

PERSONAL ARCHIVING



Preserving
Our Digital
Heritage

Edited by
DONALD T. HAWKINS
Foreword by Brewster Kahle

Contents

Acknowledgments xi

Foreword xiii

Brewster Kahle, Internet Archive

Introduction xv

Donald T. Hawkins

**Chapter 1 Personal Digital Archives: What They Are,
What They Could Be, and Why They Matter 1**

Jeff Ubois, MacArthur Foundation

Chapter 2 Personal Archiving for Individuals and Families 11

Danielle Conklin, Cotton Gloves Research

**Chapter 3 The Library of Congress and Personal
Digital Archiving 31**

Mike Ashenfelder, Library of Congress

Chapter 4 Software and Services for Personal Archiving 47

Donald T. Hawkins

Chapter 5 Digital Inheritance: Tackling the Legal and Practical Issues	73
<i>Evan Carroll, The Digital Beyond</i>	
Chapter 6 Social Media, Personal Data, and Reusing Our Digital Legacy	85
<i>Catherine C. Marshall, Microsoft Research, Silicon Valley</i>	
Chapter 7 Reading Ben Shneiderman’s Email: Identifying Narrative Elements in Email Archives	109
<i>Jason Zalinger, University of South Florida, Nathan G. Freier, Microsoft Corporation, and Ben Shneiderman, University of Maryland</i>	
Chapter 8 Faculty Members as Archivists: Personal Archiving Practices in the Academic Environment	137
<i>Ellysa Stern Cahoy, The Pennsylvania State University Libraries</i>	
Chapter 9 Landscape of Personal Digital Archiving Activities and Research	153
<i>Sarah Kim, University of Texas at Austin</i>	
Chapter 10 Active Personal Archiving and the Internet Archive	187
<i>Aaron Ximm, Internet Archive</i>	
Chapter 11 Our Technology Heritage	215
<i>Richard Banks, Microsoft Research, Cambridge, U.K.</i>	
Chapter 12 New Horizons in Personal Archiving: 1 Second Everyday, myKive, and MUSE	243
<i>Donald T. Hawkins Christopher J. Prom, University of Illinois Peter Chan, Stanford University</i>	

Chapter 13 The Future of Personal Digital Archiving: Defining the Research Agendas	259
<i>Clifford Lynch, Coalition for Networked Information</i>	
About the Contributors	279
About the Editor	287
Index	289

Personal Digital Archives: What They Are, What They Could Be, and Why They Matter

Jeff Ubois
MacArthur Foundation

Personal Archiving is a timely book highlighting some of the most interesting work that has been done in the domain of personal digital archives. It outlines recent accomplishments in digital preservation, an agenda for future research and development, ideas about the new role of archival institutions, and practical information for those wishing to archive their personal histories. By helping to build a common understanding of personal archives, this book supports collaboration between diverse types of institutions and individuals working in different disciplines.

Science and culture in the 21st century are born digital; today's Einsteins, Rembrandts, and Bachs create and record their works on digital devices. The works of yesterday's eminent individuals are rapidly being digitized, and there is a growing expectation that all of humanity's creative works are or will be instantly accessible in digital form.

The minute-by-minute, day-to-day record of our lives is becoming richer all the time. More than 1 billion camera phones will be sold in 2013, and by one estimate, more than 10 percent of all the photographs ever taken were taken in 2012.¹ Personal digital archiving may focus on genealogical data, but it includes much more: What we have written, what we have read, where we have been, who has met with us, who has communicated with us, what we have purchased, and much else is recorded digitally in increasingly greater detail in personal digital archives, whether they are held by individuals, institutions, or commercial organizations, and whether we are aware of those archives or not.

The ability to create, collect, and preserve personal archives—once a privilege of elites—has now been extended to billions of people around the world, and increasing numbers of them are concerned about the long-term prospects of their data. In the past, individuals who wished to send a message to their descendants or to future generations needed vast sums of money (and perhaps stone-masons and slaves). Not anymore. Personal digital archives democratize access to the future and, for some, fulfill the deeply human urge to leave a legacy.

There is a corollary to all this, which is that the events that will eventually be regarded as historically important are now often first recorded by ordinary individuals. The personal is becoming social and collective. Like Abraham Zapruder, the individual who by chance filmed John F. Kennedy's assassination with his home movie camera, many of us are recording events as they occur and sharing them with the world, sometimes before the mainstream media can, as witnesses present in Cairo's Tahrir Square or at the July 2005 London Underground bombings and other recent global events have done.

The ideal of the *archive* is in the zeitgeist. People casually refer to themselves as archivists; movie plots refer to archives as repositories of truth and hidden knowledge. Newspapers and radio stations routinely refer to their archives, and sometimes the only fees they charge are for access to them. Archives have become a counterweight to the sense that

we are awash in the fake; the archive promises some kind of authenticity—none more than the personal archive.

Given a little technical help with production, the memories of individuals with few obvious claims to fame can become compelling stories, an important part of the historical record, or even the basis for popular entertainment. For example, the StoryCorps archive in the American Folklife Center at the Library of Congress (www.loc.gov/folklife/storycorpsfaq.html) contains 45,000 interviews of individuals from diverse walks of life, along with many supplementary materials. With 30,000 hours of audio, it is one of the largest “born digital” collections of oral histories in the world. Radio program segments based on these interviews are heard by 14 million listeners on National Public Radio every week, and the organization is far along in its mission to provide Americans of all backgrounds with the opportunity to record, share, and preserve the stories of their lives.

Clearly, the concept of the personal digital archive is on the minds of many, from authors and artists to historians and genealogists, and from entrepreneurs and engineers to funders and managers of memory institutions. Over the last few years, a common language, shared awareness, and a new field of study centered on personal archives have begun to take shape through the work of a new community of digital archivists. But it has yet to be fully defined or realized.

Not long ago, there was little question that personal archives were the sum of published and unpublished material written and collected by an individual. That definition has become inadequate as the boundaries between what is personal and what is social have blurred. It may now be more accurate to say that personal digital archives are collections of digital material created, collected, and curated by individuals rather than institutions.

It is important that we create a shared understanding of personal archives, appropriately set expectations about them, and most of all, bridge the gap between our imagined ideal and the imperfect tools and services we use to preserve the data we care about. That will

require solving some difficult problems that are interrelated in complicated ways.

It will also turn on the resolution of two much-contested issues: copyright and freedom of expression. Each threatens the other, and few of the interests lined up on each side of those debates have personal archives and collections at the center of their concerns. Still, those who love free expression should love the personal archive. And if there is not a lot that archivists can do to affect the outcome of the conflicts over copyrights and over free expression, there are some problems that can be addressed in the more immediate term.

Five issues stand out: 1) funding and costs, 2) the relation between the commercial and noncommercial sectors, 3) the relation between individuals and institutions, 4) technology and design, and 5) culture and expectations.

Funding and Costs

Economics is one of the enduring questions surrounding personal archives. Who pays, how much, for what exactly, and why? Archiving has always been driven by the anticipated future value of information, which in the end is based on a guess. Calculating the costs associated with perpetual storage—or very-long-term access—is also incredibly difficult. And as the cost of information capture and storage continues to decline, the default course of action is to save more and more.

Governments, libraries, archives, and research organizations have collectively spent millions of dollars to develop cost models for digital archives because an ability to predict costs is essential for any organization that wants to promise long-term access. Often, these models are too complex for practical use by smaller organizations. What's needed are “pay once/store forever” services. These are not yet widely available, though a few universities now offer them to faculty. This pricing model seems likely to become more common, in part because agencies that fund scientific research and cultural production are requiring their grantees to provide a plan to preserve project data.

But there is a gap in the way various funding systems work. Science, culture, and scholarship are typically funded on a project basis for a limited term, not in perpetuity. Traditionally, libraries promised long-term access to scientific, cultural, and other scholarly publications, but much of what is being produced now does not fit as well into that system. Closing the gap between project-based funding and preservation in perpetuity is as much a matter of economics or economic understanding as it is a matter of technology.

What many people need are simpler models that make it possible to decide what they can afford to save, which will mean getting to commodity prices and to fixed unit costs, particularly for ingesting (taking in and assimilating the data) and its storage. How much does it cost to digitize an average-sized box of papers? How much does it cost to store a terabyte, forever? We need to know.

Market Solutions

Facebook's Timeline reaches a billion people. It is effectively a personal archive. Few of the more serious attempts at archiving from a traditional perspective will reach that many people as fast or as soon, but it would be foolish to assume that commercial services will continue to improve to the point where all personal archiving needs might be met by the market.

Commercial organizations have a poor record of caring for personal collections; their mission is to turn a profit, which is reflected in their terms of service. When commercial service providers state that they accept no responsibility in the event of data loss, they mean it. Commercial organizations, particularly startups that promise continuity and persistence over time, are suspect, or should be, and entrusting the task of long-term preservation solely to businesses is unwise. However comfortable it may be to imagine that data is safe with today's billion-dollar service provider, when corporate priorities change, personal data may be deleted, as former users of Yahoo!'s GeoCities and many other services can testify.

But we must not say that commercial services have no role in personal digital archives. A factory-like approach to ingest, like ScanCafe (www.scancafe.com), can usually do a better job of scanning photos than most individuals can. Software that automates metadata extraction for purposes such as speech and facial recognition will be valuable, too. Copyright clearance may be another area that is best handled by commercial services. There is also a vibrant trade in manuscripts; although digital archives may lose the aura of authenticity that a handwritten diary has, some personal digital archives are contained in physical devices. The laptop or memory card containing photos and email that once belonged to a famous author may become valuable in the way manuscripts have. The U.S. market for paper scrapbooking is declining, but in 2003, the business still earned \$2.5 billion.²

In the end, markets are about exchange, and while personal digital archives may be something we use while we are living, much of their value (or promise) is in their ability to allow communication with our descendants. It is hard to envision a market between people who are dead and people who are not yet born, and the nonmarket motivations will be critical to the future of personal archives.

New Relationships Between Institutions and Individuals

The institutions that have traditionally taken on responsibility for digital preservation—libraries, museums, and archives—must face questions over whether they have a new responsibility to collect digital personal materials. One approach has been for leading institutions to provide guidance to individuals who are interested in preserving their own digital materials. For example, the Library of Congress hosts Personal Digital Archiving Day events, which are drawing many people who have already taken on the problem of preserving their own material as best they could.

That kind of learning can go in both directions. Amateurs working outside the system often create the most interesting approaches to personal archiving. There are many fine examples of this, but the Grazian

Archive is my favorite. The general concepts behind it are explored with great wit and understanding in “The Personal Archive: On Retrieving Valuable Cultural Resources” (www.grazian-archive.com/projects/archvpt.html).

Institutions are also developing ways to get help from the public, particularly through crowdsourced cataloging projects. The Church of Jesus Christ of Latter-day Saints has a well-known digital archive focused on individuals that is now more than 10 petabytes in size and is growing by millions of images per month. Those images have been annotated by more than 100,000 volunteers (see www.familysearch.org).

Crowdfunding is another intriguing possibility. Rather than bankrupting institutions that care for cultural materials, personal archives and individual donations could become a source of support. With a “pay once/store forever” cost model, individuals might be able to endow a terabyte, rather like the more common “buy a brick” programs museums and others have used for years. Universities might also provide a terabyte with tuition and thus tie their alumni into the university in a much deeper way.

Tools, Technology, and Services

The future of personal archives will depend heavily on technologies and services developed for other purposes, particularly for capture and storage. New interfaces that include better timelines, maps, social network diagrams, automatic indexing, and data extraction technologies for face, voice, and handwriting recognition are other examples of technologies that will probably be developed for other markets but will also be useful to personal archivists.

Another set of technical needs has to do with integration of multiple sources of information and multiple types of media. Not long ago, personal archives consisted mostly of written materials, augmented by a few photographs, sound recordings, or perhaps home movies. Integrating the many kinds of digital data individuals now collect into

a meaningful whole is hard, yet it is clear that the future of personal digital archives will depend on integrating the multiple types of data we create, collect, and disseminate through our daily activities.

Institutions also have technical needs, and many of the basic functions of traditional libraries and archives have yet to be implemented in software. For example, libraries and archives have managed patron data and the privacy of it for a very long time. The American Library Association Code of Ethics from 1939 (www.ala.org/advocacy/proethics/history/index5) describes why patron data is so important: “It is the librarian’s obligation to treat as confidential any private information obtained through contact with library patrons.” That is exactly the opposite sensibility from the way in which many organizations handle data, and tools that embody the ethical sense of librarians and archivists still need to be built. The donor agreements governing what archives may do with an individual’s papers may require that materials are kept closed for some period of time, or until some number of years after the death of the author, or even after the deaths of all those mentioned in the collection. No software exists to manage that. There are also new designs required for university scholars who need to preserve their own work and to look back on the lives of eminent persons. The Mellon Foundation is now funding a series of projects in this area, but there is clearly more to be done to address the needs of different types of users.

Culture, Expectations, and the Future

Memory shapes our view of the world, and memories are contested by those in power or those who seek it. Memory institutions, such as libraries and archives, have therefore been at or near the center of debates about freedom of thought, freedom of expression, human rights issues, and privacy. The personal digital archive makes us all stakeholders in the outcome of these debates: If, as part of your practice as a personal archivist, you record your interaction with the police, there may be dire consequences.

Discussions about personal archives often turn on questions relevant to individuals, such as, “What happens to your data when you die?” or “What would you pay for a video of your great-great-grandparents?” But the question of what happens when personal archives are joined together has yet to be explored. It may be that we create a new kind of social memory or historical record, one with the potential to change public discourse and political choice. The personal testimonies of 50,000 Holocaust survivors held by the Shoah Foundation (sfi.usc.edu) are the tip of that iceberg.

The development of personal digital archives will affect the long-term preservation of cultural heritage, the future of cultural and scientific production, and freedom of expression. It will change the balance between commercial and noncommercial efforts to ensure open, long-term access to information, and change the relationship between institutions and individuals concerned with preservation. And on our best days, personal archiving may even help us fulfill the Confucian responsibility of being good ancestors to our descendants.

Endnotes

1. Jonathan Good, “How Many Photos Have Ever Been Taken?” 1000memories, September 15, 2011, accessed May 15, 2013, blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox, and “Global Mobile Statistics 2012 Part A: Mobile Subscribers; Handset Market Share; Mobile Operators,” mobiThinking, March 2013, accessed May 15, 2013, mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a#phone-shipments.
2. The market estimates vary, and they are changing, but see Eugene Fram, “The Booming Scrapbooking Market in the USA: Despite Phenomenal Growth, the Future’s Unclear,” *International Journal of Retail and Distribution Management* 33, no. 3 (2005): 215–225, accessed May 15, 2013, www.emeraldinsight.com/journals.htm?articleid=1464122&show=html.

About the Editor



Donald T. Hawkins is a blogger and writer at Information Today, Inc. (ITI), where he blogs about ITI's conferences and writes conference reports for *Information Today*. He maintains the Conference Calendar on the ITI website (www.infotoday.com/calendar.asp) and has developed databases and participated in other special projects at ITI.

Don had a distinguished career with AT&T for over 24 years, retiring in 1996 to start his own consulting business, InfoResources. At AT&T, he was the content manager for several electronic information projects. He began his career in 1971, the dawn of the online information industry, in the AT&T Bell Laboratories Library Network and spent 15 years there, developing and managing its information retrieval and current awareness services and conducting several pioneering studies in online searching, bibliometrics, and other related areas. Following his career in the Library Network, he served in the business units of AT&T on projects related to the research and development of econtent products.

Don was appointed Distinguished Member of Technical Staff at AT&T Bell Laboratories in 1986 in recognition of his pioneering work in introducing end-user searching into the Library Network. He won the prestigious UMI/Data Courier Award for excellence in writing in the online information industry in 1987 for his two articles on artificial

intelligence and online searching, and again in 1992 for his article on intelligent agents for information retrieval.

Don joined ITI in 1998 upon its purchase of *Information Science & Technology Abstracts (ISTA)* and became the editor-in-chief of *ISTA* and, upon its purchase, *Fulltext Sources Online*. He led the efforts to develop a new production process for *ISTA* based on the Microsoft Access database platform, to update and revise its thesaurus, and to improve its quality and coverage of the information science literature. When ITI sold *ISTA* and *Internet and Personal Computing Abstracts* to EBSCO Publishing, he worked with EBSCO to help integrate the two databases into EBSCO's production system and developed the information and computer science component of EBSCO's thesaurus and controlled vocabulary, which is used to index the two databases.

Don is a frequent contributor to the literature, with more than 300 publications, and he has spoken frequently at industry conferences. In June 1996, he keynoted an international conference on electronic publishing in Japan, and he has delivered plenary presentations in the U.S., as well as internationally at conferences in Australia, New Zealand, South Africa, and Israel. Recently, he contributed a chapter on the AT&T Bell Laboratories Library Network to *Special Libraries: A Survival Guide* (ABC-Clio, 2013).

Don was the secretariat for the Association for Information and Dissemination Centers (ASIDIC) before it merged with the National Federation of Advanced Information Services (NFAIS), and editor of the *ASIDIC Newsletter*. He received his BS, MS, and PhD degrees from the University of California, Berkeley.

If you enjoyed reading this chapter of *Personal Archiving*, please visit our bookstore to order a copy: books.infotoday.com

