles the recordkeeping behaviours of 125 knowledge workers through interviews, surveys, and on-site demonstrations of a participant's document management practices.[23] From her data, Henderson determined three distinct strategies. In the first strategy, individuals construct moderate folder structures through periodic cleanups or when accumulations of documents warrant them. As the folder structure is of medi-

19 Simon Scerri, Siegfried Handschuh, and Stefan Decker, "Semantic Email as a Communication Medium for the Social Semantic Desktop," in *ESWC '08 Proceedings of the 5th European Semantic Web Conference on the Semantic Web: Research and Applications,* 124–25, http://portal.acm.org/citation.cfm?id=1789410 (accessed 20 May 2011). Unlike paper-based correspondence, email adheres to a number of protocols (SMTP, POP, and IMAP) in the request-response function between an email client and server. It may be argued, however, that written correspondence adheres to protocols or etiquette in, for example, letters of condolence or invitation. But this type of protocol is more culturally imposed and socially expected than it is required for the actual transmission of information.

20 Nicholas Ducheneaut and Victoria Bellotti, "Email as a Habitat: An Exploration of Embedded Personal Information Management," *ACM Interactions* 8, no. 5 (September/October 2001): 30–38.

21 Ibid., 70. The ubiquity of email began in the mid-1990s with the debut of personal email accounts such as Hotmail and Yahoo!, which offered free but limited disk storage space on central servers. With the advent of Google Mail (Gmail) in 2004 and its one-gigabyte (one-thousand megabytes) capacity of free storage per user, email accounts have quickly come to act as personal repositories of information more so than simple inboxes since individuals are rarely forced to delete messages as a result of limited storage space. By 2009, the per-user free storage space had grown to 7.3 gigabytes.

22 Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka, "Email in Personal Information Management," *Communications of the ACM: Personal Information Management* 49, no. 1 (January 2006): 68.

23 Sarah Henderson, "Personal Document Management Strategies," in *CHINZ '09 Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human–Computer Interaction* (2009), http://sarahhenderson.info/tags/personal-document-management/ (accessed 11 June 2011), 69–76.

um depth, browsing as opposed to executing a text-search query is employed to locate documents ("Filing"). A second identified strategy involves relatively disorganized clusters of documents for which low numbers of folders are created and are unsystematically used. In this second strategy, the computer desktop is the primary area of organization and can therefore be browsed with relative ease, while information that has been filed away haphazardly is retrieved by text-search query ("Piling"). The final strategy demonstrates organized folders with low numbers of unclassified documents, whereby individuals often pre-empt the influx of documents through the creation of folder categories in advance. In this strategy, individuals browse to re-find documents but rely on context, such as parent folders, to locate information within their deep folder structures ("Structuring").

In their study of "finding" and "reminding," Deborah Barreau and Bonnie Nardi report on the collated findings of two separate investigations of the methods employed by managers, graphic artists, programmers, administrative assistants, and librarians to organize and find files on their personal computers. Barreau and Nardi highlight two basic recordkeeping strategies: location-based finding and logical (text-search query) finding.[24] They conclude that the deployment of these two strategies correlates with the type of electronic information the user is working with, which they classify as "ephemeral," "working," or "archived."[25] Study participants demonstrated that ephemeral and working items are often retrieved by location-based finding because this type of information is frequently used or serves a reminding function and therefore receives prime real estate on the computer desktop. Archived items of less immediate relevance or utility, however, are rarely organized in a systematic way and rely on text-search queries for retrieval. Barreau and Nardi conclude that while individuals keep archived information for extended periods of time, study participants indicated that selecting and establishing logical filing schemes of keywords and carefully built structures for this information often failed.[26] Users, the authors suggest, "prefer filing by location because it aids in helping them find what they need as well as serving a crucial reminding function."[27] Accordingly, location-based storage, on the desktop for instance, "assumes a small information collection (basically what the user can remember) and does not scale to large archived collections" that comprise, for example, numerous folders and multiple files.[28] In short, the way individuals

24 Barreau and Nardi, "Finding and Reminding," 39–43.

25 Ibid., 40–43. These three types of information are further defined as ephemeral information (short shelf life), working information (shelf life of weeks or months), and archived information (shelf life of months or years).

26 Ibid., 42.

27 Ibid.

28 Scott Fertig, Eric Freeman, and David Gelernter, "Finding and Reminding Reconsidered,"

use their information "is a primary determinant of how it will be organized, stored, and retrieved in the personal workspace," and because archived collections of information are "often needed in a context that is different from the one in which it was created" or, in other words, "secondary" to ephemeral and working information, they are spatially separated as a unique type of information.[29]

Profiling the email use of sixty professional office workers, Wendy Mackay found that individuals typically view email as a tool for task management or information management. Those viewing email as a mechanism to support a variety of time and task management activities often do not read all incoming messages, limit the number of times they read email per day, and reduce message volume by unsubscribing from mail lists.[30] These "prioritizers" run contrary to those users who attempt to read all incoming mail, save a large percentage of messages, and maintain many email folders to which messages are transferred on an ad hoc basis. These "archivers," she notes, seek to keep rather than delete, view email as a means for supporting information management, and are subsequently overwhelmed by messages in their inboxes. Mackay's study reveals how the email organization patterns of individuals are influenced by how they think about the functions of email; those who view email as an information store typically retain all of their messages and do not view the deletion of messages "as particularly useful."[31]

In their seminal exploration of email management, Steve Whittaker and Candace Sidner state:

> Email is one of the most successful computer applications yet devised. Our empirical data show, however, that although email was originally designed as a *communications* application, it is now being used for additional functions that it was not designed for, such as *task management* and *personal archiving*. We call this *email overload*.[32] [Italics in the original].

ACM SIGCHI Bulletin 28, no. 1 (January 1996): 67.

29 Barbara H. Kwasnik, "How a Personal Document's Intended Use or Purpose Affects Its Classification in an Office," in *Proceedings of the ACM-SIGIR 12th Annual International Conference on Research and Development in Information Retrieval* (June 1989), 210, http://portal.acm.org.proxy1.lib.umanitoba.ca/citation.cfm?doid=75334.75356 (accessed 11 June 2011). See also Fertig, Freeman, and Gelernter, "Finding and Reminding Reconsidered," 67–68.

30 Wendy E. Mackay, "More than Just a Communication System: Diversity in the Use of Electronic Mail," in *Proceedings of the 1998 ACM Conference on Computer-Supported Cooperative Work* (New York, 1998), 344–53.

31 Ibid., 350.

32 Steve Whittaker and Candace Sidner, "Email Overload: Exploring Personal Information Management of Email," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (New York, 1996), 276–83, http://portal.acm.org/citation.cfm?id=238530&dl=ACM&coll=DL&CFID=27124289&CFTOKEN=32910868 (accessed 20 May 2011).

Examining the inboxes of twenty office workers, Whittaker and Sidner identified four types of email messages: those requiring the user to execute some action (to dos); messages with considerable informational content requiring thorough examination (to reads); informational messages of undetermined significance (intermediate status); and threads of asynchronous communication (ongoing correspondence).[33] The authors note that in rationalizing an overwhelming amount of these types of information, individuals tend to perform one of three recordkeeping strategies. Those who make use of email folders are categorized as users who frequently delete and archive email items ("frequent filers"), whereas those users who periodically delete and archive (every one to three months) are categorized as "spring cleaners."[34] Users who neither delete nor archive email items are categorized as "no filers." Following the deployment of these strategies, users engage in the process of maintaining a filing system, which Whittaker and Sidner state "is a cognitively difficult task" involving the overhead costs of consistently creating folders, adhering to unfailing naming conventions, and remembering the definitions of folders and their contents for every transfer of an item out of the inbox ("multiple folder definitions").[35] Also, creating new folders may not be useful if pre-existing ones are synonymous ("duplication"). Furthermore, created folders with only one or two messages do not significantly reduce the complexity of email management, while folders with too many messages become just as unmanageable ("failed folders") as the inbox itself.[36]

By collecting longitudinal and cross-application data on personal recordkeeping strategies, Richard Boardman and Martina Sasse provided insight into how personal recordkeeping behaviour evolves over time.[37] After profiling thirty-one participants, Boardman and Sasse argued that in the creation of documents, individuals will generally file either occasionally, extensively, or upon creation or accumulation of digital information. The authors note that those participants exhibiting high organizational effort ("pro-organizing")

33 Ibid., 278.

34 Ibid., 280–83. Archiving is defined by Whittaker and Sidner as the process of transferring an email item from the inbox to a separate folder.

35 Ibid., 279. Ten years later, two PIM researchers conducted a similar study comparing it to the research of Whittaker and Sidner and discovered a tenfold increase in email archive size as well as a significant increase in the amount of email folders. See Danyel Fisher, A.J. Brush, Eric Gleave, and Marc A. Smith, "Revisiting Whittaker & Sidner's 'Email Overload' Ten Years Later," in *Proceedings of the 15th International Conference on Intelligent User Interfaces* (November 2006), 309–12, research.microsoft.com/pubs/69394/p309-fisher.pdf (accessed 20 May 2011).

36 Whittaker and Sidner, "Email Overload," 280.

37 Richard Boardman and Martina Angela Sasse, "Stuff Goes into the Computer and Doesn't Come Out: A Cross-tool Study of Personal Information Management," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI*, 6, no. 1 (New York, 2004), 583–90.

tend to do so across their collections of email, files, and web bookmarks, with similar overlap found across all three collections among those individuals exhibiting low organizational tendencies ("organizing neutral").[38] Revisiting these classifications in the second phase of their study, Boardman and Sasse noted that changes in recordkeeping tendencies are relatively subtle and tend to be adjustments to existing pro-organizing strategies rather than major changes such as an individual changing their email strategy from "no filer" to "spring cleaner."[39] The authors conclude that an individual's organizational strategy is also influenced by his/her method of retrieval. For example, a person relying primarily on folder-based browsing in retrieval will invest time in filing for the "cost of filing is offset by predicted benefits at retrieval time."[40]

Additional PIM studies have identified similar recordkeeping strategies, including "cleaners and keepers," "folder-less cleaners and folder-less spring-cleaners," "neat and messy," as well as "sporadic and end-of-session filers," among others.[41] While these studies differ somewhat in how they classify user behaviours, there are three prominent attributes of personal recordkeeping to be considered. First, personal digital items are either active or dormant, with the former logically situated for regular re-encounters and the latter relegated to more obscure locations. As active digital items will inevitably become inactive, agglomerations of dormant files fall victim to a "poverty of attention," competing against a growing amount of active information for cognitive processing.[42] Second, the process of re-finding personal digital items invari-

38 Ibid., 584–88. Along with email and files, web bookmarking is the third major area of PIM research. Web browsers support ways of returning to previously viewed web content by adding URLs to "Favorites" (Internet Explorer) and "Bookmarks" (Firefox). Alternatively, a person might email a URL link to himself or herself or paste it in a document.

39 Ibid., 588–90. In *Phase 2* of the study, the authors tracked the evolution of recordkeeping strategies across all three digital information collections over an average period of 286 days.

40 Ibid., 589.

41 See Jacek Gwizdka, "Email Task Management Styles: The Cleaners and the Keepers," in *SIGCHI Conference on Human Factors in Computing Systems* (Vienna, 2004), 1235–38; Olle Balter, "Strategies for Organizing Email," in *People and Computers XII: Proceedings of HCI '97*, ed. H. Thimbley, B. O'Conaill, and P.J. Thomas (London, 1997), 21–38; Thomas W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems* 1, no. 1 (January 1983): 99–112; David Abrams, Ron Baecker, and Mark Chignell, "Information Archiving with Bookmarks: Personal Web Space Construction and Organization," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (California, 1998), 41–48.

42 In 1971, economist Herbert Simon stated, "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention." Herbert Simon, "Designing Organizations for an Information-Rich World," in *Computers, Communications and the Public Interest*, ed. Martin Greenberger (Baltimore, 1971), 40–41, quoted in Thomas H. Davenport and John C. Beck, "The Attention Economy,"

ably relies on the execution of queries based on keyword attributes recovered from human memory (search-based system) or contextual and spatial cues encountered through browsing (location-based system). Both approaches, however, are encumbered by an individual's capacity to recall precise detailed information from the confines of human memory, or an individual's ability to maintain proficient and persistent organizational schemes within hierarchical folder structures. Third, in all recordkeeping strategies reviewed in this section, there is little mention of individuals consciously deleting digital items, which implies that dormant items are, by default, segregated from new items and kept for further cognitive processing rather than destroyed by their creators. In digital environments, retention is the norm and destruction the exception for the simple reason that there is often no motive to destroy files when limits on virtual space and financial cost cease to be determining factors.

Despite the number of items ostensibly allocated in digital silos of inactivity, a percentage of these items is often released to ensuing phases of recordkeeping upon the fulfillment of distinctively personal disposition criteria. The next section of this paper examines why people choose to preserve re-encountered digital items and how individuals extend the production of digital documentation to the creation of digital records.

Personal Appraisal and Notions of Value

Professional archivists engage in two types of appraisal when acquiring the records of private individuals: macro-level appraisal to determine the historical and cultural significance of a personal collection in relation to an acquisition or collection development policy; and micro-level appraisal to determine the value of documentary forms within that personal collection in order to separate the wheat from the chaff. As active participants in the construction of documentary heritage, archivists strive to inform their appraisal decisions as much as possible on "value ascribed by those contemporary to the material."[43] Yet in digital environments, the opportunities to locate and exploit conceptions of value within their temporal context dissipate at a far quicker rate than in analog domains and make the rediscovery of value more difficult. For instance, while archivists were previously able to appraise physical media (journals and diaries, literary drafts, poems, and correspondence, or photographs and audiovisual recordings) fifty years or more after their

Ubiquity (May 2001), http://ubiquity.acm.org/article.cfm?id=376626 (accessed 10 October 2011).

43 Hans Booms, "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," *Archivaria* 24 (Summer 1987): 104.

creation, this interval of time has significantly contracted with information technology becoming "essentially obsolete every 18 months," and digital storage media expected to become obsolete within no more than five years.[44] Concurrently, the passage of time between the creation of personal records and their appraisal by professional archivists reflects evolving mediations of value facilitated by implicit and explicit private appraisal decisions. As such, these values and private appraisals remain indispensable contextual elements of personal digital records and are critical to the understanding and explanation of their provenance.

In her discussion of the long-term fate of personal digital archives, Microsoft researcher Catherine C. Marshall and her colleagues state:

> Why do we need to discuss the value of digital materials separately instead of bringing in best practices from physical information management? The reason is straightforward: digital belongings accumulate at a far more precipitous and unmanageable rate than physical belongings do.[45]

Marshall suggests that as individuals rapidly create and accumulate digital materials there are occasions when conscious decisions about what to save and what to delete become unavoidable in the effort to establish control over their technological environment. These decisions, Marshall affirms, may be linked with private assessments and designations of value that are identified through examining the source of the digital item, the actions taken by or upon it, and by its disposition.[46]

For instance, a calculation of how often a digital item is used or replicated, how much time and effort went into its creation, with whom the digital item is shared, and the ability to reconstruct its source over time produces five general notions of value: demonstrated worth, creative effort, labour, emotional impact, and stability.[47] Marshall argues that in the personal digital archive, value is not wholly ascribed to items at inception, but rather accrues with use, in custody, and transmission over time. Here, Marshall leverages

44 Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Information," *Sixty-third IFLA Council and General Conference* (August 1997), 3, http://archive.ifla.org/IV/ifla63/63kuny1.pdf (accessed 15 May 2011); and Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272, no. 1 (January 1995): 42-47.

45 Catherine C. Marshall, Sara Bly, and Francoise Brun-Cottan, "The Long Term Fate of Our Digital Belongings: Towards a Service Model for Personal Archives," in *Proceedings of Archiving 2006*, http://research.microsoft.com/apps/pubs/default.aspx?id=75527 (accessed 15 May 2011).

46 Catherine C. Marshall, "Rethinking Personal Digital Archiving Part 2: Implications for Services, Applications, and Institutions," *D-Lib Magazine* (March/April 2008), http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html (accessed 15 May 2011).

47 Marshall, Bly, and Brun-Cottan, "The Long Term Fate of Our Digital Belongings."

the tacit appraisal decisions of individuals and expresses them heuristically to advocate an organic archiving system that expedites the cognitive process of distinguishing "between items that are valuable and items that have simply accumulated" in the personal digital archive.[48]

In a series of surveys and interviews with fifty office workers, Steve Whittaker and Julia Hirschberg of AT&T Labs Research examine how individuals evaluate the significance of paper-based information and the motivations behind the archiving of that information.[49] Whittaker and Hirschberg concluded that individuals measure their decisions to archive against five broad criteria: reference value, legal and administrative value, immediate availability, reminder of encountered information, and distrust of external information storage.[50] However, the authors posit that individuals archive information for reasons beyond routine business functionality and factor emotional and sentimental reasons into the decisions they make regarding the retention and disposition of recorded information. For example, the authors note that documents such as reviews of published papers, successful research prototypes, and reference documents containing personal annotations have "little relevance for likely activities, but they [people] still cannot part with [them], because it is part of their intellectual history."[51] One participant acknowledged that although they had no identifiable need for their archived papers and could not articulate why they kept them, "sentiment … or something" prevented them from throwing the papers out.[52] Emotional or sentimental reasons for archiving, while not easily rationalized in business settings, are nevertheless contemplated by individuals in the appraisal of their recorded information.

The broad abstraction of "sentimental value" is deconstructed in a PIM study seeking to design technologies to support the preservation of analog and digital objects within familial environments. In their excavation of the home archive, Microsoft researchers David Kirk and Abigail Sellen explain that sentimental value is "tied to the notion of constructing or bolstering a sense of identity, through knowing who one is by keeping hold of memories and reflections of the past" ("Constructing the persona").[53] Study participants also informed the authors of the value of sharing a collective past, whereby

48 Marshall, "Rethinking Personal Digital Archiving Part 2."

49 Steve Whittaker and Julia Hirschberg, "The Character, Value and Management of Personal Archives," *ACM Transactions on Computer Human Interaction* 8 (2001), 150–70. The study coincided with an office move, which the authors believed would motivate their participants to make clear decisions about retention and disposition of materials.

50 Ibid., 155–56.

51 Ibid., 156–57.

52 Ibid.

53 David Kirk and Abigail Sellen, "On Human Remains: Excavating the Home Archive," *Microsoft Technical Report* (June 2008), http://research.microsoft.com/apps/pubs/default.aspx?id=70595 (accessed 15 May 2011), 5.

certain objects such as photographs supported family "connectedness" or when objects are designated to be passed down through successive generations ("Connecting with a shared past"). Similarly, participants also attribute importance to objects intended for broader historical considerations by persons unknown in what may be referred to as personal or familial legacy value ("To preserve a legacy").[54] Kirk and Sellen conclude that the impulse to honour the past permeates many of reasons why people archive; for example, photographs appraised for their utility in the construction of persona or personal memory also express reverence for others in the photograph and may serve "to elevate others into family consciousness (In Honorium)."[55]

A comparable study observes that individuals archive for multiple reasons and in multiple ways, yet communicate five common motivations: to store and retrieve information for later use ("Finding it later"); as a testament to personal and professional achievement ("Building a legacy"); to facilitate access by others ("Sharing resources"); out of anxiety about losing important information ("Fear of loss"); and as a reflection or expression of themselves ("Identity construction").[56] The study also identifies a strong connection between these motivations and how an individual structured his or her archives. For instance, a study participant expressing anxiety over loss of information carefully archived his entire filing system to support easy incremental backups to avoid catastrophic loss of information in a hard-drive crash.[57] In comparison, study participants who rigorously documented life or career milestones or who kept information expressing personal achievement, did so not in the interest of future information retrieval or out of fear of loss, but rather in the interest of contributing to a grander narrative of accomplishment or to reinforce a sense of self based on perceived values of dedication and personal worth.[58] The investigators of this study argue that there is an important relationship between the design of personal archives, the motivations involved in keeping them, and the value(s) one ascribes to digital materials, such as value pertaining to personal legacy and value involved in the continuing construction of self over time.

54 Ibid., 5–9.

55 Ibid., 6–7.

56 Joseph Kaye, Janet Vertesi, Shari Avery, Allan Dafoe, Shay David, Lisa Onaga, Ivan Rosero, and Trevor Pinch, "To Have and to Hold: Exploring the Personal Archive," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1–6, http://jofish.com/writing/tohaveandtohold.pdf (accessed 20 May 2011). In 2006, the research team published the results of this study, which involved interviews with forty-eight individuals as well as on-site visits of their personal archives.

57 Ibid., 5–9.

58 Ibid., 6. The archives of one study participant contained an antiquated laptop computer, which had been used to write her PhD dissertation some years earlier. The authors note that the laptop failed to support ongoing work and the information within it was inaccessible, yet this digital remnant still served as a reflection of accomplishment and reinforced a sense of self based on perceived values of education and dedication.

In 2008, the Digital Lives project[59] reported on a series of in-depth interviews conducted with twenty-five individuals from the fields of politics and the academic arts and sciences to determine how and why private individuals archived in the twenty-first century. In their initial findings, the Digital Lives team concluded that both utilitarian and emotional factors are involved in private appraisal of digital items and their subsequent retention. For instance, the research team found that "quite often interviewees could not specify exactly how a document would be used later but still felt that, as long as there was a possibility that it might be of use, it was worth keeping."[60] This evaluation of digital items for potential future usage, the authors note, is coupled with the appraisal of items for their emotional sentimental value, which individuals calculate by the time and effort expended in creating the item and by contextual factors, such as personal memories, that surround its continued use and custody. In a later study of more than 2,000 completed interview responses, the Digital Lives team found that the most prominent reasons for archiving a digital item were: as a witness to creativity, for sentimental reasons and personal memory, for future reference, to share with colleagues, as a record of past activity or events, and in the interest of posterity.[61] Of these explanations, individuals tended to value the witnessing of creativity above all else, which "might speak for creativity as a core human value and need or as self-validation."[62]

In 2010, Canadian literary archivist Catherine Hobbs stated:

> To see more of the personal context of documentation, we need to understand as much as possible of the creator's intention and thoughts about documenting. Archivists should consider all psychological factors involved when individuals make/keep/destroy documents. The creator's relationship to documentation could involve many emotional and practical aspects….
>
> What were the choices to and motivation for creating the record?[63]

59 Between September 2007 and March 2009, the Digital Lives project explored the evolving nature of personal digital collections and their relationship with research repositories. The project team was drawn from the British Library, University College London, and the University of Bristol.

60 Peter Williams, 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/ (accessed 10 May 2011).

61 Leighton John et al., "Digital Lives: Personal Digital Archives for the 21st Century," 44–45.

62 Ibid. This information is derived from two surveys conducted by the Digital Lives Project. One survey was directed at academics and another at members of the digital public. Sixty-three percent of academics and 45 percent of the digital public cited "as a witness to creativity" as the main reason for archiving digital items, followed by "sentimental reasons" and "personal memory," cited by 15 and 26 percent respectively.

63 Catherine Hobbs, "Reenvisioning the Personal: Reframing Traces of Individual Life" in *Currents of Archival Thinking*, ed. Terry Eastwood and Heather MacNeil (Santa Barbara, 2010), 227–28.

Assessing the impulses and intentions, appraisal decisions, and designations of value within the broader context of private records creation has for some time been the province of personal records archivists. Yet as Figure 1 reveals, PIM research identifies many of the same values commonly attributed to records by personal archivists and serves to complement and reaffirm the already established value folksonomies of personal archives. Value folksonomy is defined here as a classification of terminology often used by both individual records creators and archivists to describe the values of personal archives. Taxonomy is far too formal a term to be used in discussions on the value of personal archives. PIM Identifications of Value are derived from the PIM studies examined, while Archival Identifications of Value are based on discussions on the values of personal records appearing in the archival literature.[64] Given the precipitous rate at which digital material is created and accumulated, PIM research regularly examines how individuals decide what constitutes meaningful or insignificant items in the construction of their personal archives. This is of direct relevance to the provenance of personal archives and as such should be leveraged by those seeking to preserve it in the interests of documentary heritage. The Digital Lives surveys, admittedly an archival-based initiative, are analogous to the PIM studies reviewed in the second section of this paper. They serve as a testament to what can be discovered about personal digital records through proactive engagement with contemporary records creators.

64 See Richard J. Cox, "The Record in the Manuscript Collection," *Archives & Manuscripts* 24, no. 1 (May 1996): 52–61; Verne Harris, "On the Back of a Tiger: Deconstructive Possibilities in 'Evidence of Me,'" *Archives & Manuscripts* 29, no. 1 (May 2001): 8–21; Catherine Hobbs, "The Character of Personal Archives: Reflections on the Value of Records of Individuals," *Archivaria* 52 (Fall 2001): 126–35; Sue McKemmish, "Evidence of Me…," *Archives & Manuscripts* 24, no. 1 (May 1996): 28–45; and Riva A. Pollard, "The Appraisal of Personal Papers: A Critical Literature Review," *Archivaria* 52 (Fall 2001): 136–50.

PIM Identifications of Value

Identity Value: Items involved in the continued construction and expression of the self.
Personal Memory Value: Items facilitating re-encounters with the personal past.
Personal and Familial Historical Value: Items recording individual and shared narrative.
Emotional and Sentimental Value: Items witnessing creativity and achievement, and eliciting emotion.
Functional Value: Items serving as reference or administrative utility in day-to-day life.
Posterity and Legacy Value: Items perpetuating a conception of a life or career for external consideration.

Archival Identifications of Value

Evidential Value: Records documenting functions, activities, and transactions of persons.
Informational Value: Records providing information on significant persons and of potential use for historical and sociological research.
Narrative Value: Records involved in the process of storytelling and autobiography.
Societal and Cultural Value: Records documenting the contemporary character, personality, intimacy, beliefs, and spirituality of individuals within society.

Figure 1: Value Folksonomy for Personal Archives

Personal Preservation of Digital Records

Following mediative processes of personal recordkeeping and private assessments of value, the meaning and purpose of digital materials evolve beyond their initial use and become much more than simple documentation. Individuals engage in initiatives to preserve their valued digital records for extended periods of time. Far from being a one-time static event, these preservation practices command an ongoing personal commitment to keep track of diverse digital records with unique curatorial needs occupying multiple storage areas. Interestingly, the preservation of personal digital records is the one area of the precustodial environment that memory institutions consciously seek to influence through awareness-raising campaigns and other outreach activities. For example, the Library of Congress website contains a series of WebPages dedicated to personal archiving. Individuals can download a best-practices brochure entitled “Preserving Your Digital Memories” and view a digital pres-

ervation awareness video that states: "No matter what type of file you want to save, they all require the same essential preservation strategy: identify what you want to save; decide what is most important to you; organize the content; save copies in different places."[65] But is it really that simple, and do best practices such as the ones proposed by the Library of Congress operate effectively in everyday digital life? The next section investigates the realities of personal digital preservation through an analysis of pertinent PIM literature about localized and online digital preservation.

Localized Digital Preservation

In 2007, Microsoft researchers Gordon Bell and Jim Gemmell stated that a $600 hard drive could hold one terabyte (one trillion bytes) of data, which is "enough to store everything you read … all the music you purchase, eight hours of speech and 10 pictures a day for the next 60 years."[66] In contrast, Neil Beagrie of the British Library contends that with current exponential increases in levels of computing power and storage capacity it will soon be possible "to envisage individuals being able to store the equivalent of all the texts in the Library of Congress on their PC."[67] All of these bits and bytes of personal data are housed within localized preservation environments consisting of the HDD of desktop and laptop PCs, USB drives, CDs, DVDs, as well as within the memory media of external devices such as digital photo and video cameras, media players, smartphones, and, more recently, tablet technologies. Furthermore, as digital information is inherently bound to the software necessary to interpret and render its binary sequences of zeroes and ones, personal data are further distributed across multiple file formats, many of which are proprietary in nature. In short, these localized environments contain an ever-increasing amount of digital files stored in myriad physical (hardware) and logical (software) carriers.

65 Library of Congress, *Preserving Your Digital Memories and Why Digital Preservation is Important for You,* http://www.digitalpreservation.gov/you/ (accessed 30 June 2011). Video transcript http://www.digitalpreservation.gov/videos/personal_archiving/index.html (accessed 30 June 2011).

66 Gordon Bell and Jim Gemmell, "A Digital Life," *Scientific American* 296 (March 2007): 58–65. As of 2011, a one-terabyte external hard drive is sold by major Canadian electronics retailers for approximately $100 to $150, with external two-terabyte hard drives available for marginally more.

67 Neil Beagrie, "Plenty of Room at the Bottom? Personal Digital Libraries and Collections," *D-Lib Magazine* 11, no. 6 (June 2005), http://www.dlib.org/dlib/june05/beagrie/06beagrie.html (accessed 2 June 2011). Beagrie quotes Gordon Moore's seminal article on the increase of computing power "where there is roughly a doubling of the number of transistors on integrated circuits every 18 months for the same unit cost." See also Gordon E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* 38, no. 8 (April 1965): 1–4, download .intel.com/museum/Moores.../Gordon_Moore_1965_Article.pdf (accessed 2 June 2011).

Conventional strategies for personal archiving in local environments involve backups (performed manually or automatically), exporting files to external storage media (CDs and DVDs), and the preservation of entire platforms (the computer, its peripherals, and installed software).[68] Periodic backups are the most basic form of data replication; redundant copies are created and stored to protect against loss caused by user errors, disk or other hardware failures, software errors, and natural disasters.[69] As an archiving behaviour, backing up also refers to the replication of data stored in media players, digital cameras, and smartphones, which involves the synchronization of data distributed among these external devices and the system hosting the backup. A similar form of replication involves copying valuable digital items to external hard drives, CDs, DVDs, USB flash drives, and other contemporary storage media, which are then labelled and placed in a physical storage area. As a long-term archiving strategy, this digital content may be migrated to successive physical carriers, depending on the preservation regime followed by the individual.[70]

Online Digital Preservation

While it may also be referred to as cloud storage or Web 2.0 storage, all online digital preservation operates on the premise of a client-server relationship: an individual's digital records are stored on a server infrastructure they neither own nor control as far as how often data are backed up or how long they are retained.[71] Quite often, individuals take advantage of free or moderately priced online storage and distribute the custody of their records across multiple online services (multiple servers), resulting in silos of digital storage as opposed to centralized repositories. Conventional strategies for online personal archiving include the following: the email-repository strategy; storing records on commer-

68 Catherine C. Marshall, "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field," *D-Lib Magazine* (March/April 2008), http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html (accessed 2 June 2011).

69 Ann L. Chervenak, Vivekanand Vellanki, and Zachary Kurmas, "Protecting File Systems: A Survey of Backup Techniques," in *Proceedings of the Joint IEEE and NASA Mass Storage Conference* (March 1998), 17–31, http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.7765 (accessed 2 June 2011). People may protect their file systems with either a full backup, thereby copying the entire contents of the file system, or with an incremental backup, which copies only those files that have been modified since the previous backup.

70 A final strategy is the full retention of an entire computer platform (in place of exporting data and software), which may occur upon the purchase of a new and faster computer system.

71 Personal computer users (the client) request data from a more dynamic and often third-party computer (the server), which responds by sending the requested data back to the user. Clients may request data from a variety of computer servers responding with database information, WebPages, email, or streaming media. Individuals also participate in this client-server relationship in social media when publishing blogs, updating Facebook status, tweeting, or posting images to Flickr or videos to YouTube.

cial file-sharing platforms, social-networking services, or a blog/podcast publishing service; and soliciting remote storage from online service providers.

With the almost unlimited storage capacity of email and its proven utility as a personal information management tool, it seems logical that individuals would extend archiving practices to their email. Email provides the ability to send and store many of the same file classes found in localized preservation environments through attachment or embedding options. Keeping these files online is in itself a preservation measure; however, individuals can choose to repatriate their files from commercial computer servers to bring them back under their control in localized storage in a process that may be referred to as "data liberation."[72] Email client software uses one of two protocols for message retrieval: Internet Message Access Protocol (IMAP), which stores messages on a mail server; and Post Office Protocol (POP), which stores messages on the HDD of the personal computer.[73] In order to capture and preserve an email account(s), a user must enable POP with their mail provider (such as Gmail or Hotmail), access their email account with client software (Thunderbird or Outlook), locate and export the email files (.msf, .pst, .eml, etc.) within the personal computer's file system, and preserve these files in a localized environment.

An extension of the email-repository strategy is to upload and store digital records with commercial file-sharing platforms (for example, YouTube, Google Docs, or Flickr), with a social-networking service (Facebook, or Google+), or with a blog/podcast publishing service (Blogger, Twitter, or iTunes). As they do with email, individuals may choose to keep their digital files on Web 2.0 platforms or recapture this content for localized storage. Although some content uploaded to Web 2.0 applications may indeed be digital surrogates of files already kept in localized preservation, content such as tweets or Facebook posts irrevocably circumvent the HDD of the personal computer, while some content uploaded to YouTube, Flickr, or Blogger may be the only existing versions. Repatriating personal websites and more advanced Web 2.0 content may be as simple as re-downloading files to a personal computer or clicking the export button, yet some content may require a more specialized software application such as Warrick, a utility that searches the Internet Archive, Google, Bing, and Yahoo for cached versions of WebPages and stores

72 Data liberation is a relatively new phenomena brought about by the ubiquity of Google products. The Data Liberation Front is a team of engineers at Google who seek to inform the public about how to export data from Google products such as Gmail, Picasa, Blogger, and YouTube if and when people decide to gain more control over their data. See http://www.dataliberation.org/home (accessed 30 July 2011).

73 IMAP messages remain on a web server until the user deletes them from the inbox or other mail folders. When POP messages are retrieved from the mail server and stored locally, they are typically deleted from the mail server. Google Mail, however, provides POP users with the option to keep messages on their server even after they have stored them locally.

them on the user's personal computer.[74] In the case of social media, individuals can use utilities such as ArchiveFacebook, a web browser add-on that captures and exports all file data from their Facebook account to a directory on their personal computer, thereby making it available for localized preservation.[75]

A final online preservation strategy is to deposit digital items with commercial cloud storage services:

> Cloud computing is a model for enabling ubiquitous, 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.[76]

Cloud storage is built on the client-server architecture: an individual transfers files from one personal computer to a commercial server, where the files are stored and can be accessed from any personal computer with an Internet connection. This system is similar to using File Transfer Protocol (FTP) client software to transfer files over the Internet by uploading them to an FTP host server. Unlike the email-repository strategy, cloud storage provides for larger files and greater numbers of files, often at no financial cost or for a nominal storage fee. For example, Dropbox allows users to store 2GB of data on their servers for free but incrementally charges for amounts exceeding that limit.[77] Other cloud storage providers, such as Mozy, ChronicleofLife, MyEvents, Legacy Locker, and Making Everlasting Memories, provide ongoing storage

74 Frank McCown, Joan A. Smith, Michael L. Nelson, and Johan Bollen, "Lazy Preservation: Reconstructing Websites by Crawling the Crawlers," in *Proceedings of ACM WIDM* (2006), http://www.cs.odu.edu/~fmccown/pubs/lazyp-widm06.pdf (accessed 15 July 2011). See also http://warrick.cs.odu.edu/. As search engines crawl and index the Web, encountered pages are temporarily stored in a cache. Cached pages reduce network traffic and improve the responsiveness of the Web by providing clients with access to the search engine copy rather than the original server copy. Put differently, cached pages sit between the client request and the original server response. Another software program that can be employed for website harvesting is the HTTrack website copier; see http://www.httrack.com/ (accessed 15 July 2011).

75 Frank McCown and Michael L. Nelson, "What Happens When Facebook Is Gone?" in *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital libraries* (2009), http://portal.acm.org/citation.cfm?id=1555440 (accessed 15 July 2011). See also the Mozilla Firefox add-ons page at https://addons.mozilla.org/en-US/firefox/addon/archivefacebook/ (accessed 15 July 2011). This add-on captures and typically stores the data in C:\Documents and Settings\user\Application Data\Mozilla\Firefox\Profiles\. Similar services include TwapperKeeper, Twistory, and Keepstream.

76 Peter Mell and Timothy Grance, "The NIST Definition of Cloud Computing (Draft)," National Institute of Standards and Technology, US Department of Commerce (January 2011), www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf (accessed 15 July 2011).

77 See Dropbox, homepage, http://www.dropbox.com/ (accessed 15 July 2011). Dropbox currently offers 50GB, 100GB, and 350+GB of storage space. Dropbox also creates an exact mirror of a user's cloud content on his/her local PC or laptop HDD.

and access, automatic and scheduled data backups, records of usernames and passwords, and encryption for all types and forms of personal files in exchange for subscription fees ranging from five dollars per month to one dollar per megabyte of permanent storage space. Repatriation of files stored in the cloud typically involves the subscriber's transfer of content back to the local environment or a request for a disk copy of their digital materials from the commercial service.

The degree to which personal digital preservation strategies, whether local or online, are successful is also subject to determinants both within and beyond the control of individuals. For instance, Microsoft researcher Catherine Marshall has found that although individuals often have the best intentions of adhering to preservation principles, everyday practices may belie these principles when people begin to rely implicitly on sporadic backups and unsystematic file replication as the primary modes of preservation.[78] Similarly, as digital preservation is an ongoing series of curatorial actions, individuals may not be able to invest the required effort in all of their digital records all of the time, and some records may fall victim to what Marshall identifies as "benign neglect" brought on by the burden of digital stewardship.[79] Additionally, as the status and terms of service of email providers, social-networking sites, file-sharing platforms, and cloud-storage providers change over time, accounts and profiles may be deleted or deactivated by these entities without proper notification being sent to those who have uploaded and otherwise entrusted their data to a service provider. Similar issues arise when companies discontinue the production of required peripheral computer hardware such as eight- or 5¼-inch floppy disks, or when the array of proprietary digital formats becomes so confusing as to make individuals unsure of the consequences of their format choices.[80]

Conclusion: Implications for Archivists

This paper has introduced and discussed PIM perspectives on personal computing environments to reveal actualities of personal recordkeeping strategies, identifications of value, and digital preservation practices heretofore

78 Marshall, "Rethinking Personal Digital Archiving, Part 1." Adding to the complexity of backup strategies, Marshall notes, "It is surprising how many people have made backups without ever trying to perform a restore, or have restored files and accidently overwritten them."

79 Ibid. Marshall describes benign neglect as setting aside digital records that matter in a special place and hoping they will be there when the time comes to retrieve them.

80 Marshall, "The Long Term Fate of Our Digital Belongings." Another element affecting the success of personal preservation strategies is the potential of attacks from malicious software (malware) that is able to circumvent even the most trusted anti-virus software and system firewalls.

undisclosed by archival scholarship in the areas of both personal and digital records. A recent analysis of personal records by archivist Jennifer Meehan reiterates the potential of PIM in archives:

> [T]he archivist must also seek to understand the sequence in which the records may have been created, how the creator may have used the records, and to what end(s). Doing so requires the study and analysis of the creator, not only in terms of *what* he or she was doing, but also *how* and *why*.[81]

Indeed, comprehending the intricacies of the precustodial environment has been of the utmost concern to archivists as it elucidates the provenance of personal records and is, especially in the digital era, integral to their ongoing preservation. An issue, however, has been a stark inability to examine the nascent stages of personal records creation as they develop *in the here and now*. Moreover, the information about twenty-first-century precustodial environments that does exist is typically gleaned from seemingly analogous personal papers created before the digital era.[82]

PIM research provides a lens through which archivists may view not only contemporary recordkeeping practices of individuals but also private appraisal decisions and personal digital preservation strategies closer to the point of their actual performance. Like researchers in PIM studies, archivists will also need to keep pace with society as it moves away from the dominance of personal computers toward increasingly cloud-based computing environments. At the same time, archivists may also mine past PIM research to discover how society has been creating records with personal computers and associated technologies over the past twenty-five years.[83]

Some of the personal recordkeeping practices discussed in this paper have been derived from PIM studies taking place in work environments as opposed to only home settings, and yet these should not be conflated with institutional records and archival management practices for a number of reasons.[84] First,

81 Jennifer Meehan, "Rethinking Original Order and Personal Records," *Archivaria* 70 (Fall 2010): 38–39. Emphasis in original.

82 Ibid., 40–43. For example, in her excellent discussion of the original order of personal records, Meehan references the archival collections of three individuals whose lives, and therefore records creation, ended before many digital documentary forms discussed in this article even existed.

83 PIM emerged as a unique branch of human–computer interaction research in the 1980s with works such as Thomas W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems* 1, no. 1 (January 1983): 99–112, and M.W. Lansdale, "The Psychology of Personal Information Management," *Applied Ergonomics* 19, no. 1 (March 1988): 55.

84 The workplace studies include these articles cited earlier: Henderson, "Personal Document Management Strategies" (knowledge workers); Barreau and Nardi, "Finding and Reminding" (managers); Mackay, "More Than Just a Communication System" (full-time researchers,

discussions of government and corporate digital archives often centre on appraisal criteria and systematized records classification according to high-level functions, activities, and business processes, with little consideration given to how these methods and directives are applied and followed by actual human records creators on a day-to-day basis.[85] PIM studies, on the other hand, focus on how individuals interact with their computing environment, with little if any discussion of business functions or directives and standards for corporate recordkeeping. Although PIM research may at times be carried out in formal office settings, it privileges the discovery of day-to-day personal computing behaviours in those office settings over the analysis of the functions those offices perform in an organization. And, as Catherine Hobbs observes, there is an interplay of the personal and the professional in private archives, whereby it is possible to identify how private life influences one's work.[86] Inverting this observation, activities learned and performed in occupational life, such as choices of hardware and software or strategies for email and file management, may inform how digital records are created, managed, and preserved in private life. While the precise mechanics of personal recordkeeping and archiving at home and at work may not be exactly alike, there will undoubtedly be points of convergence and overlap as, with the growing predominance of mobile devices and cloud computing, the boundaries of these two realms may at times become "very hard to delineate clearly, if at all."[87]

In her often-cited 1996 article, "Evidence of Me …," Australian archivist Sue McKemmish asks: "What characterises the recordkeeping behaviour of individuals and what factors condition that behaviour? What range of 'personal recordkeeping cultures' can be identified?"[88] PIM studies have exposed patterns of personal recordkeeping behaviour that have begun to answer some of the questions archivists have been asking about personal records for the better part of sixteen years. As PIM research continues to evolve to examine all types of people across many different types of environments, these findings may come to produce a mosaic of personal recordkeeping and archiving behaviours from which archivists can draw to assist them in their work with twenty-first-century personal fonds.

managers, computer scientists, academic professionals, and administrators); and Whittaker and Sidner, "Email Overload" (office workers).

85 See Inge Alberts, Jen Schellinck, Craig Eby, and Yves Marleau, "Bridging Functions and Processes for Records Management," *Canadian Journal of Information and Library Science* 34, no. 4 (December 2010): 365–90.

86 Hobbs, "Reenvisioning the Personal," 223–24. To illustrate this point, Hobbs refers to the personal archives of Canadian writer Carol Shields. Hobbs was able to pick up on the interplay and connection between events occurring in Shields' daily life and in her work as a novelist.

87 Leighton John et al., "Digital Lives: Personal Digital Archives for the 21st Century," 32.

88 McKemmish, "Evidence of Me …," 29.

For example, the personal recordkeeping strategies identified in a number of PIM studies reviewed in this paper may be incorporated into archival workflows in the areas of intellectual and physical arrangement of personal archives; this information may come to guide archivists through, for instance, the reconstruction of original order. This reconstruction of original order may mean replicating the folder directories of a frequent filer or pro-organizer as they appear at the point of acquisition. Conversely, reconstructing the seemingly chaotic, but nevertheless meaningful, original order of records belonging to a no-filer who relied on keyword searching for the rediscovery of his files may require that the archivist ascertain how certain desktop search engines indexed files and what filtering options were available to retrieve data from personal computer hard drives.[89] Maintaining some semblance of original order is also important for users of personal digital archives as it provides cues for discovering the relationships between records that would otherwise be lost if the records were haphazardly reorganized by archivists.

While archivists may not be able to analyze personal recordkeeping behaviours the same way they do corporate records management, they may come to develop frameworks for analyzing patterns of those behaviours based on personal information management research and thereby gain better precustodial understandings of the meta-cultures and practices of personal recordkeeping. Additionally, appraisal decisions disclosed by PIM research participants may influence the way archivists conduct their appraisals of personal digital archives. Criteria such as identity, personal memory, familial and historical factors, emotional and sentimental factors, as well as posterity and legacy value should serve as benchmarks in the process of selecting records within a personal digital archives for permanent retention. PIM research is also relevant to archival work in discovering the true breadth of personal digital fonds or collections. Never before have the records that constitute a personal fonds or collection been so widely distributed across multiple locations, and as shown in this paper, PIM research elicits key information on how and where individuals preserve their valuable records in both online and offline environments. This is especially germane in an era where personal digital records are being kept by non-familial third parties more for their commercial value than for their cultural and historical value. Indeed, archivists must know where to look for personal digital records before they can appraise and select them in the interest of producing a truly representative personal fonds or collection.[90]

89 Before a keyword search can be executed, an index database (list of files and their locations) must be built. While this index building often occurs automatically, certain computer drives must be manually indexed by the user.

90 While it is true that PIM studies do not address the issue of locating those individuals within society whose records warrant archival preservation, these studies do identify why, how, and where individuals archive their digital records in certain temporal and technological contexts

The archival profession has a long tradition of incorporating ideas and approaches from other academic disciplines, such as using diplomatics in establishing the authenticity of electronic records, collaborating with libraries in the development of metadata schema, and adopting the Open Archival Information System reference model as a digital curation standard. Absorbing and applying elements of PIM in archival endeavours is an extension of this profitable tradition. While information technology specialists and corporations can leverage PIM research to develop new and better hardware and software applications, so too can archivists exploit this information to better understand how documentary forms are created, accumulated, used, and preserved by individuals in the digital present while, at the same time, cultivating new and more effective means of archiving personal digital records.

Jordan Bass *is an archivist with the Faculty of Medicine Archives, Neil John Maclean Health Sciences Library at the University of Manitoba. He is a graduate of the MA program in archival studies at the University of Manitoba, where his thesis focused on meeting the challenges of archiving personal records in the digital age. His current research interests focus on discovering cultures of recordkeeping and archiving behaviours in personal digital environments.*

and, as such, should inform the mediation of personal digital records in both the precustodial and archival environments discussed at the beginning of this paper.