## Table of Contents

Introduction to *I, Digital*

*Christopher A. Lee.* ......................................................... 1

**Part 1—Conceptual Foundations and Motivations**

And Now the Twain Shall Meet: Exploring the Connections between PIM and Archives

*Christopher A. Lee and Robert Capra.* ..................................... 29

Ghosts in the Machine: Towards a Principles-Based Approach to Making and Keeping Digital Records

*Adrian Cunningham.* ............................................................. 78

Challenges and Opportunities for Personal Digital Archiving

*Catherine C. Marshall.* .......................................................... 90

Evidence of Me . . . in a Digital World

*Sue McKemmish.* ................................................................. 115

**Part 2—Specific Genres and Document Types**

“Three Backups Is a Minimum”: A First Look at Norms and Practices in the Digital Photo Collections of Serious Photographers

*Kristina M. Spurgin.* ............................................................. 151

Collecting the Externalized Me: Appraisal of Materials in the Social Web

*Christopher A. Lee.* ............................................................. 202
Part 3—Implications for Memory Institutions

Take It Personally: The Implications of Personal Records in Electronic Form
Rachel Onuf and Tom Hyry .................................................. 241

Making It Usable: Developing Personal Collection Tools for Digital Collections
Leslie Johnston ................................................................. 257

Curating the I, Digital: Experiences at the Bodleian Library
Susan Thomas ................................................................. 280

Bibliography .................................................................. 307

Contributors .................................................................. 363

Index ............................................................................. 367
Introduction

Christopher A. Lee

…only if we better understand personal recordkeeping practices can we hope to do a better job at capturing and preserving those records for posterity.

—Adrian Cunningham, 1996

Many archivists and archival institutions have a collecting mission that includes personal papers, manuscripts, and other noninstitutional materials. Despite a massive increase in the volume and complexity of personal digital collections, the literature designed specifically to guide archivists’ thinking about personal digital materials has long been limited to a few scattered journal articles and research project websites. I Digital aims to fill this gap. It explores issues, challenges and opportunities in the management of personal digital collections, focusing primarily on born-digital materials generated and kept by individuals, as opposed to electronic records that are generated within and managed by formal organizational recordkeeping systems. It represents a convergence and synthesis of literature and thinking about how cultural institutions can grapple with new forms of documentation and how individuals manage (and could potentially better manage) the digital information that is part of their contemporary lives. One of the book’s main objectives is to expose the archival profession to an active body of highly relevant research that has traditionally been published outside of the archives literature and not typically represented at archives conferences. It also seeks to expose other audiences (e.g., those with an interest in personal
information management) to the principles, insights and activities of the archival profession.

An Obligatory Note on Terminology

*I, Digital* bridges discussions from a diverse set of cultures and disciplines. An endemic characteristic of such efforts is “creative tension” in the use of various terms. In cases when it is considered particularly important to do so, authors of the chapters often provide working definitions of specific terms. The most obvious divide is between the language of Personal Information Management (PIM) and Archives and Records Management (ARM); even these very labels are subject to dispute.

There is also a diversity of terms used within the literatures of the collecting professions (librarians, archivists, museum curators) over what to call the *set of traces* that are left behind by individuals as they engage in the activities of their lives. One distinction that has received considerable attention—particularly in the U.S. literature—is between institutional records (records whose primary use is related to documenting and supporting the activities of an organization) and “manuscripts” (materials whose primary use is related to the activities of an individual in his or her personal life). The British Library, for example, favors the term *eManuscripts* for digital materials created by individuals. The term *manuscript* carries specific mechanical connotations (being written by hand) that are often not strictly accurate. An alternative term, *personal papers*, is often used instead. However, a great deal of personal material is not actually captured or retained on paper, so this is also not ideal.

A phrase that one will frequently encounter in this volume is *personal archives*. It is important to note that Australian authors tend to use the term *archives* much more holistically than U.S. authors. In the Australian archival literature, archives are all records of “continuing value,” regardless of where they reside or who is currently caring for them.¹ By contrast, the North American archival literature tends to consider archives to be the places, collections or set of associated activities that reside on the receiving end of the “Archival threshold.” From this perspective, many types of documentation can serve as “personal records,” but they are only personal archives when
they have been transferred to the care of professional archivists within a formally recognized repository.

Another commonly used phrase is *personal recordkeeping*. The term *recordkeeping* is usually used to include both activities that are often labeled *records management* and activities that are often labeled *archival administration*—i.e., recordkeeping encompasses care for records both within their creation environment and in any environments that are used to care for them thereafter. The scope of personal recordkeeping is very close to the stated scope of much writing about PIM, with the primary difference being the strong emphasis of the former on the *record* as the primary unit of analysis. A record is a subset of all information created by an entity (individual or organization) that has been designated as requiring a certain degree of coherence and integrity over time, in order to serve as evidence or documentation of something.

I have chosen to use the term *personal collection* within the title of this book, in an attempt to find the closest thing to neutral ground that might exist within this complex ecosystem of rich, sometimes complementary, but sometimes conflicting, terminology. Roughly speaking, a personal collection is a sizable aggregate of an individual’s personal traces that he/she or someone else has identified and attempts to manage over time as a relatively coherent unit in order to reflect something important about that individual. This is certainly not a formal definition, nor is it even a working definition that is applied consistently throughout the book. Instead, it is simply a working notion of a common area of interest that has served as a point of intersection for writings by a very thoughtful set of authors who have contributed to this volume.

**Trends Related to Personal Digital Collections**

Historically, individuals and families have accumulated and managed personal collections. Most of these collections have been relatively small and have not left the homes of the collectors. Many of the materials have been lost due to the “hazards of time.”² However, collections of a few prominent individuals and families made the transition into collecting institutions. Many cultural institutions were initially seeded by personal collections of influential people.
Over the past half century, five trends have dramatically changed the nature and status of personal collections: professionalization; professional convergence around digital preservation; increased storage capacity available to individuals; distribution of collections; and research focusing on individual stories.

First, work within collecting institutions has become increasingly professionalized. This has included specialization, professional education, professional associations, conferences, journals, specialized language, and social norms. As work becomes more professionalized, there is more consensus about what types of activities fall within the scope of the profession and which do not. This is what Andrew Abbott has called the “jurisdiction” of a profession. Professionalization has dramatically raised the expectations for how institutions must care for and provide access to collections in order to be considered “professional” in their activities. For example, the archival literature reflects numerous widely held expectations about engagement with records creators, formal transfer agreements, standardized description, environmental controls, storage materials, access controls, reference services, and outreach; there is a diversity of views about appropriate approaches to implementation and degree of emphasis on the various expectations, but it would be very unusual for anyone in a professional forum to flatly dismiss their relevance to archivists.

Similarly, museum curators and librarians operate within their own sets of expectations about the scope and nature of their work. Setting a clear boundary around a set of professional activities is a powerful way to focus attention, mobilize resources, and establish sufficient coherence across the activities to get good things done. However, the boundary also increases the “transaction costs” for interacting with people and groups who are not part of the profession. The fragmentation of libraries, archives, and museums, for example, can inhibit those institutions from collectively responding to the challenges and opportunities of personal digital collections. A strong focus within the professional archival literature on methods and procedures within the context of formal repositories can also inhibit archivists from working with everyday citizens, genealogists, and various other types of amateur enthusiasts to collectively ensure that personal digital archives can be meaningfully and appropriately used in the future. In short, long-term care for personal digital collections faces the two sides of the professionalization
Introduction

coin: focused development and progress within associated professions has spawned many valuable resources, concepts, and methods that could be mobilized to address personal digital collections; while these same developments make it difficult to mobilize across the professions. Collective efforts will require concerted boundary spanning, and I, Digital is one step (hopefully, among many) in that direction.

A second trend is that previously distinct communities have come to recognize that they share challenges associated with long-term care of digital resources. Digital objects are different from purely spatio-temporal objects in that they are recreated each time they are used, based on interactions of numerous technological components. In order to ensure meaningful access and use of digital objects over time, one must plan for and respond to changes to the technological components, the ways the components interact, user needs, and mechanisms for use. Digital preservation is a multilayered, complex endeavor. There has been increasing recognition—particularly since the mid-1990s—across both collecting institutions and other types of organizations responsible for stewardship of data over long periods, that many of the fundamental issues of digital preservation are shared across contexts and could be better served through communication across institutional and professional boundaries. In many ways, this trend has the potential to counterbalance the professional fragmentation issues discussed above.

A third important trend is that individuals have gained more ability to create and store materials that they find meaningful, useful, or simply more convenient to keep than to discard. In the developed world, a typical consumer can conveniently create a vast array of information artifacts that were previously much more difficult and expensive to produce, including text, images, moving images, audio, structured data, microcontent (e.g., tags, bookmarks), and many combinations thereof. Although computers regularly overwrite many types of data, the default computer behavior related to user-created files usually is to retain them until the user makes a conscious act to delete them. Various factors make it increasingly less likely for individuals to delete files that they have created, resulting in a proliferation of data in personal digital archives:

- Deleting takes time, attention, and a conscious commitment to the idea that something will definitely no longer be needed.
• The cost and size of the storage medium required to store a given number of bits has decreased dramatically every year and is likely to continue decreasing for the foreseeable future, which means that many individuals do not have to worry about running out of space.

• The number and types of storage media that an individual may have at his/her disposal (e.g., thumb drives, telephones, cameras, mobile phones, desktop computers, laptop computers, external drives, and disks) also continues to grow.

• Increasingly sophisticated technologies for both searching over and reusing digital data provide a strong case for keeping things “just in case” one might find them to be useful in the future.

A fourth, very closely related, trend is that personal collections have become more widely distributed than in the past. The issue of personal fonds being spread across multiple locations is certainly not new, but the magnitude of this phenomenon is. In addition to the various types of storage media discussed above, there has been a significant movement of many types of personal information into “cloud computing” environments. Whereas many organizations are entering formal arrangements in which they pay “cloud providers” to host their content, the services upon which individuals rely are often provided for “free.” Rather than providing direct financial payment, individuals usually pay by allowing the company access to their personal data, willingness to view advertisements, or both. Such arrangements rarely provide the individuals with much control over how their information will be managed or retained over time. Individuals often use different computing environments to carry out different tasks; this can be because of limitations of the technology (interoperability issues) or based on deliberate decisions of individuals (e.g., switching to email when a blog exchange has become too personal, maintaining different social media profiles for different social contexts, keeping some family photos in a shared online space while keeping others on a hard drive).

Finally, researchers have placed considerably more emphasis than they did several decades ago on the importance of personal stories, voices, and perspectives. Many fields of study in the social sciences and humanities now use personal accounts as sources of data and recognize aspects of personal subjectivity to be legitimate topics of research. Numerous schools
of thought in social theory, literary criticism, and historiography have also questioned positivist or “grand narrative” approaches, advocating instead for a focus on individual interpretations and perceptions. Social historians, feminist scholars, family historians, and genealogists all rely heavily on personal information traces (within archives or elsewhere) to answer their questions. This trend foregrounds the importance of personal digital collections and suggests many communities of interest that have a strong stake in appropriate treatment of personal digital information.

These five trends have a significant impact on the system of professions and professional activities related to personal digital collections. The trends greatly enable many new activities, but they can also impose serious constraints on collaboration and innovation. The chapters of *I, Digital* were written within this context.

**Building on a Diffuse Literature**

The chapters of *I, Digital* build on—and will hopefully contribute to—several different bodies of literature. Although it is impossible to completely disentangle them, I have found it useful to consider five associated areas of literature: the administration of personal papers and manuscript collections; electronic recordkeeping; digital data recovery and forensics; personal information management; and the design of tools for user-generated collections.

**Administration of Personal Papers and Manuscript Collections**

Libraries and archives have a long history of collecting personal papers and manuscripts. In the United States, there were numerous organizations (particularly historical societies) that focused on the care of both organizational records and manuscript materials long before the archival profession was formally professionalized. Ever since the formation of the National Archives and the Society of American Archivists in the 1930s, personal papers and manuscripts have been part of the archival profession’s charge, but their status has often been ambiguous. Several prominent authors have argued that the primary focus of the profession should be on official government records. Some have gone so far as to dismiss personal papers and manuscripts as not being true archival records and thus not
worthy of professional archival attention. Much ink has been spilled onto the pages of North American archival journals over the status of both non-institutional records and the “manuscripts tradition” more generally. The Australian archival literature has been characterized by a similar tension, addressed most explicitly in a special issue of Archives and Manuscripts in 1996 on “Personal Recordkeeping: Issues and Perspectives.”

One consequence of the profession’s ambivalence is that practical guidance and empirical findings about the archival treatment of personal papers and manuscripts has tended to be underrepresented in the published professional literature. This has been particularly true in the last several decades, as notions of evidential value, recordkeeping systems, and institutional accountability have driven much of the research. Ironically, this has occurred at the same time that the boundary between personal and official recordkeeping systems has continued to blur. Given how much of records management is carried out by individuals “at the desktop,” and the extent to which personal and organizational materials are intermingled on those same desktops, all records professionals can benefit from a more informed view of personal information practices and principles. In 2006, Toby Burrows lamented, “Though the range of issues relating to personal electronic archives has been relatively well-documented, there is as yet little in the way of systematic investigation of solutions and approaches.” However, there are positive signs that this is changing.

A growing number of librarians and archivists have been gaining hands-on experience with personal digital collections. Several informative conference papers and publications have come out of this stream of activity, often in the form of case studies. Two important, very relevant projects have been Digital Lives and Personal Archives Accessible in Digital Media (Paradigm). The Digital Lives project aimed “to explore how individuals use, organise and manage their digital collections.” The Paradigm project investigated “issues involved in preserving digital private papers through gaining practical experience in accessioning and ingesting digital private papers into digital repositories, and processing these in line with archival and digital preservation requirements.” The Paradigm project’s most visible product has been a “Workbook on Digital Private Papers.” Another recent project generated a white paper based on a series of site visits and meetings of those working with the born-digital components of three significant literary
collections. In her book *Electronic Records in the Manuscript Repository*, Elizabeth Dow summarizes important considerations and discusses strategies for approaching collecting scenarios in which “the materials are in fact the donor’s so the curator works with them only as a consultant, armed with knowledge, but no actual power.” And in “Managing Electronic Records in Manuscript Collections,” Michael Forstrom articulates the Beinecke Library’s approach to describing electronic records.

**Electronic Recordkeeping**

The archival literature has given at least some attention to issues of computerization since the 1940s, and by the late 1970s and early 1980s the archival literature included many reports on the actual and potential use of computers to support the internal operations of archives. Numerous articles and reports have also called for members of the archival profession to revise their theories and practices, and take a more active role in addressing the management and preservation of digital objects. Despite these calls for action, active discussion of electronic records in the archives literature is still surprisingly young and under-developed.

The Society of American Archivists (SAA) made important, early contributions with the publication of proceedings of a 1979 conference about machine-readable records and *Archives and Manuscripts: Machine-Readable Records*. The Records and Archives Management Programme (RAMP) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) also issued publications that advanced early thinking on electronic records. Several other important books about electronic records were introduced into the archives and records management literature in the 1990s, and SAA published a set of eight electronic records case studies in that decade.

The electronic records literature has grown considerably since the 1990s. There have been many books, as well as significant treatment in the serial publications of records management professional associations, including *Information Management* published by ARMA International. Professional archival journals have also included a substantial number of pieces that address electronic records, but the attention to personal digital collections in archival journals has been relatively limited. Much of the recent attention to electronic records management literature has been driven by concerns about legal compliance, with the enactment of policies and mandates including
the U.S. Health Insurance Portability and Accountability Act (HIPAA) (1996),
European Data Protection Directive (1998), Information Security Management,
Code of Practice for Information Security Management (2000), U.S. Sarbanes-
(2006), many prominent court cases, and the growth of a massive electronic
evidence discovery industry. Legal compliance has become a driving factor
for electronic records management in many organizations, leading one
author to conclude that all other organizational “considerations must yield
to information’s role in litigation processes.” Several initiatives have created
guidance for record creators (within organizational contexts), including
Managing the Digital University Desktop (MDUD) and the Preservation of
Electronic Mail Collaboration Initiative (EMCAP).

When the archival literature on electronic recordkeeping has directly
addressed personal records, the discussion has usually focused on issues
of professional jurisdiction rather than elaborating specific principles or
methods for dealing with personal electronic records. Frequent topics of
discussion have been whether manuscript curators have been marginalized
by the major electronic records projects of the 1990s, whether a definition
of archival records as conveyors of evidence is inconsistent with the goals
of manuscript curators, and whether an excessive focus on accountability
diminishes the larger societal mission of archives as conveyors of social
memory. As with so many professional issues, competing interpretations of
fundamental terms are major sources of tension and confusion. It should be
no surprise that an archivist responsible for manuscript collections would
take issue with the statement that “archivists are in the evidence business” if
she interprets “evidence” to mean solely “that which is admissible in a court
of law.” Likewise, “archives are instruments of accountability” would also not
sit well with her, if accountability is taken to mean only “ensuring that public
officials appropriately discharge their legal duties.” The working reality of
many archivists—who are often responsible for both institutional and non-
institutional records—requires working definitions of core terms that are quite
broad and multifaceted. Recordkeeping practices serve multiple interests,
objectives, and “accountabilities” along many dimensions at various levels
simultaneously: legal, institutional, professional, ethical, cultural, social,
historical, and personal.
Contemporary scholarship about electronic records has been strongly influenced by the continuum model of recordkeeping. The archival literature on electronic records over the past several decades has placed a strong emphasis on understanding and engaging records creating environments, rather than waiting passively for records to eventually cross the archival threshold. One of the fundamental questions for the profession—and one that has received very little attention—is how to successfully apply continuum thinking to electronic records that are created outside of formal organizational recordkeeping structures.

Archivists and librarians have recently begun to focus considerable energy and resources on web archiving, i.e., identifying, harvesting, and preserving materials available from the Web. As with materials residing on personal computers, the boundaries between personal materials and official records are often very unclear. Professionals engaged in web archiving must often consider annual reports, organizational publications, blogs, wikis, and YouTube videos all in the same set of selection activities, rather than insisting on a stark line between the official and unofficial. Several recent initiatives have focused on appraisal of “social web” materials, which are often characterized as forms of personal expression but also document much of the “business” of contemporary society.

Digital Data Recovery and Forensics

The literature on digital archives tends to place a great emphasis on the “virtual” (i.e., intangible) nature of electronic resources. Computer systems have “an illusion of immateriality by detecting error and correcting it,” but it is essential to recognize that digital objects are created and perpetuated through physical things or phenomena (e.g., charged magnetic particles, pulses of light, holes in disks). This materiality brings challenges, because data must be read from specific artifacts, which can become damaged or obsolete. However, the materiality of digital objects also brings unprecedented opportunities for archival description, interpretation, and use. Digital resources are composed of interacting components that can be considered and accessed at different levels of representation (bitstream; file as accessed through a filesystem; files as rendered through specific applications; records composed of multiple files; larger aggregations such as websites). Some of the “types of historical indications that an electronic text may contain”
include program of origin, system of origin, traces of transmission, and typists and their interaction with programs. To ensure integrity and future use, archivists must make decisions regarding treatment at multiple levels of representation.

There is a substantial body of information within the underlying data structures of computer systems that can often be discovered or recovered. Recovery of data from physical media has been a topic of discussion in the professional library and archives literature for several years. There is also a large and quickly expanding industry associated with digital forensics, which focuses on the discovery, recovery, and validation of information from computer systems that is often not immediately visible to common users. Several authors have recently investigated the use of forensic tools and techniques for acquiring digital collections in libraries and archives. The Prometheus and Presidential Electronic Records Pilot System (PERPOS) projects have developed software for data extraction, focusing on needs of specific collecting contexts. Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) and FutureArch are exploring workflows that include forensic copies from digital media. Another project, Computer Forensics and Born-Digital Content in Cultural Heritage Collections, has also provided a significant contribution to this discussion.

**Personal Information Management (PIM)**

Personal Information Management refers to “both the practice and the study of the activities people perform in order 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).” It “places special emphasis on the organization and maintenance of personal information collections (PICs) in which information items, such as paper documents, electronic documents, email messages, web references, and handwritten notes, are stored for later use and repeated reuse.” After emerging as a distinct area of study in the 1980s, PIM has become a very active area of research in recent years. Several professional associations have been involved in hosting a regular PIM workshop series since 2005. PIM was the topic of a special issue of *Communications of the ACM* in January 2006,
and a piece in the *Annual Review of Information Science and Technology* in 2007.

Despite all of the recent energy and attention to PIM, the literature has been relatively constrained in its focus. Although PIM is intended to be inclusive of all aspects of personal information management, the primary focus of the literature has traditionally been the retrieval of personal information. However, several authors have broken that mold in recent years by examining issues of selection, management, and sharing of information over time. Several of the *I, Digital* authors are contributing to this new stream of research within PIM.

The archival literature about personal papers often emphasizes the relatively unruly ways in which the materials tend to be organized and acquired. However, the story does not end there. The study of personal recordkeeping should not be “a study about chaos but one about the impulses driving individuals and families to create, maintain, and use their own records.” The PIM represents a significant body of research that suggests definite patterns of behavior in the personal creation, management, and use of information.

### Tools and Support for User-Generated Collections

When discussing software to support electronic recordkeeping, the archives and records management literature has often emphasized large, enterprise-wide systems. Two very visible sources of industry guidance for such systems are DOD 5015.02 and the Model Requirements for Management of Electronic Records (MoReq). Such guidance documents—and tools designed to fulfill their requirements—can be valuable for electronic recordkeeping in large institutional environments. However, it is unclear how they could support the management and preservation of and access to materials created by individuals outside such formal structures.

One potential response is to provide tools to individuals that will allow them to engage in better recordkeeping (creation, capture, management and preservation) practices themselves. Richard Cox has argued that “archivists might have a better chance of being successful in the preservation of private, non-organisational records if they worked with software manufacturers to create commercial products that individuals could acquire that enable the long-term maintenance of their electronic files and the more easy transfer
of these personal papers to real or virtual repositories.” Archivists can also contribute to, support, and take advantage of efforts to develop open-source applications and services for use by individuals.

There have been several recent initiatives to develop tools to allow individuals to better manage their personal digital collections. Major software vendors have also entered this arena with products such as Apple’s Time Machine, Norton 360, and Windows Live OneCare, which help individuals to keep their bits safe. A particularly active area of development has been tools and services for personal web “archiving.” ContextMiner and ArchivePress are designed for building collections from the social web, such as YouTube and blogs.

Finally, there are various targeted documents designed to provide guidance for individuals to manage their own digital materials as cultural resources. The Image Permanence Institute has created several such documents. In relation to online personal content, the Electronic Frontier Foundation provides a service called TOSBack to inform consumers of the terms of service related to systems that they use, and Google maintains a site called the Data Liberation Front that provides instructions for getting information into and out of Google services.

What This Volume Is Not About

We have become a nation of snoopers.

—U.S. Representative Cornelius Gallagher, 1966

This volume does not place much emphasis on the vast databases of relatively discrete and structured data about individuals that are being collected, managed, and mined by the private or public sector—i.e., “the systematic capture of everyday events” that is generating “a vast network of stored information about the populace that must rival the infinite account books of Heaven.”

Now that so many human behaviors and interactions are carried out through the generation and exchange of symbolical representations in digital form, individuals leave behind a vast number of micro-traces. According to one estimate, less than half of the “portion of the digital universe created by individuals . . . can be accounted for by user activities—pictures taken, phone calls made, emails sent—while the rest constitutes a digital ‘shadow’—
surveillance photos, web search histories, financial transaction journals, mailing lists, and so on." In other words, “data is the perspiration of the Information Age,” and plenty of organizations are standing by to soak up the sweat. Private sector examples include credit card companies, credit agencies, marketing companies, voter data aggregators, search providers, and online retailers. Public sector examples include data warehouses of tax collecting agencies and intelligence agencies. The massive amount of data related to individuals and their transactions is evidenced by numerous companies that report or are believed to be running web server farms that number in the tens of thousands. The aggregation of the various “shards of data” held by public and private entities about a person can be considered to be his or her “digital dossier.”

While such data have both contemporary importance and significant potential future research value, the data are often outside of the control of either individuals or records professionals, and are generally managed as diffuse units of data, rather than as “personal collections.” Even more importantly, the diffuse digital dossier about a person “fails to capture the texture” of her life. The digital dossier does not usually serve one of the most important purposes of personal archives: giving voice to the individual. Contributions to I, Digital focus on materials that are directly, of (and often purposely shared by or with) individuals, rather than all digital information that is about individuals. For example, rather than investigating the various forms of surveillance of individuals by corporate or government actors, this volume places considerable emphasis on “participatory surveillance.”

Another noteworthy set of activities that are not a direct focus of this volume are those that allow individuals to submit their own stories and items to larger, shared repositories. Prominent examples include the BBC’s WW2 People’s War project and Virginia Tech’s April 16 Memorial website. Such projects hold great promise for representing and preserving individual voices that would be missed by collecting driven exclusively by institutional recordkeeping requirements, much as oral history efforts have done for many years. However, they do not directly address the much wider universe of digital materials being created and collected by individuals, which is the focus of I, Digital.
Summary of Contributions to this Volume

*I, Digital* brings together insights, findings, and perspectives of authors with expertise in various aspects of the management of digital information. The first part of this volume is devoted to conceptual foundations and motivations. The first chapter—the one you are currently reading—explains the scope and motivation for *I, Digital*, as well as setting it within the context of related literature.

Robert Capra and I have jointly authored the next chapter, which is entitled “And Now the Twain Shall Meet: Exploring the Connections between PIM and Archives.” Capra is a Post-Doctoral Fellow, and I am an Associate Professor at the School of Information and Library Science, University of North Carolina, Chapel Hill. In this chapter, we describe how, over the past two decades, two complementary, but relatively unconnected, streams of literature have emerged: electronic recordkeeping (ERK) and personal information management (PIM). The former builds upon long-standing principles and practices in archives and records management (ARM), while the latter has its strongest roots in human-computer interaction and information science. A fundamental issue that cuts across both areas of literature is how to make meaningful use of information artifacts when one is no longer embedded in the original context of creation, capture, or use. The ARM literature formulates this as secondary use, while the PIM literature focuses on refinding and reuse. In this chapter, we identify and elaborate on areas of commonality and difference in these two areas of study. We conclude with several open research questions and argue that there is great potential for research that is attentive to the complex set of factors that PIM and ARM can collectively bring to these questions.

In “Ghosts in the Machine: Towards a Principles-Based Approach to Making and Keeping Digital Records,” Adrian Cunningham, Director of Strategic Relations at the National Archives of Australia, argues that “it is time to take a step back from discussing time-bound, technology-specific practical strategies and technical solutions and to focus instead on a more enduring, principles-based approach to dealing with the complex realities of digital recordkeeping.” He proposes twelve principles, which are adapted from a set of principles developed by the International Council on Archives for institutional recordkeeping. In the process of articulating and justifying...
the relevance of the twelve principles, Cunningham makes a strong case for collaboration and mutual learning among those responsible for personal records and organizational records.

Cathy Marshall, who is a Senior Researcher at Microsoft Corporation and an Affiliated Researcher at Texas A&M University, explores “Challenges and Opportunities for Personal Digital Archiving.” She begins with a discussion of two 1995 quotations, one from John Seabrook and the other from Jeff Rothenberg. She contrasts the visions from those authors with the state of contemporary personal collections, which are likely to be composed of “thousands and thousands of pictures, hours and hours of undifferentiated digital video footage, music both purchased and shared, personal finances, the beginnings of what will one day be extensive medical records, email messages important and trivial, and countless other files representing day-to-day interactions with the computer and with other people by means of the computer.” Personal digital collections are not only diverse but also “apt to span many repositories, storage media, and file systems.” Management of personal digital archives over time “is a problem that extends far beyond accurately rendering obsolete formats.” The rest of Marshall’s discussion is organized around three questions: What’s in a personal digital collection (as a matter of selection and appraisal)? What is “the technological basis” for individuals’ storage, maintenance and preservation of their personal digital collections? What are individuals’ stewardship practices and what practices should they be taught? The result is a sober assessment of strategies and options for stewardship of personal materials.

Sue McKemmish is a Professor at Monash University’s Faculty of Information Technology. In 1996, she published a very reflective and influential piece called “Evidence of Me . . . ” in which she explored the nature of personal recordkeeping and social mandates for “witnessing to individual lives.” In her chapter for this volume, entitled “Evidence of Me . . . ” McKemmish revisits her earlier piece within the context of new technologies, record forms, and practices. Building on Derrida’s claim that “mutation in technology changes … the possibility of archiving,” she explores the relationship between personal and public recordkeeping, as well as the place of records within a wide array of potential “memory stores” about individuals. Along the way, McKemmish investigates and challenges many conventional archival assumptions about personal archives.
The second part of *I, Digital* is devoted to particular types, genres, and forms of personal traces; areas of further study; and new opportunities for appraisal and collection. “‘Three Backups Is a Minimum’: A First Look at Norms and Practices in the Digital Photo Collections of Serious Photographers” was authored by Kristina Spurgin, who is a Doctoral Candidate at the School of Information and Library Science, University of North Carolina, Chapel Hill. Spurgin argues that amateur photographers fall at the boundary between three different arenas that have been represented in previous literature: “a) institutions and organizations; b) casual snapshotters collecting mainly family photos; and c) unknown individuals sharing, tagging, and annotating photo collections on the Web.” Spurgin investigates the “norms and expectations of the social world of photography” and implications for the practices of amateur photographers. She reports on preliminary analysis of a year of discussion threads from an online forum devoted to discussing collection management and digital photography workflow. She asks whether “there is anything new and important to learn from studying the personal collection management of amateur photographers,” and the answers are very compelling. She places this investigation within a larger context of personal collections related to activities pursued as “serious leisure,” and argues that “photographs alone do not comprise the photography-related collections of amateur photographers.”

In “Collecting the Externalized Me: Appraisal of Materials in the Social Web” I argue that, with the adoption of highly interactive web technologies (frequently labeled “Web 2.0”), forms of individual documentation and expression are often inherently social and public. Such online environments do allow for personal documentation, but they also engage external audiences in ways not previously possible. This opens up new opportunities and challenges for collecting personal materials, particularly within the context of archival appraisal. The chapter explores the nature and characteristics of user-generated content on the Web and various ways in which principles of archival appraisal can be operationalized in an environment in which collecting takes the form of submitting queries and following links. As with any other case of archival appraisal, the fundamental questions relate to what one is attempting to document and why.

The final section of *I, Digital* addresses the practical implications of the issues raised in the previous chapters for the strategies and practices of
professionals who work in memory institutions. Tom Hyry and Rachel Onuf published a piece in 1997 called “The Personality of Electronic Records,” in which they made a strong case for widening the professional discourse on electronic records to be more inclusive of concerns related to personal materials. In this volume, Hyry, who is Director of Special Collections at the University of California, Los Angeles (UCLA), and Onuf, who is Archives Analyst for the Archivists’ Toolkit project, have again joined forces in “Take it Personally: The Implications of Personal Records in Electronic Form.” They explore the major opportunity of digital recordkeeping to reflect “digital contexts” and for “future social historians and others invested in documenting regular folk.” They argue that the archival profession “has come full circle, back to the idea that the records we need to preserve are rooted in larger systems—but now in online networks, rather than organizational recordkeeping systems—and archivists need to focus on holistic capture of this data in order to meaningfully document the individual.” They conclude by elaborating several recommendations: explicitly collect the digital; develop enhanced access systems; and embrace and acquire new skills.

In the chapter by Leslie Johnston, Digital Media Projects Coordinator at the U.S. Library of Congress, titled “Making it Usable: Developing Personal Collection Tools for Digital Collections,” she reports on her earlier work at the University of Virginia (UVa) Library. Johnston explains that a major assumption of the UVa Library “while developing its Digital Collections Repository was that the library was responsible not only for collection and repository building, but also for the creation of tools for collection use.” She reports on the development of a User Collection Tool to organize personal digital media collections, a PageComber tool to gather images from the open web, and the Collectus digital object collector tool to develop portfolios of all object types and create presentations. This work, and Johnston’s articulation of the development process, provides many lessons for other memory institutions that hope to actively engage with record creators, and better integrate digital library systems with the environments being used by individuals in support of their scholarly activities.

In “Curating the I, Digital,” Susan Thomas, Digital Archivist at Oxford University’s Bodleian Library, examines how the Bodleian Library has been adapting and evolving to address contemporary personal archives. She offers an overview of the Bodleian’s activities, both in terms of everyday practices
and two very influential projects that I discussed earlier: Paradigm and futureArch. The efforts of Thomas and her colleagues serve as a compelling and informative model for memory institutions across the globe that are undergoing the transition toward incorporating born-digital materials into their regular operations and practices.

When it comes to personal collections, we live in exciting times. Individuals are living their lives in ways that are increasingly mediated by digital technologies. While this mediation presents many technical challenges for long-term preservation, it also provides unprecedented opportunities for documenting the lives of individuals. There is a growing community of practice related to the acquisition and management of personal digital collections, with many of the participants being archivists, special collections librarians and manuscript curators. The PIM research community also continues to engage in valuable research, with many signs that they can serve as valuable collaborators with archival researchers. I believe that *I, Digital* can serve as a catalyst in a process of growing attention to and progress with personal digital collections, which is already underway. I look forward to great things ahead.
Notes

6 Abbott, System of Professions.
14 Elizabeth H. Dow, Electronic Records in the Manuscript Repository (Lanham, MD: Scarecrow Press, 2009), xii.


See http://www.ils.unc.edu/digitaldesktop/.

See http://www.records.ncdcr.gov/emailpreservation/.

Abbott, *System of Professions*.


For example, see Diigo (http://www.diigo.com), Find it! Keep it! (http://www.ansemond.com) and Spurl (http://www.spurl.net), as well as several Firefox add-ons and archiving features built into many of the major blog hosting services.


See http://archivepress.ulcc.ac.uk.


59 Solove, *Digital Person*.


61 Solove, *Digital Person*, 49.


Contributors

Robert Capra is a Post-Doctoral Fellow and adjunct faculty member in the School of Information and Library Science at the University of North Carolina at Chapel Hill. His research interests include human-computer interaction, personal information management, and digital information seeking behaviors, tools, and interfaces. He holds a PhD in computer science from Virginia Tech and MS and BS degrees in computer science from Washington University in St. Louis. At Virginia Tech, he was part of the Center for Human-Computer Interaction where he investigated multiplatform interfaces, information re-finding, and digital library interfaces. Prior to Virginia Tech, he worked in corporate research and development, spending five years in the Speech and Language Technologies group at SBC Communications, focusing on voice user interfaces, speech recognition, and natural language processing.

Adrian Cunningham has worked at the National Archives of Australia since 1998, where he is currently Director of Strategic Relations and Personal Records. Adrian was the Secretary of the International Council on Archives (ICA) Committee on Descriptive Standards (2002–2004), and is the Treasurer of the ICA Pacific Regional Branch, the convenor of the Australian Society of Archivists Descriptive Standards Committee, the Chair of the AGLS Metadata Working Group, and a member of Standards Australia’s Committee IT/21, Records Management. Adrian previously worked at the Office for Government Information Technology and as a Private Records Archivist/Librarian at the National Library of Australia, the Pacific Manuscripts Bureau, and the State Library of New South Wales. Adrian was the President of the Australian Society of Archivists from 1998 to 2000, and was named a Fellow of that Society in 2007.
Tom Hyry is the Director of Library Special Collections at the University of California, Los Angeles. From 2006 to 2010, Hyry was Head of the Manuscript Unit in the Beinecke Rare Book and Manuscript Library at Yale University. Prior to that, he spent nine years in Yale University Library’s Manuscripts and Archives Department, where he began as an entry-level archivist and finished as the Head of arrangement and description. Hyry earned a master’s degree in information and library studies from the University of Michigan and a bachelor’s degree in history from Carleton College.

Leslie Johnston is the Manager of Technical Architecture Initiatives for the National Digital Information Infrastructure and Preservation Program at the Library of Congress. She was previously Digital Media Project Coordinator at the Library of Congress. Prior to that, she served as Database Specialist at the Getty Research Institute, Documentation Coordinator for the Historic New Orleans Collection, Academic Technology Specialist for Art and the Cantor Arts Center at the Stanford University Libraries, Head of Instructional Technology and Library Systems at the Harvard Design School, and Head of Digital Access Services at the University of Virginia Library. She founded the online journal eSpectra, edited the print journal Spectra, and served on the board of the Museum Computer Network. She holds a BA in anthropology and an MA in archaeology from University of California, Los Angeles.

Christopher A. (Cal) Lee is an Associate Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill. He teaches courses on archival administration; records management; digital curation; understanding information technology for managing digital collections; and the construction of policies and rules for digital repositories. He is one of the lead organizers and instructors for the DigCCurr Professional Institute, which is a week-long continuing education workshop on digital curation, and teaches professional workshops on the application of digital forensics methods and principles to digital acquisitions. His primary area of research is the long-term curation of digital collections. He is particularly interested in the professionalization of this work and the diffusion of existing tools and methods (e.g., digital forensics, web archiving, automated implementation of policies) into professional practice.
Sue McKemmish holds the Chair of Archival Systems at Monash University, and is Director of the Centre for Organisational and Social Informatics. She is engaged in major research projects relating to recordkeeping metadata, information portals, Indigenous archives, and the nexus between memories, communities, and technologies. She directs the postgraduate teaching programs in archives and records at Monash University, and has published extensively on recordkeeping in society, records continuum theory, recordkeeping metadata, and archival systems.

Cathy Marshall is a Principal Researcher at Microsoft Research’s Silicon Valley laboratory after a six-year stint in Microsoft’s product divisions. Her research interests include personal information management, the preservation of and long-term access to personal and scholarly digital belongings, and interaction with digital media. She was a member of the research staff at Xerox PARC and is an affiliate of the Center for the Study of Digital Libraries at Texas A&M University. Marshall has delivered keynote addresses at numerous international venues. She recently completed a book called *Reading and Writing the Electronic Book*, and has published more than eighty peer-reviewed papers. She also has been awarded nine patents, and has nine other patents pending.

Rachel Onuf teaches Preservation Management for Simmons College Graduate School of Library and Information Science, and works as a consultant for a range of collecting repositories. Much of this work involves preservation planning and collection assessment, and is grounded in her experience in various research libraries, including the Historical Society of Pennsylvania. She earned her MILS with a concentration in archives from the University of Michigan, and an MA from the University of Virginia, with a focus on the nineteenth-century South. She lives in western Massachusetts, where she also works with cows and teenagers on a school farm.

Kristina M. Spurgin is a Doctoral Candidate at the School of Information and Library Science, University of North Carolina at Chapel Hill. Her college photography professor said she could likely make a career out of photography, but would have to shoot a lot of chicken feed bags to make a living. She has never done professional photography. She studies how people make sense of personal information management, catalogs electronic resources, and teaches.
Susan Thomas is Digital Archivist at the Bodleian Library at the University of Oxford, where she is responsible for developing the library’s capacity to curate, preserve, and provide access to born-digital archives. She is also an Honorary Lecturer at the University of Dundee’s Centre for Archive and Information Studies, where she teaches the management and preservation of digital records for the MLitt and MSc programs. Thomas holds a BA in English language and literature from the University of Manchester and an MA in archives and records management from the University of Liverpool.
The letters f or t following a page number indicate a figure or table.

A
abandonment of old files as personal archiving strategy, 98
access policies to archives
and accessibility of electronic records, 245–246
and creation of personal portfolios, 265
in User Collection Tool (UCT), 261
access systems. See browsing; searching
accessibility of electronic records
long-term uncertainty of commercial enterprises for, 132, 134, 215, 251
and maintenance of records, 84–85
personal records, 93
and services to researchers, 245
accessioning procedures for digital files, 289
accidental loss as strategy for deleting files, 102, 112
accountability, 56, 126
action information in PIM research, 38–39
amateur enthusiasts
difficulty of professionals responding to, 4–5
in photography, 152, 154–157
ambiguity in classification of information, 59–60
annotation tools, 267, 300
annotations to official records, 141
appraisal, 54, 134, 142, 216–218
arbitrariness of digital records
organization, 101
archives and records management (ARM)
core PIM concepts in, 35–37
terminology of, 2
archivists
failure to incorporate born-digital into archives, 244
need for technical skills in, 244–245, 253–254
See also precustodial relationships with content creators
arrangement and description
in archives, 43
interfaces for, 298–299
multiple points of view, 123–124
user-supplied, 248
audience
for blogs, 208–209
for different profiles of person, 223–224
for social network sites, 210
See also primary and secondary value of records
Australia’s Stolen Generation of Indigenous children, 124, 137–141
authentication of digital records
and access for researchers, 299
in appraisal process, 295
and personal collections, 271
as research skill, 247
authoritative text in shared set of values, 175
backup systems
ambiguity between duplication of current files and long-term storage, 174–175, 188
failure to use, 102, 104, 112, 167, 284 by serious photographers, 173, 187–190, 193
BEAM. See Bodleian Electronic Archives and Manuscripts (BEAM) repository
belongingness among photographers, 152, 155, 166
benign neglect as weeding approach, 107, 110, 111–112
best practices of photographers compared to digital preservation professionals, 175–176, 194
bias in collecting interests and underrepresented populations, 205–206
blogs and bloggers
comments on, 215, 250
most popular, 211, 212–213t
as online personal archive, 130, 133
as personal journals, 208, 213, 244
as user-generated content, 207
Bodleian Electronic Archives and Manuscripts (BEAM) repository, 282 and accessions of digital files, 289–294
multiple kinds of media, 296
boundaries between personal and work life increasingly blurred, 81
records having both personal and organizational characteristics, 46–47, 142
and third party rights, 293
browsing
as finding/refinding operation, 37
by home mode photographers, 167, 179
interface for in User Collection Tool, 262
See also searching
bureaucracy, growth of, 45
businesses, documentation of web activities of, 206
C
Cairo Project, 282, 286, 302n6
camera clubs in amateur photography, 155
capture of metadata, 86, 87, 282
capture of records in accessioning workflow, 290 in cloud computing, 293, 294 and context, 250–251 in continuum theory, 120 for digital records, 130
categorization of information, See classification
celebrities, 223
chain of custody and electronic records, 34–35
classification, 35, 58–59
classification taxonomies, 86
closure periods, 287
cloud computing
capture techniques for, 293, 294 control of information in, 6 risks with, 284, 289
See also commercial service providers
collaborative search, 43, 48
collabularies, 250, 255n24
collecting hobbies, 154
collection development, proactive, 193, 282, 286–289
collection management in photography, 163–164, 173, 194
collection of web content, approaches to, 225–228, 226t
collective memory, 48, 135
Collectus, 263–269, 264f, 265f, 266f, 273f
collaborative search, 43, 48
collabularies, 250, 255n24
collecting hobbies, 154
collection development, proactive, 193, 282, 286–289
collection management in photography, 163–164, 173, 194
collection of web content, approaches to, 225–228, 226t
collective memory, 48, 135
Collectus, 263–269, 264f, 265f, 266f, 273f
commercial service providers
difficulties in collecting from, 285 long-term uncertainty of, 132, 134, 215, 251 and privacy issues, 47–48
See also cloud computing
web storage devices, 95, 189
See also cloud computing
communities of records, 133, 147n69
confidentiality. See privacy and confidentiality
c content creators. See records creators

c context of records
in ARM and PIM, 49–51, 62
captured with personal records, 250–251
in continuum theory, 120, 121
in discussions with records creators, 285
and evidentiality of records, 127
as index to memory, 101
loss of in use of feeds, 209
and messiness of classification, 60
and original order, 34
and privacy of individuals, 231
and quality metadata, 84
and web collecting, 228–229
continuing value, concept of, 40, 49,
55–56, 284
See also value of archives
continuum theory
and evidential role of information, 127, 130
overview, 11, 118, 119–122, 119f,
145n22
principles of, 82, 89n12
and social web, 229–230
copies of digital objects, 51, 168
See also duplicates
copies of paper objects, 245
copyright. See intellectual property and
copyright
correspondence, web-based, 207–209,
243, 251
See also letter-writing, role in
personal identity
costs of maintaining digital archives, 282, 301
costs of maintaining paper collections, 248
costs of storage, 6, 51, 55, 97, 244
course management systems, 260
crawl frequency in web collecting, 227
creation of records. See records creation
process
creators of records. See records creators
crowd sourcing
for organizing archives, 110, 113n28
overemphasizing uncommon but
spectacular events, 217–218
as web collecting strategies, 231
currency of knowledge, 191–192
current information in PIM research, 38–39
custody of distributed records across
multiple platforms, 82

D
data gathering by commercial services, 132
data loss
due to degradation of media, 93, 281
due to failure to backup, 102, 104,
112, 167, 284
due to losing track of location of
items, 106
and fragility of content on web
services, 215
data mining, 14–15, 228, 247, 253
data safety as advantage of distributed
storage, 104–105
data synchronization services, 189–190
dataspace, concept of, 65
date in organization of photographs, 177
degradation of media, 93, 281
deletion of records
difficulty of, 5–6
and email storage quotas, 284
Paradigm guidelines for, 182–184
as strategy for organization, 36, 57
as unsystematic, arbitrary process,
99–100
See also weeding of records
demographics of user-generated
content, 218–221, 224–225
depositor, definition, 302n10
See also precustodial relationships
with content creators
description. See arrangement and
description
deselection of materials. See weeding of
records
destruction of records. See weeding of
records
digital archaeology as research skill, 247,
291–292
Digital Asset Management System (DAMS), 297
digital cameras as recordkeeping devices, 130
digital data recovery and forensics, 11–12
digital diplomacy as research skill, 247
digital divide, 251–252
Digital Images Archiving Study, 182
Digital Lives project, 8, 80, 252, 275
digital negatives. See RAW files (photographic file format)
digital preservation
in accessioning process, 290–294
and intellectual property issues, 272
issues in, 5, 276
long-term uncertainty of commercial enterprises for, 132, 134, 215, 251
photo collection guidelines, 168
in PIM, 40, 64, 92–93
policies for, 297–298
and reuse of information, 53–54
See also long-term preservation
digital rights management technology, 84–85, 95
digitization of holdings, 246
curation of, 297
directory structures
assistance from information professionals, 194
for organizing photographs, 177
disempowered populations, traces of on web, 205–206
disk crashes as method of pruning collections, 102–103
distributed repositories, 82, 106–107, 222
DNG image format, 182, 185–186
donors. See precustodial relationships with content creators
duplicates, 295, 299
See also copies of digital objects
See also personal electronic recordkeeping (PERK)
email, organization of, 57, 60, 284
eManuscripts, 2
empowerment of others in shared archival spaces, 133
emulation, 94–95, 292
encrypted records, 85, 190
environments crawled in web collecting, 226
ephemeral information, 39
ethical issues
data mining, 228
expunging materials from archive, 296
sensitive materials, 286
third party rights, 294
evaluation criteria for photographic tools, 163, 191
evidence of personal activity, records as, 83–84, 126–129
EXchangeable Image File (Exif), 180
expected use and choice of preservation method, 95
Extensible Metadata Platform (XMP), 180
extraction techniques in transfer of files, 285–286
F
family information practices, 48
fandom, management of information in, 154
fears of loss, 166, 167, 188
feeds and loss of contextual information, 209
file formats
for long-term preservation, 184–187, 191, 276
multiplicity of, 79, 86, 181–182, 227
file naming conventions, 86, 177–178, 298
filing and piling strategies in PIM theory, 35, 37
filter blogs, 208
filtering of duplicated data, 295
finding aids for hybrid archives, 298
finding and refinding
and context of information, 49–51
keyword searching, 37, 59
location-based finding, 50, 59
in photography, 180
as PIM activity, 30, 36, 166
variable methods for, 37
See also searching
Flexible and Extensible Digital Object
and Repository Architecture (Fedora), 263
Flickr. See photo-sharing services
fonds as organizing principle, 6, 33, 65
Forensic ToolKit (FTK), 292, 295–296
forensics, digital, 12, 292, 295–296,
304n27
forgetting and remembering, 125, 183
formats. See file formats
fragmentation of inter-blog
conversation, 207
full-text search as organization strategy,
57
futureArch project, 282, 290
fuzziness and ambiguity in classification
of information, 59–60

G
genealogy, management of information
in, 154
geographic representation of blogs, 211
ghost Twitters, 294, 304n31
granularity and provenance of levels of a
record, 33
group information management (GIM), 48

H
hard drives, external, as backup media,
188–189
hardware
obsolescent, 11–12, 191, 290–292
for photographic files, 169–170
reliability of, 190
hashing, 295
home life, changes in, 45–46
See also work/life interactions
home mode photography, 154
browsing by chronology, 179
deferred plans to label photos, 179
file formats for long-term
preservation, 184–187, 191
file naming practices, 177
photo sharing practices of, 156, 167
strategies on organizing from serious
amateurs, 158
unsystematic backing up, 188
See also photography
human-computer interaction, 30, 35
hybridity of digital and print records,
281, 284
I
identity construction in personal
collections organization, 166
identity management, 224–225
See also profiles, online, uses of
incentives for good personal digital
curation, 62
inclusion and exclusion in archives, 125,
142
incommensurate ontologies, 124–125,
142
Indigenous ways of knowing, 123–124,
126, 129, 142
individuality and group membership,
118–119
information creation and sharing as
serious leisure, 154
information fragmentation, 49
information overload, 36, 55–56, 93
informed consent and web collecting,
228, 229
institutional and organizational records
archives as, 242
blended with personal records, 81,
128–129
vs. manuscript tradition, 2
See also public archives traditio
institutional or corporate memory as
focus of ARM, 48
intellectual property and copyright
and collection management tools, 272
and long-term preservation, 255n26
in Paradigm project, 191
and personal archives, 95, 190
and selection decisions, 55–56
intentionality of creation of traces, 30
interactive recordkeeping and digital
technologies, 131, 132–133, 142
interfaces for discovery tools, 259, 262, 298–299
International Color Consortium (ICC), 180
International Press Telecommunications Council (IPTC), 180
interoperability across platforms over time, 86
interpretation of records and selection/retention decisions, 56

J JPEG file format, 184, 187
jurisdiction of a profession, 4–5, 10

K “keep everything” collection policy
as common de facto strategy, 97–98
and computer as memory prosthesis device, 101
as keeping operation, 36, 37
in PIM, 53, 183
Keeping Found Things Found (KFTF) studies, 275
keyword searching as finding technique, 37, 59
keywording, 179
See also tagging
knowledge acquisition as benefit of serious leisure in photography, 155, 156
knowledge management and PIM concepts, 35
Koorie Archiving System (KAS), Australia, 137–141

L legacy, establishment of, 166
legacy digital files, 192, 289
letter-writing, role in personal identity, 116
See also correspondence, web-based
levels of access and user-supplied description, 248
levels of activity in social media, 211, 218–221
librarians, scope and nature of work of, 4
licensed digital content, 191, 290–291, 296–297
lifelong learners, strategies on organizing from, 158
links in web archives, 207, 218, 225
location-based finding, 50, 59
long-term management of records and expectation of individual bloggers, 207–208
and social memory, 83–84
strategies for entire continuum, 81
time scale in PIM and ARM, 37–40
uncertainty of commercial services, 132, 134, 215, 251
long-term preservation
copyright enforcement and, 255n26
for photographic file formats, 184–187, 191
privacy of personal information and, 255n26
problems of long-term care for records, 5
vs. current optimization, 61
See also digital preservation
long-term value. See continuing value, concept of
lurking on online spaces, 166, 214

M Mandela Archive, 135–136
manuscripts tradition, 2, 7–9, 242
See also public archives tradition
media degradation, 93, 281
See also migration of records
memory, collective, 116, 119, 125
memory prosthesis device, computer as, 101
metadata
for archival repositories, 65
assistance from information professionals, 194
capture of, 86, 87, 282
collection of in record surveys, 284
and context of records, 84
and copyright, 190
desired characteristics, 267
and enhanced access, 246, 253
and federation of records in distributed repositories, 106, 111
loss of in migration of data, 180–181, 291–292
Index

for photographic collections, 180, 181–182, 193
in PIM, 31
for records from online services, 294
self-documenting, 179–182
use of by scholars, 262, 266–267
Metadata Working Group, 182
meta-frameworks for integrated
personal and public recordkeeping, 142
migration of records
and degradation of media, 93, 281
and format transformation, 276
and incompatibility with older
equipment, 191
interoperability across platforms over
time, 86
and management of electronic
records, 84
and photographic metadata, 180–181
use of open formats for, 86
See also transfer of files
mobile devices, as recordkeeping
deVICES, 130, 284
motivations for PIM, 62
multiple provenance. See provenance
multitasking, effect on workflow, 59–60
museum curators, 4

N
naming conventions for files, 86, 177–178, 298
networked information and work/life
interactions, 46
noise and “keep everything” approach, 53
norms and best practices in shared set
of values, 175–176, 194

O
offsite storage of backups, 189
Omeka tool for building collections, 277
online commercial services. See
commercial service providers
online image viewing tools, 262
online profiles, uses of, 223–225, 229
open source environments and
recordkeeping behavior, 130
OpenRAW Working Group, 195
optical media
as backup devices, 188–189
forensic imagers for, 292
oral records, 117, 123, 139
organization and management activities
as enjoyable, 156, 173
for photographic collection
management, 163
organization of records
in continuum theory, 120
as low-priority activity, 58
strategies used for, 56, 177
organizational records. See institutional
and organizational records
orienteering process in PIM, 50
original order, concept of, 33–34, 60, 292
outreach to community by institutional
archives, 158
ownership of personal information, 32, 37, 48–49

P
Pachyderm tool for slide show
development, 274
PageComber, 271–272, 272f
page-turner tools, 263
paper records, collection of, 84, 244–245, 248–249
Paradigm project (Personal Archives
Accessible in Digital Media)
backup, 187–190
consultation with professionals, 192–195
formats, choice of, 184–187
guidelines for creators of personal
archives, 168–169, 176–184
as model, 252, 281–282
overview, 8, 20, 281–283
and photography, 171, 176, 190–192
parallel provenance. See provenance
participation patterns in online spaces,
166, 214–215
password protection for photographic
collections, 190
peripheral devices, 94, 284
persistence of content, concerns about, 62
personae, identification of, 227–228, 287, 294, 303n16
personal archives
  as collective memory, 135
definitions, 2–3
effect of digital media on, 249
fragmentation of information across a
diversity of locations, 204–205
See also manuscripts tradition
personal archiving services, 110
personal collection management. See
Personal Information Management (PIM)
personal electronic recordkeeping
(PERK), 18–19, 63
See also personal recordkeeping
personal expression and disclosure, 208,
244
personal information collection (PIC),
definitions, 12, 36–37, 58, 65
Personal Information Management
(PIM)
  contents of typical collection, 96–104
  contextual information in, 49–51
core concepts of, 31–37
definitions, 2, 3, 258
as difficult, drab, and undesirable,
153–154
and electronic recordkeeping, 16
as enjoyable, 156
intersections and differences with
ARM, 38–39t
organization of information, 57–60
ownership of materials, 45–49
photographers’ use of, 157
review of literature, 12–13
selection principles, 17, 51–56
system requirements, 60–61
time scale, 38–40
user tasks to support, 41–44
personal journals, blogs as, 208, 213,
244
“personal papers” as alternative term for
manuscripts, 2
personal recordkeeping
definitions, 2–3
relation to public recordkeeping,
121–122, 123, 125
role in witnessing and memorializing
individual lives, 116, 129, 166
See also personal electronic
recordkeeping (PERK)
personal space of information (PSI), 36,
37
personal storage devices (USB drives),
95
personal stories as source of research
data, 6–7
personal work files, 38–39
photography
  as amateur pursuit, 154–157
  collection management practices,
157–159
difficulties of intellectual access to,
157
ethos of, 175–176
guidelines for long-term preservation,
168, 184–187, 191
personal collections in, 166–167,
171–173
Photo.net forums, 160–161, 164–166
as serious leisure, 151–154
technical skill required for, 173–174
template for study of, 163–164
workflow in, 170–171
See also home mode photography
Photo.net, 160–161, 162t, 164–166
photo-sharing services
  in photography, 155–156, 157, 159,
166, 180, 184, 189, 190
privacy of, 250
usage of, 91, 95, 105
  as user-generated, 206
physical carriers of electronic data,
preservation of. See hardware
“physical” elements of hybrid archives,
288
“piling” as organization strategy, 37, 57,
59
PIM. See Personal Information
Management (PIM)
Pittsburgh Project, 242
Planets project, 298
pluralization of records, 121, 122, 130
points of view, multiple
  in archival description, 123–124
  in continuum theory, 121
See also worldviews, incompatible,
collection of
popularity as selection criterion, 217–218
portability as need in PIM, 62
portfolios, creation of, 263–264
postcustodial approach to management, 229–230, 237n109, 242
postmodern thinking on archives, 123–124, 123–126, 129, 142
power law distribution of web participation, 210–213
power relationships in traditional archives, 125
precustodial relationships with content creators
contacts with, 158, 242, 251, 285
maintenance of relationship over time, 287, 303n18
and unexpected data in digital files, 296
preservation. See digital preservation
preservation of current state of information, 62
preservation professionals as source for serious photographers, 192–195
primary and secondary value of records in ARM literature, 30, 43–44, 44t
difficulty in determining, 51
and manuscripts tradition, 242
Principles and Functional Requirements for Records in Electronic Office Environments, 83
privacy and confidentiality
and context of information, 231
and data mining, 228
decoupling from personal, 249–250
and long-term preservation, 255n26
maintaining in donations of archives, 287
management of, 37
in Paradigm project, 191
and personal archives, 95
and selection decisions, 55–56
and serious photographers, 190
and services to researchers, 253
for third parties and interactivity of online spaces, 293
and users of social networks, 225
processing of paper collections, 248
professionalization of personal collection management, 4
profiles, online, uses of, 223–225, 229
See also identity management; personae, identification of provenance
multiple provenance, 121, 123–124, 134, 139, 142
parallel provenance, 123–125, 126–127, 131
in PIM, 31–33, 50
and web collecting, 228–229
PSD files (Photoshop file format), 184
public and private roles of individuals, 118–119
public archives tradition
and advent of digital records, 242
social memory in, 56
viewpoint of government privileged in, 124
vs. personal archives, 128–129
See also institutional and organizational records; manuscripts tradition
published statements as guide to selection of records, 217–218
purging of records. See deletion of records
Q
query submission in building collections, 225
R
RAW files (photographic file format)
and legacy digital files, 192
long-term preservation of, 184–187, 195
multiplicity of file formats for, 181–182
size and complexity of files, 168
workflows for managing, 170–171, 172f
reading rooms for use of hybrid archives, 253, 300
record naming conventions, 86, 177–178, 298
record surveys, 284–285
recordkeeping systems
definitions, 117
embedded into workflow, need for, 81
as integral to conduct of one’s personal activities, 85
need to be unobtrusive, 61
records
in continuum theory, 120
definitions, 3
as dynamic objects, 122
in PIM, 31
in postmodern archival thinking, 126–127
relation to other forms of recorded information, 129
records continuum. See continuum theory
records creation process, 81, 120, 130, 242, 244
records creators
in ARM, 32
co-creatorship between individuals and organizations, 131, 139
and difficulty of harvesting from web, 222
intentionality of records creation, 85, 89n18
intermingling of personal and public personae, 287
interventions by archivists, 252–253
involvement in record creation process, 81
multiple creators, 123–124, 133
and screen names, 293, 304n31
See also precustodial relationships with content creators
records management as profession, 55, 118, 133–134
recovery of data, 12
redaction techniques for confidentiality, 287
redundancy of backup media, 189
refinding. See finding and refinding
reliability of backup media, 189
remembering and forgetting, 125, 183
reminders, 62
repositories for personal archives
central repositories, advantages of, 104–105, 106
multiplicity of, 91, 93
representation as goal of collecting, 213, 216–218, 219
research methodology, new methods of, 247–248, 258
researchers. See scholarship and research
retention scheduling, 54, 63
retrieval methods, variability of, 37
reuse of personal data by commercial services, 47–48, 132
role-playing games, management of information in, 154
S
sampling techniques in web collecting, 230–231
Scholar’s Box project, 276–277
“scholar’s workbench” tool, 274–275
scholarship and research
access to materials for, 299–300
lack of digital records for, 281
new methodologies for, 247–248, 258
privacy issues in, 253
tools for, 247, 274–275, 277, 300
search results, views of, 262
searching
as finding/refinding operation, 37
and improvements in tagging, 250, 255n24
in photographic collections, 167
tools for, 53, 253
User Collection Tool, 262
See also finding and refinding
secondary value of records. See primary and secondary value of records; value of archives
security of records
in Digital Asset Management System (DAMS), 297
in distributed repositories, 104–105, 106
and long-term storage of encrypted records, 85, 87
in transfer of archives, 286
selection policies
and goals of collecting, 213
in hybrid archives, 282
individuals’ criteria for vs. collectors, 203, 203t
of photos for further editing, 184
in PIM and ARM, 51–56
popularity as, 217–218
and size and complexity of collection, 258
self-actualization as result of serious leisure, 152
self-assessment, 152, 167
self-disclosure, 225
self-documenting data, 179–182
self-enrichment, 152
self-expression, 152, 156, 249
Semantic Web technologies, 158
sensitive information, 142, 286
See also privacy and confidentiality; redaction techniques for confidentiality
sentimental value of bad photos, 183
shared archive space, 138, 141
shared community values among photographers, 155, 161, 175–176
shared control of recordkeeping, 134
shared workspaces and parallel provenance, 131
sharing of information in PIM and ARM literature, 42–43, 62
in social networks, 206–207, 210
See also photo-sharing services
shopping cart feature for building persistent sets, 263
sidecar files for photographic metadata, 182
simultaneous multiple provenance. See provenance
size and complexity
difficulty of appraisal of collections, 222
of photographic collections, 157, 168–170
scaling of selection process, 258
slide shows, creation of, 265–266, 268, 274
Snaprs (solo amateur photographers), 156, 173, 174, 190
social capital, knowledge and expertise as, 162
social groups formed by photographers, 152, 156, 159, 166–167, 194
social justice issues, 126
social media, 202–231
becoming personal public archives, 131
and communal organizing, 110, 113n28
demographics of users, 220–221t
as networked publics, 210
online communities for maintaining personal networks, 204
storytelling in, 132
See also user-generated content on web
social memory as selection criterion for public records, 56
software
compatible file formats in photography, 184–187
obsolescent, 290–292
for user-generated collections, 13–14
software licenses, management of, 191
sorting and scanning as retrieval methods, 37
sorting options in User Collection Tool, 262
spring cleaning as model for document retention, 57, 101
standards, need for, 86, 134, 161, 180
stewardship of individuals for their own records, 97, 100–101, 102–103, 112
as task of ARM, 48–49
storage, web-based. See cloud computing; commercial service providers
storage strategies
cost of, 6, 51, 55, 97, 244
and development of ARM profession, 54
multiplicity of, 6, 105–106
offsite storage, 189
personal storage devices (USB drives), 95
piling as, 37, 57, 59
in PIM research, 38–39
recovery of items from, 110–111
storytelling, 117–118, 129
subjectivity as topic of research, 6
surveillance using digital technologies, 132
tagging
by archives users, 215, 248, 253
and classification of email, 60
improvements in searchability, 250, 255n24
as organization strategy, 57
by serious photographers, 179–180
in social network spaces, 209
tools for researchers, 300
technology
effect on personal recordkeeping, 17, 65–66, 243
effect on reuse of information, 53–54, 64
and long-term management of records, 5, 84
and long-term storage for digital collections, 40
need for neutrality of, 86
obsolescence in, 102–103, 104, 281
and reconceptualization of archival endeavors, 82–83
template analysis, 160, 163–164, 199n50

The Digital Asset Management (DAM) Book, 168, 169, 175, 192
TIFF file format, 184, 187, 195
“tombstone” metadata, 267
tools and support for user-generated collections
desired characteristics, 266–268
implications for institutions, 19–20
need to develop, 13–14, 62–63, 246–248, 276

traces
as clues to provenance in PIM literature, 50
definitions, 2
intentionality of creation of, 30
and “keep everything” approach, 53
on social web, 203, 222
transactionality in continuum theory, 120, 127
transfer of files
in donation of archives, 285–286
to new computer, 98–99
See also migration of records
trusted digital repository services, 134, 142

Twitter
archived by Library of Congress, 251
ghost twitters, 294, 304n31
as online personal public archive, 130, 209, 250

U
underrepresented populations, 205–206, 211, 251–252
University of Virginia Digital Collections Repository, 258–260
usability, 61, 87
User Collection Tool (UCT), 261, 261f, 262f, 263f, 269, 270f
user tasks in PIM and ARM, 41–44
user-generated content on web, 206–215
active vs. passive interactions, 214–215
demographics of, 218–221
micro-contributions, 206–209
power law distribution of, 210–213
See also social media

UVaCollab learning management tool, 270, 271f

V
value of archives, 2, 242
See also continuing value, concept of; primary and secondary value of records
value of information
difficulty of assessing in advance, 101
informational vs. evidential, 222
as judged by those contemporary to the material, 216–218
and strategies for organization of, 57
variability of, 103–104
value (monetary) of manuscript collections, 288
variants of the same item, identification of, 299–300
version control in photography collections, 178–179, 185, 193
video sharing, 206–207
Virtual Exhibition Tool (VET), 274
virtual personal archives, 131
virtual reading rooms, 253
volume of information. See size and complexity
W
wayfinding process in PIM, 50
Web 2.0 technologies as tools for collective intelligence, 130
web collecting strategies, 222–231
depth-first vs. breadth-first appraisal, 228–231
and documenting those not creating digital records, 251–252
logistics of, 225–228, 226t
personae as provenance, 223–225
web extraction technology, 227
web storage devices. See cloud computing; commercial service providers
weeding of records
benign neglect as, 107, 110, 111–112
destruction of records, 85
in photographic collections, 182–184
during processing, 296
Wikipedia, demographics of user-generated content of, 219
wisdom of the crowd. See crowd sourcing
witnessing
in digital world, 129–131
in face of determination to destroy memory, 116–117
as function of personal collections organization of, 116, 129, 166
as a kind of recordkeeping, 115–116, 117–118, 122
Workbook on Digital Private Papers, 168
workflow technologies
backup routines in, 188
changes in, 81
design of, 171–173, 172f
in management of digital photography, 170–171
and multitasking, 59–60
role of provenance in, 32
and version control, 179
work/life interactions, 46–48, 49, 81
worldviews, incompatible, collection of, 124–125, 142
See also points of view, multiple
write blocker devices, 292
Y
YouTube, 215, 219