

From *Personal Archiving: Preserving Our Digital Heritage*, edited by Donald T. Hawkins, Information Today: 2013. **Use of this chapter is for noncommercial purposes only.** The volume is available for purchase from:

<http://books.infotoday.com/books/Personal-Archiving.shtml>

Chapter 13

The Future of Personal Digital Archiving: Defining the Research Agendas

Clifford Lynch

Executive Director, Coalition for Networked Information

In this brief concluding chapter, I will first try to frame and situate the various developments that are now being considered (appropriately or otherwise) under the catchphrase “Personal Digital Archiving” and explore some of the many places where these developments interconnect with other (often not academic) fields of endeavor. In part, my approach will be through the autobiographical lens of my own encounters and engagement with various streams and tributaries of research over the last decade and a half that have converged into what is now personal digital archiving.

These include efforts to digitally capture life as it is lived – Gordon Bell’s “Total Recall” work, ARPA’s Lifelogs, Microsoft’s SenseCam. This work was strongly influenced by information retrieval research in many ways, but reached into very interesting potential applications such as helping people with deteriorating memories. The major UK study on digital lives, led by Jeremy Leighton John, started from consideration of how to acquire and manage new special collections that included extensive digital materials. It grew to a broader inquiry into the nature of these new digital materials and their implications, not only to cultural memory organizations but to a wide range of scholars and to the general public, culminating in the milestone Digital Libraries Conference hosted by the British Library in 2009. (In an ironic and depressing testimony to the ephemeral nature of digital content, as of the end of 2012, none of the links to this material at the British Library web site seem to work.)

Over the past decade, personal information management has emerged as a new subfield of information retrieval research. A team at the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland, College Park, has taken the lead in exploring applications of digital forensics tools (primarily developed to support the intelligence and law enforcement communities) to capture and understand personal digital archives. Since 2010, the Personal Digital Archiving conferences organized by Jeff Ubois and hosted by the Internet Archive have served as primary focal points for people interested in personal digital archiving (the 2013 meeting moved to the University of Maryland, College Park). These conferences have provided a broad and generous stage for a very wide array of work in digital archiving, personal and otherwise, as well as developments in related fields such as genealogy, analysis of social media platforms, media studies, and public history. They brought together researchers, engineers, activists, entrepreneurs, archivists, librarians, and collectors among others.

But looking back over this vast body of work with the benefit of hindsight, much of it feels like *context* to the study of personal digital archiving, or *external viewings* of personal digital archiving from well-established vantage points situated in other disciplinary traditions rather than a direct engagement with the core issues themselves. Personal digital archiving as a

field of study still demands clear definition and delineation, and my hope in writing this chapter is that, as someone not intellectually committed to any specific one of the many adjacent disciplines, I can help to advance this definitional process.

The remainder of the chapter sketches what I believe are the central research agendas for personal digital archiving today, as these are shared among researchers, archivists, librarians and curators. I focus on issues that seem to me to be specific to personal digital archiving and do not much consider the many broader fundamental problems in digital preservation (such as ensuring that a Microsoft Word or Excel document created in 2008 is still meaningfully readable in 2058). Progress in basic digital preservation will be important to personal digital archiving, and will, to some extent, probably shape developments, but the personal digital archiving area introduces new issues – many not primarily technical in nature – that go far beyond digital preservation.

While this concluding essay has not been explicitly coordinated with the other chapters in the volume in hand beyond a sharing of chapter authors and titles, it is my hope that it will help the reader to synthesize many of the perspectives presented in the earlier chapters, some of which look at personal digital archiving explicitly from the perspective of well-established older adjacent fields such as the development and management of special collections in libraries.

Scoping the Personal Digital Archiving Challenges [A]

For the last three decades I have been trying to understand the ways in which information technology and ubiquitous computer communications networks are reshaping the scholarly and cultural record of our civilization. I have also been concerned with the effects of these technologies on individuals and how they live their lives: how they communicate and share, remember, and learn. And, of course, the migrations and connections between the private sphere and the public (or collective) record are central.

I should also note at the outset that, even within the overall “western” tradition, there are great variations in both legal and social traditions and understandings from nation to nation. Thus personal digital archiving can be explored within a national context, or from a comparative international perspective. The international perspective has an increasingly pragmatic element, as the network crosses national boundaries and legal and regulatory frameworks very casually. Issues beyond the western traditions are almost completely unfamiliar to me, and I will not pretend to offer an understanding of how these considerations might broaden the research agendas in question here.

Digital Records of Individuals [B]

It is clear that for many people, much of their communication is now recorded in electronic mail, communication and social media systems like Twitter, MySpace, Facebook, Tumblr, Flickr and the like, and even in phone logs and occasional records from cell phones or multiplatform systems like Skype. Communication and sharing have become constant and casual, and much more day-to-day experience is documented through these tools. With digital cameras, and then particularly with the merger of cell phones and cameras, this collection of records takes on a much more visual dimension, incorporating enormous numbers of still and moving images. Documents of all types, including to-do lists, may well persist in digital form. Bills, invoices, statements of account and other records of commercial transactions have moved into email or various online systems.

Interesting contested areas exist: search logs, records of purchases, and other kinds of interaction history may be held by individuals or by businesses who will share them back to the individual under various terms. Related (particularly in important public policy ways) but slightly out of scope here are the vast and ever growing set of records *about* specific individuals and their activities held by retail businesses, credit bureaus, insurers, the medical system, background check services, governments and others. In many cases it is very difficult for the individuals in question to even obtain copies of these records, let alone address issues about their accuracy or dissemination.

New forms of personal records are moving into the digital realm, notably medical and genomic information. With the rise of the new generation of instructional technology (massive open online courses (MOOCs), advanced learning management systems and course delivery platforms, and similar developments), individuals will create learning records as detailed and perhaps even more revealing than medical records today, and some individuals will doubtless at least try to have shared custody of these records with the learning platforms.

A small but growing number of individuals have embraced various of these trends and have deliberately tried to intensify the amount of data collected; we find movements such as “sousveillance,” (a term popularized by Professor Steve Mann, a play on “surveillance” that speaks to individuals surveilling themselves and events around them from the bottom up), life logging, “quantified individuals,” (popularized by Professor Larry Smarr, among others), some of the personal genotyping (and soon, sequencing) sharing communities organized by companies such as 23andMe.com, as well as those keeping records from time and task management systems.

The actual custody and storage of all of this information is very messy, and it is getting more complex as time goes on. A decade ago, most of it was on local storage media, though an important class of information exists that is housed by a service provider that, in some sense, shares ownership or use of the data, and may or may not let the individual even download copies in useful forms. Today, we find more information scattered about the network, in some cases in remote storage offered as a service, but in many other cases actually embedded in some network-based system – at stores, banks, hospitals, social media platforms – that attends to its housing, structuring, and sharing. Very few of these network systems make any meaningful stewardship or preservation commitment, though all too often users assume that such commitments exist, in spirit if not in an explicit contract. And, of course, some information lingers in discontinued or abandoned services, or vanishes (for example, consider the fate of some of the early social media platforms, such as GeoCities, as case studies.)

More and more of the cultural products that individuals acquire, enjoy, collect, share, and keep – books, music, videos, games – are moving to electronic form. In digital form, they are often jumbled together with materials that individuals have themselves created, along with records of their acquisition and use (purchase histories on Amazon, playlists and play frequencies for music, etc.). It is interesting to think about how scholars have exploited the (fairly rare) records of book acquisition and building of personal collections by important historical figures, or occasionally have been able to systematically study the annotations that those historical figures may have added to the works that they owned, and then compare these scholarly practices to what will be enabled by today’s records of digital lives. .

Individual back-up practices and even commitments to maintaining meaningful backup vary tremendously, and until the last few years, effective ongoing back-up strategies for casual computer users were substantial technical challenges. Factoring in the complexities of information stored in a wide range of independent systems and services, the interactions between

computer security threats and information persistence, and the long time horizons involved in the records of a lifetime, it is inevitable that for many people, at least some material will be lost over the years.

Clearly, there is a vast research area in understanding the nature of these personal records and how they are changing year to year; understanding how people think about these collections of records, what importance they assign to them, and how these views change over time; and understanding the issues involving the distinctions between personally held and controlled records and records about individuals that are held in remote systems, sometimes with limited access. It is important to study not only what people are doing with personal records they create today, but also what has happened to the personal records that they created a decade or two ago. There is an enormous variation in individual behaviors in all these areas, and it is valuable to try to correlate behavior with various kinds of demographics, to try to understand how behaviors develop, how they are learned, and how and why they change.

Many useful pathways exist that can advance this research agenda. A variety of humanistic and social science techniques can be applied to understand individuals in depth or to survey large groups of individuals. But there are also methods involving the instrumentation of personal computers, social media and e-commerce platforms, and various software packages that can provide a complementary set of insights; unfortunately the vast majority of this information, when it exists, is viewed as highly proprietary. Strategies for opening up some of these system-side data sources could offer a very high payoff.

An additional research agenda deals with pragmatic advice and best practices that can be offered to the broad public for dealing with life in the digital world, for ensuring the long-term survival and usability of the electronic records that they create, and the cultural materials that they acquire. This also connects with guidance about how to maintain privacy where desired and how to secure identity and property in the digital environment. Libraries, in particular, are increasingly being called upon for advice in this area. Technology developers can also benefit from this agenda: it is interesting, for example, to see the very welcome vendor investments in developing easy to use consumer-oriented integrated backup, recovery, and migration software after long decades of neglecting these problems (consider, for example, Apple's Time Machine, Migration Assistant, and the automatic recovery features incorporated in its recent operating system releases).

How the Private is Managed over Time [B]

The issues raised by the death of an individual and what happens to his or her possessions impose a good deal of clarity on the nature and scope of those possessions. Fifty years ago, when someone died, it fell to the executor or an heir to sort through the deceased's personal papers. Typically they were stored at home, with perhaps a few critical documents held by a family attorney or in a bank safe deposit box. In very rare occasions, if the deceased had been a writer or perhaps a composer, copyrights would be recognized as an integral and specific part of the estate, and one might see something like a "literary executor" put in place specifically to deal with copyrights and the fate of unpublished materials.

Today, for the vast majority of the general public, simply determining the scope of an estate that includes digital materials (private, shared, and acquired) scattered across a wide range of services and storage is a formidable task. Resolving issues around ownership, inheritance, and meaningful transfer of access, possession, or control (the clumsiness of the language here itself

suggests the complexity of the problem) is a tangle of legal, contractual, technical challenges further complicated by a lack of overall social consensus in many areas. Tragically, many of the cases that have shown the shortcomings of the current level of understanding and practice have been those involving young people who were sometimes enthusiastic users of social media platforms, and who died suddenly, often in military service, leaving grieving parents, spouses or friends to try to salvage their memories.

Note that at the time of an individual's death his or her digital life is always a mixture of the deliberate (intentionally saved and retained) and the accidental (once saved and never subsequently weeded, or just there by happenstance and never cleaned up). This mix will vary. Determining the intent of the collector/creator may be difficult or impossible in many cases, which will greatly complicate the interpretation of these materials. Further, identifying things that were being kept for personal sentimental value is very hard, and the longer-term significance and importance of such materials may be very difficult to evaluate. Understanding the current situation and helping to structure better practices and solutions through legal and public policy means (including perhaps new best practices in estate planning) with respect to digital possessions of all kinds is an important research agenda. But the research agenda is actually much broader and deeper and extends to the public understanding of what constitutes a "digital estate", what expectations surround the disposal of that estate, and the extent to which this is in conflict with current legal understandings or commercial practices. For example, to what extent can a collection of digital music or ebooks be inherited in the way that physical books and sound recordings have been, and how is this transfer actually accomplished?

The Ambiguity of Shared Materials and Spaces, New and Old [B]

Individuals have always been contributors to, and custodians of, collections of shared, but generally not public, materials, which are typically viewed as belonging to a family (photo albums, narratives, letters, documents or diaries related to deceased or elderly family members). Sometimes, usually less formally, one encounters similar materials for a group of friends of long standing. And, of course, these groups, family or friends, have been subject to fracture through feuds, divorces, etc. that have created controversy and hard feelings about who has access or can have copies of the material in question. One advantage for digital materials, at least sometimes, is that everyone can have their own copy of shared materials easily and cheaply, as opposed the treasured family photo album (the negatives being long lost) or the collection of handwritten letters that were typically held "in trust" by one individual. How are family records managed when third party social media or media sharing platforms are in play, and who should be responsible?

The boundaries of family are being stressed by many developments. More frequent divorces and remarriages, and longer lives create more complex extended families. The ability of the internet to re-connect people has made this much more complex in recent years. What is one to make of an (alleged) distant cousin who contacts you on a genealogy site offering to trade information about a common great-great-great grandmother you have both been researching? (In effect, this is an offer to merge personal or family databases in a very controlled way: commercial genealogy sites today host much more than genealogy, using such records as an armature for all kinds of documentary material about individuals and families.)

The internet can reverse diasporas originating from a village in the Balkans or Somalia; the individuals involved are linked by a complex web of relationships based on kinship and

family, friendship, and shared heritage. It can re-assemble most of the members of an elementary school class, or a group of high school students that graduated together, fifty years afterwards. What, if anything, is the common digital archive of such a group, and how is it assembled and managed?

On current social media platforms, it is quite common for *ad hoc*, dynamic groups (of not just close friends or family but sometimes rather casual acquaintances) to collectively create a corpus of inter-linked material. The nature of social media platforms is to facilitate the movement of material, often by sharing (copying), from one circle of people to another. When is it appropriate to share in this way? When can someone demand that material be taken down, and is that likely to be effective? What are the tacit social understandings about archiving and re-use? The law is not enough (and not particularly clear on these points). How do third party platforms alter the picture, where terms of use (contracts) may also come into play, and where the platform operator itself may claim some rights?

Sometimes the common interests that bind are those of love, friendship, or even just collective endurance and participation in disasters or other great events. Other times, however, these common interests are defined by shared intellectual interests, common political or social beliefs, artistic collaborations or similar ideas, perhaps also reinforced by shared experiences. Any of these can form the basis for the creation of shared, private digital collections. The period of creation or aggregation can be relatively brief (a common disaster, an infatuation), enthusiastic, passionate, trusting, but the unraveling and bickering about the ongoing fate of this shared collection can span lifetimes.

Imagine a continuum that begins with very small groups, families and intimate friends and extends through to formal *organizations* which define and structure information-sharing interactions and preservation, establishing frameworks surrounding shared materials through organization policies, terms of employment or participation, records management, government, legal or financial regulations, and related strictures. Between families and friends and formal organizations there is a fascinating ambiguous territory that has been incredibly fertile in the past, offering a home for intellectual movements, creative collaborations, artistic collectives, protests and political movements. Think of the Bloomsbury circle, the Beats, the Surrealists and the Futurists, the range of activist movements in the 19th and 20th centuries, and the like. But the nature and variety of this territory, along with family archives, was historically quite circumscribed and simple. In the digital era, in the time of social media platforms, it has become much more complex and variable. We need to understand much better what content and social structures shape this space in the digital world, and the connections between the personal and the collective in this area, as well as the space of families, intimates, and friends.

Individuals typically cannot be understood in isolation, only in the context of their times and their relationships, families, friends, lovers, colleagues, etc. So it is with the personal digital archives that individuals create. These archives overlap with those of their families, friends, and others with whom they may share common cause. The evidence so far suggests that these interpenetrations and overlaps are so intimate and so complex that personal digital archives cannot be considered in isolation. Indeed, to understand the nature of personal digital archives we must also understand these small-scale, intimate collective digital archives and how they relate to the personal, which is an essential part of the personal digital archiving agenda.

How, When and Why the Private becomes Public [B]

The private becoming public is probably the most poorly understood, almost certainly the most poorly studied, and yet perhaps the most important area within the overall personal digital archiving research agenda. It is central both in understanding the broad nature of the cultural record and personal digital archiving specifically, because it focuses on the connecting pathways between one and the other.

There is no *mandate* and no *requirement* that personal digital archives or digital lives ever become public. The heirs of some public figures (and sometimes the public figures themselves) despise and distrust biographers; characterizations like “jackals” are common. Throughout history, it has been common for people to ask that various parts of their letters, personal papers, manuscripts and the like be burned or otherwise destroyed; sometimes they have seen to this personally, and other times they have relied on executors or heirs. In the latter case, sometimes their wishes are honored, and sometimes they are not. The choices that have been made over the centuries have had a profound effect on the materials that have reached us as part of the cultural record. The ethical dilemmas, the alternatives and consequences in this area have been well studied, at least anecdotally and in some cases comparatively.

Today we operate in a legal framework that distinguishes ownership of a specific physical copy of a work and the rights to the *content* of the work under copyright. After enough time has passed, copyright ceases to protect materials, even unpublished works, and they fall into the public domain (though given the legislative propensity to extend copyright terms, this may take a *very* long time). After that, materials are kept private only by possession and access control. I am pessimistic that very much genuinely private material will ultimately find its way into the collections of cultural memory organizations by accident in the way that we have centuries old manuscripts, diaries, and similar materials that simply survived by chance until copyright expired and their contents belong to the culture as a whole (though the *artifacts* themselves may well be owned by specific memory organizations). The situation where artifacts pass to memory organizations even though their contents remain under copyright, and use of these contents is highly constrained, is going to be unusual in the world of personal digital archives, I think. Perhaps someday we will see antiquarians making treasure hunts through two-century-old memory sticks, but that seems unlikely. In the digital world, I believe that most personal digital archives will become public only through deliberate action, not by chance and the passage of time.

Ownership of an artifact typically persists as long as the artifact survives, but there are some interesting counterexamples and specialized transition paths for objects that can be characterized as “national patrimony” to reach the cultural memory sector. These are rare, and particularly foreign to American law (though we do recognize international treaty obligations in this area, and we also have legislation such as the Native American Graves Protection and Repatriation Act). But this, too, is an area that demands fresh examination in the digital world, particularly when it is possible to create very sophisticated digital representations of objects, and then decide who gets the object, and who gets a (perfect) copy of the digital representation. We also need to consider what it might mean to establish a regime where material in a personal digital collection or archive might be considered national patrimony.

So, if the vast majority of the movement of personal digital archives and collections from the private to the public sphere is going to occur through deliberate action, how does this happen? Clearly it is connected to the questions about inheritance and estate discussed earlier. A personal digital archive might be contributed or sold to a library or archive (indeed, today important collections of personal papers are pricing in seven figures). To be most effective, this

transfer probably should take place with the active collaboration of the individual in question, while she or he is still alive, rather than posthumously in collaboration with the individual's estate. The recent acquisition of the Salman Rushdie archives (including the ongoing engagement of Rushdie himself) by Emory University is an excellent case study. Note that the marketplace for the acquisition of personal archives is very different than it was in the pre-digital world, where individual items might be sold and resold; here the stock in trade is rights to a *collection*, most commonly with some help from the creator in gathering it, which will become ever more necessary as large segments of a "digital life" leave an individual's personal machine for locations on the web, in social media systems, in the cloud, etc.

In the future, the acquisition of personal digital collections is going to be more and more about relationships and intent, and about an acknowledgement (grudging or otherwise) for the interests of the scholarly and biographical world. Consider, as an illuminating thought experiment, the market for acquiring Thomas Pynchon's (as far as I know imaginary) personal digital archive (or pick your own favorite important literary author). Would it make a difference in value if it were known that he wrote (and edited for publication) all of his books since 1995 in Microsoft Word *with change tracking enabled, and had saved these versions*? One can only imagine the range of unintended consequences in the race to identify and recruit individuals that will be part of the acquisitions program of particular special collections in libraries and archives early in their professional lives, and the kinds of constraints to which they might be asked to agree.

Some of this research agenda is framed primarily by memory organizations that are trying to manage the transition from private personal digital archives to a curated, relatively public environment hosted by the library, archive or museum. The issues include negotiation of suitable contracts or deeds of gift, privacy, embargoes, and redaction; the appropriate application (if any) of forensic technologies; and of course (more and more) how to handle material that is spread across various servers on the network rather than housed on local storage media that can simply be copied. Legitimately, I think that the research agenda should include symmetrical guidance on what prospective donors to the cultural record need to consider in making choices and negotiating such donations. The research agendas here also recede away from personal digital archives themselves to best practices not just in acquiring but in subsequently curating and providing access to these collections, and in the ways that scholars can make best use of them. We see questions about:

- When and how digital lives should be linked to one another – typically spanning across different cultural memory organizations – and when they should stand alone; to what extent should these links be created computationally and to what extent should they be created by human intellectual analysis; when is the creation of such linkages the appropriate work of curators, and when of scholars. Note that the options here are much more extensive, and it is much easier to make progress than it has been historically with physical collections of personal papers, which are often extensively fragmented and scattered across memory organizations,
- The appropriate use of sense-making and information retrieval tools and technologies, including forensic tools originating in the law enforcement and intelligence communities and those developed to support legal discovery in large-scale electronic records collections, both by curators and by scholars, to gain an understanding of what is in a given personal digital archive and to locate specific materials of interest.

Beyond the portion of a digital life that is basically viewed as “owned” by the individual (including the parts that interconnect with shared social spaces), there are other segments of personal digital archives, lives, and collections that also need to be examined. One part is the collection of cultural materials that an individual has amassed during his or her lifetime – ebooks, sound recordings, downloaded web pages, video games, and so much more. Some of this is licensed in consumer markets in very structured ways, and may or may not be able to be moved into a collection that is public in some sense. Sometimes inventory records are as important as the materials themselves, particularly if the underlying ebooks, videos or other materials are still widely available; annotations and marginalia are another very important resource.

But digital collections can be much more complex. Consider music: one might find a personal collection that includes commercial music from multiple sources such as Amazon or the Apple iTunes store, live materials acquired from various online sources, audio tracks taken from YouTube, and ripped CDs (that may or may not have been entirely legal in the first place). It is a safe bet that the provenance information will not always be quite as thorough as a lawyer might ideally like, although the metadata may be extensive and add very considerable value. This is a challenge at two levels: first, for the cultural memory organization that is interested in capturing the inventory, evolution, and use of a digital collection by its owner, and second, and equally important, ensuring the survival and availability of perhaps ephemeral, “grey,” or similar material that the collecting individual chose to collect, curate, and preserve which might otherwise not find its way into the long-term cultural record. We may need new legal provisions for cultural memory organizations that undertake stewardship of these collections to seek safe harbor from liability for copyright infringement, at the very least, if we are to be able to genuinely preserve what is in these personal digital archives.

Equally problematic, but for very different reasons, are the massive amounts of information *about* a given individual that are held by corporations, government, the medical establishment, and similar sectors. In some cases, a living individual has some right to at least obtain copies of these records (though often in unhelpful printed forms rather than as digital files); the passage of these rights through inheritance is complex, messy, and limited. Some bits – airline frequent flyer records, credit card histories, perhaps even credit bureau records – will likely vanish within a few years after the individual’s death (or when the holding corporation becomes cognizant of the death, which can take more years) because they have vanishingly small commercial value. (This area has yet to be fully exploited; imagine a potential marketplace in credit card records, itineraries, or interactions with Amazon, for dead public figures. Many of these records are relatively small; it may cost more to delete them promptly than to treat them with benign neglect.) Records of interaction with commerce or social media sites may persist longer. We do not yet know how rapidly the contributions from old interaction records decay in value as recommender systems use these interaction databases to make suggestions (and the answer is likely to be highly proprietary, and perhaps vary wildly from one applications domain to the next).

Often the contributions of the dead are heavily linked to the ongoing, active contributions of the living in social media spaces, and there seems to be no pressure to remove material by the platform operators; indeed, some social sites appear willing to host memorial pages indefinitely and presumably view them as assets.

Medical records are likely to be an emerging battleground in the coming years; historically, access to these has been controlled by a byzantine mix of Federal and State legislation in the United States that seems to serve everyone poorly, but until today the interested

parties have been limited to a few biographers; family members, mainly concerned understanding about hereditary conditions; and people pursuing lawsuits, usually around alleged errors in end-of-life care. If we look at the medical records that will be much more common in the near future, they will include genotyping or gene sequencing data, detailed machine readable medical history records, perhaps prescription or insurance claim information, tests, and imaging. Whether the individual is dead or alive, this is prime material for data mining on a large scale. Some countries, such as the UK, are setting policy about access to these medical records as a matter of national public health strategy, international competitiveness as a partner with multinational pharmaceutical firms, and medical entitlement cost control strategies. There is a great deal of money at stake. It is not clear to what extent, if any, the deceased has control over the disposition of his or her medical records; while they are alive they usually have some right to view or perhaps obtain copies of these records, but typically do not “own” them, and the health care or insurance providers that do claim ownership may be free to destroy them under some circumstances. One could imagine a very desirable – though perhaps currently impossible – future where an individual can choose to place his or her medical records (before or after death) in a genuinely public research commons, perhaps somewhat like signing up to become an organ donor.

Dossiers held by various government agencies on deceased individuals will be another point of contention, especially as these dossiers apply to more and more people and expand seemingly without limit. There is some precedent in the treatment of military records, FBI files, and similar material. We can see a mix of public policy considerations and pragmatic issues shaping the research agenda about how this part of an individual’s “digital life” might move from the private to the public sphere.

Personal digital archives are, in a very real sense, an optional, even accidental, part of our collective cultural record. I know of very little systematic research (as opposed to endless anecdotal case studies) on how these collections have been acquired and integrated into the cultural record historically. There is a pressing need to understand what is happening at present. These personal collections, and now personal digital archives, are the signature elements that distinguish many of the genuinely great research collections housed in libraries and archives. They lend irreplaceable depth and richness to the broad cultural record.

Both within specific scholarly communities and as a broader society, we need policy discussions about how many digital lives we need to capture, and what organizations should take responsibility for collecting them. This has connections to the evolving missions and strategies of not just national and research libraries but local historical societies, public libraries, and similar groups. In some cases, we can see traditional cultural memory organizations seeking personal digital archives of individuals they deem important for one reason or another, but we also see new players – StoryCorps, the Internet Archive, and the BBC in the UK – that are soliciting contributions from the general public as well as commercial organizations such as Ancestry.com and trying to establish lines of business maintaining digital lives. There are new ideas afoot in areas such as public history, and we need to understand the collections necessary to support these activities, where they will be housed, and how they will be curated and funded.

Indeed, one can easily imagine a near future where we see expectations established among the general public about a “*right* to archive one’s digital life” in some appropriate service. This leads us directly into a final set of questions about the extent to which genealogy, factual biography, national biographical dictionaries (and their successors: think about Wikipedia), birth and death records, and similar materials belong to the public or private infrastructure, how they

are governed, and who they represent. Personal digital archives, as they move from private to public, have a rich interconnection with all of these developments.

Mapping these connections and developments leads, ultimately, to a new, but still poorly understood and poorly articulated research and public policy agenda surrounding the broader nature and evolution of the intellectual and cultural record, which goes far beyond the scope of this chapter. This is much more than personal digital archives made public. But understanding the role that personal digital archives can play in contributing to this record, and the mechanisms by which they might contribute to it, is clearly a very useful source of insight.

Acknowledgements [A]

Don Hawkins has been amazingly patient and supportive yet gracefully persistent in seeing this chapter to conclusion. My thinking on these issues over the past few years has benefited from discussion with people too numerous to list here, but I want to particularly recognize the contributions of the Buckland/Lynch/Larsen “Friday Seminar” at the University of California, Berkeley School of Information on November 30, 2012 in discussing a draft of this chapter. Cecilia Preston, Joan Lippincott and particularly Diane Goldenberg-Hart provided helpful suggestions on this manuscript; Michael Buckland, who has been an invaluable collaborator as I’ve developed my thinking in this area, was kind enough to offer not only extensive and insightful comments but an extraordinary thoughtful set of broader reflections in response to a late draft of this essay. Thank you all.