

We are all Archivists: Encouraging Personal Digital Archiving and Citizen Archiving on a Community Scale

Leslie Johnston
Library of Congress
101 Independence Ave SE
Washington, DC. USA 20540-1300
011-1-202-707-2801
lesliej@loc.gov

ABSTRACT

The more we interact online and manage digital lives, the more we build a born-digital record of our personal contexts and interests. There is a clear mandate that libraries, archives and museums must assume responsibility for educating the public about strategies for personal digital archiving and personal curation, and for exploring new approaches to processing, preservation, and access. Personal digital items initially have value to the individuals who generated them, but once those items are transferred to a collecting institution they will have a collective value to society. We also have a less well-established mandate to work with those citizen archivists who are taking the initiative to save large collections of digital ephemera and the at-risk output of underserved communities. Large amounts or small, all digital materials have the potential to be vital to cultural history studies in the future.

Categories and Subject Descriptors

K.4.0 [Computers and Society]: General

General Terms

Documentation, Experimentation.

Keywords

Personal Digital Archiving, Citizen Archiving, Web Archiving, Digital Preservation.

1. INTRODUCTION

We are all archivists. We are all makers. We are all participants in many communities. There is a seeming increase in interest in “citizen” activities: citizen journalism, citizen science, and citizen archiving. We are all creating personal histories that increasingly include born-digital records and some type of online activities. The more we interact online and manage digital lives, the more we also seek to digitize our analog histories, as well as strive to make some sort of record of our personal contexts and interests.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iPRES2011, Nov. 1–4, 2011, Singapore.

Copyright 2011 National Library Board Singapore & Nanyang Technological University

Initiatives to provide education about support personal digital archiving best practices often dovetail with initiatives to collate or collect personal digital materials. Interest in citizen archiving has moved from the analog to the digital as the awareness of the short life spans of formats, media, and online services becomes more widespread. There is an increased engagement of individuals and communities in digital preservation: personal initiatives are growing into large-scale efforts. How should the digital preservation community encourage and incorporate these activities?

2. “CITIZEN” EFFORTS

We have all seen examples of “Citizen Journalism,” from the tweeting in Iran during the 2009 elections, commuters who were quick to send email with photos to media outlets immediately following 9/11 or the 7/7 London Bombings, bloggers reporting on local news in small communities where newspapers are no longer easily sustained, or something as seemingly straightforward as the ability for anyone to comment on and initiate a discussion about published news stories online. There is an almost unprecedented involvement of the general public in the process of collecting, reporting, analyzing and disseminating news and information. [1]

We might be less informed about “Citizen Science,” a category of scientific work in which individuals or networks of volunteers with little or no scientific training perform scientific research tasks such as observation, measurement, or computation. Prior to the twentieth century, from the Enlightenment through the Victorian era, most scientific endeavors were undertaken by “amateurs,” both trained and self-educated. Science educators encourage people of all ages that they participate in research, be it species counts, environmental and atmospheric monitoring, astronomical observation, or volunteering on an archaeological excavation.

We are of course familiar with “Citizen Archivists,” a term once used primarily to refer to amateur collectors, genealogists, and family historians who amassed physical collections of photographs, personal papers, newspapers, maps, films, recordings, etc., documenting an era, event, place, community, family, or an individual. Such collections are the core of many of our institutional collections.

But Citizen Archivists are now preserving all things digital as well, from personal histories to software to games to the web itself. And we should be encouraging this.

3. PERSONAL DIGITAL ARCHIVING

One clear topic to be addressed is personal digital archiving. As Richard Cox notes in his paper on digital curation and citizen archivists:

“It is likely that the increasing use of digital formats will enhance interest in the preservation of personal archives and that this will strengthen the public’s awareness of the importance of archives, records, and information management.” ... “The exciting aspect of rethinking how archivists will work in preserving personal and family archives is that it may re-open a much greater possibility for reaching the public with a clearer sense of the archival mission, an objective archivists and their professional associations have struggled to do for several decades with very mixed results. It is, however, clearly the case that the public itself is actually sowing the ground for archivists to seed.” [2]

The single most extreme individual effort in personal digital archiving is certainly that of Gordon Bell and his MyLifeBits lifelogging project.¹ [3, 4, 5] After responding to a colleague’s request to digitize one of his books, he started to consider what it meant to have a hybrid personal history, part paper-based and part digital. He considered the combination of a small wearable camera and the Memex proposed by Vannevar Bush [6] and began to document his days with a SenseCam² along with digitizing his personal archives.

For everyone else, they know they need to save their personal digital output and history and records for future generations but are wondering how to do it. Cathy Marshall described the issues succinctly:

“... people archive their personal digital belongings by relying on a combination of benign neglect, sporadic backups, and unsystematic file replication. Even the most valuable of their digital assets -- files representing considerable investments of effort, significant emotional worth, or actual cash expenditures -- are often in danger of being lost. Distributed storage, uncontrolled accumulation of digital materials, a lack of standard curation practices, and an absence of long term retrieval capabilities all point toward an incipient digital dark age.” [7]

There is a recent but growing body of research on personal digital archiving. Some of this research builds on related research in Personal Information Management (PIM) strategies, the study of personal digital information seeking and management, sometimes described as “keeping found things found.” [8,9,10,11,12] Cathy Marshall, Neil Beagrie, Jeremy Leighton John and others have written extensively on personal digital archiving strategies and requirements. [13, 14, 15, 16]

In that vein, there is an equally important emerging area of study on methodologies for the acquisition and management of personal

digital archives at collecting institutions, as well as on the use of digital forensics tools in the appraisal, capture, management, description, and preservation of personal digital collections. [2, 17, 18, 19, 20, 21, 22] Conferences on the issues are taking place internationally³. Awareness of the issue has moved out into the popular press. [23, 24, 25, 26, 27] Many institutions have instituted digital forensics programs and labs to process collections, including the British Library⁴, Emory University [25], Stanford University⁵, and Oxford University⁶.

But what about advice for individuals? Many commercial services have launched to assist in personal archiving and to provide “legacy services” in case of incapacitation or death. [27, 28, 29] Where are collecting institutions in this realm?

A number of initiatives and organizations have released resources or tools to assist in personal digital archiving. While some issues in the file format persistence have been presented in a lighthearted way [30], there are now a number of resources to advise both individuals and institutions on the comparative sustainability of different types of files and media storage. The Library of Congress maintains its “Sustainability of Digital Formats” site.⁷ The National Library of Australia has created the prototype Mediapedia, which documents storage media.⁸

The Paradigm Project released its “Guidance for Creators of Personal Papers.”⁹ The SALT project¹⁰ at Stanford has been prototyping digital self-archiving “legacy” tools for faculty. Penn State University is developing a model for curating digital intellectual lives¹¹. FamilySearch has released a guide for preserving family history records. [31]

The Library of Congress has launched two personal digital archiving initiatives. In 2010 the Library hosted its inaugural Personal Archiving Day, initially held as an ALA Preservation

³ To date, two personal archiving conferences have been held at the Internet Archive (<http://www.personalarchiving.com/>), one was sponsored by the Digital Lives project (<http://www.bl.uk/digital-lives/conference.html>), one was sponsored by the Digital Curation Centre (<http://www.dcc.ac.uk/node/9219>), one was hosted by the University of North Carolina, Chapel Hill (<http://ils.unc.edu/callee/emanuscripts-stewardship/index.html>), one on forensics was hosted by the University of Maryland (<http://mith.info/forensics/>), and one was hosted by the Library of Congress on home movies (<http://www.centerforhomemovies.org/homemoviesummit.html>), among others.

⁴ See http://britishlibrary.typepad.co.uk/digital_lives/2011/03/the-emss-lab-20.html

⁵ See <http://lib.stanford.edu/digital-forensics>

⁶ See <http://www.bodleian.ox.ac.uk/beam>

⁷ See <http://www.digitalpreservation.gov/formats/>

⁸ See <http://www.nla.gov.au/mediapedia/>

⁹ <http://www.paradigm.ac.uk/guidanceforcreators/guidance-for-creators-of-personal-digital-archives.pdf>

¹⁰ See <http://sites.google.com/site/stanfordluminaryarchives/>

¹¹ See <http://www.personal.psu.edu/esc10/blogs/E-Tech/2011/02/personal-digital-archiving-201.html>

This paper is in the public domain as a work created by a United States Federal employee.

¹ See <http://research.microsoft.com/en-us/projects/mylifebits/>.

² See <http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/>.

Week event. Library staff members were available to answer questions about physical and digital formats, ranging from the care of daguerreotypes to the digitization of audio and video tapes to the preservation of digital images. The first event was so popular that it was repeated at the 2010 National Book Festival, and again during the 2011 ALA Preservation Week. At every event Library staff members are asked for more extensive guidance, showing that there is a huge demand for advice in this area.

To accompany the events, the Library launched its “Personal Archiving: Preserving Your Digital Memories” site, providing downloadable and printable information sheets with tips for various formats.¹² The personal archiving pages are now among the most popular on the digitalpreservation.gov site.

4. CITIZEN DIGITAL ARCHIVING

Libraries are still developing strategies for archiving all digital collections, personal or organizational, born-digital or converted from paper. What all collecting organizations know is that they cannot do it all themselves.

The U.S. National Archives has not only issued a public call for citizen archivists to help explore the collections¹³, they hosted an event – “Are You In? Citizen Archivists, Crowdsourcing, and Open Government¹⁴ -- to encourage participation in all forms of citizen projects. The open government movement in the United States has inspired people to participate in the digitization and transcription of documents to increase their discoverability and use. The best known individual in this realm is Carl Malamud, whose non-profit publicresource.org¹⁵ is working toward the publication of public domain information from local, state, and federal government agencies. Among his best-known efforts are the digitization of over 550 government and publishing a 5 million page crawl of the Government Printing Office.

The New York Public Library is among the most recent organization to crowdsource a digital collection transcription effort, aiming to transcribe the dishes listed on 40,000 menus¹⁶. The earliest of these online efforts was likely the Fine Arts Museums of San Francisco Imagebase tagging project that began in 1997. One of the most extensive is the Flickr Commons, where more than fifty international organizations have shared images in order to crowdsource improvements to the metadata¹⁷.

As vital as these efforts are for the improvement of digital collections, it is citizen efforts in the identification and acquisition of collections that we must encourage and collaborate with. We are all familiar with the efforts of the Internet Archive¹⁸ which started out as a grassroots effort to archive the web, but now

provides a repository for all manner of personal digital collections, from texts to video and audio to software and games.

We are also familiar with the work of Rick Prelinger, and his efforts to build a film archive and a library books, periodicals, printed ephemera and government documents¹⁹. His interest in creating a collection of ephemera, from non-theatrical films to zines to maps and plans, from textbooks to maps to government documents, coupled with his opening his library to the public, hosting traveling public showings from his film archive, digitizing portions of the collection and sharing it with the Internet Archive and the Library of Congress, has inspired likely thousands of others to build their own collections, both physical and digital.

Perhaps the least well-known, outside a group of rabid supporters, is Jason Scott, the founder of the related initiatives [textfiles](http://textfiles.com)²⁰, [ArchiveTeam](http://www.archive-team.com)²¹, and [URLTeam](http://urlteam.com)²² among others. His cooperative, volunteer projects are an example of true grassroots digital archiving, stepping in to save marginalized collections, including bulletin boards and software manuals and web sites of imminent risk of shutdown. Scott is one of the most eloquent speakers on the risk of loss:

“A wonderful thing happened in the 1980s: Life started to go online. And as the world continues this trend, everyone finding themselves drawn online should know what happened before, to see where it all really started to come together and to know what went on, before it's forgotten.”²³

“Archive Team is a loose collective of rogue archivists, programmers, writers and loudmouths dedicated to saving our digital heritage. Since 2009 this variant force of nature has caught wind of shutdowns, shutoffs, mergers, and plain old deletions - and done our best to save the history before it's lost forever. Along the way, we've gotten attention, resistance, press and discussion, but most importantly, we've gotten the message out: *IT DOESN'T HAVE TO BE THIS WAY.*”²⁴

“Official” collecting institutions cannot afford to ignore the efforts of these citizen archivists, and, in fact, we must collaborate with them. A citizen digital archivist is often the most knowledgeable and motivated to save otherwise unrecognized at-risk materials. We need to encourage such efforts as well as recognizing them, incorporating them and their work into our distributed digital preservation community.

5. CONCLUSIONS

We need to recognize the importance of individual efforts in preservation. There is a clear role for the enthusiast in identifying what needs to be curated. For example, nineteenth-century collectors of books bound their collections for their use, and to share their collections with others which inadvertently contributed to future preservation. A more recent model is wikipedia, in which enthusiasts share their personal, curated knowledge. Another example is the work of amateur genealogists in sharing

¹² See <http://www.digitalpreservation.gov/you/>

¹³ See <http://blogs.archives.gov/online-public-access/?p=2661>

¹⁴ See <http://blogs.archives.gov/aotus/?p=2938>

¹⁵ See <https://public.resource.org/>

¹⁶ See <http://menus.nypl.org/>

¹⁷ See <http://www.flickr.com/commons?phpsessid=ea7b4da468f5935f24b65f41dbfc356f>

¹⁸ See <http://www.archive.org/>

¹⁹ See <http://prelingerlibrary.blogspot.com/>.

²⁰ <http://www.textfiles.com/>

²¹ http://www.archive-team.com/index.php?title=Main_Page

²² <http://urlteam.com/>

²³ <http://www.textfiles.com/statement.html>

²⁴ http://www.archive-team.com/index.php?title=Main_Page

their curated family histories. The personal role in preservation is significant--some people care enough to keep stuff alive; institutions may not, often because they are not aware of the importance of the content to one or more communities. Is it necessary for an institution to accomplish all digital preservation? We do not need to do it all ourselves, and we already recognize that we cannot. In some cases, our institutions may be in the way. We need to create a new sense of sensitivity at our organizations to grassroots efforts. There are individuals curating at the fringes of our communities, but embedded within their communities. Our organizations need to give them support, and help them with a preservation strategy, whether it is guidance to individuals or collaboration with collecting institutions.

In the same way that institutions have encouraged citizen archiving of physical collections, we must encourage citizen digital archiving. First, we must educate ourselves about personal archiving requirements, recognizing the widespread need for skills in the community to work with a wide variety of formats and technologies needed to export and retrieve content from its silos – email programs, online images sites, blogs, and social media – and from potentially obsolete media. We must then undertake personal digital archiving education in our local communities. Hold personal digital archiving events that provide instruction on digitization and training on the best ways to preserve born digital and digitized files. Introduce people to the need to archive the web and to the use of web archiving tools and services. Partner with public libraries, historical societies, genealogical groups, and local history museums. We should reach out to the communities that are archiving their specialized histories, providing that same archiving advice and expertise, encouraging them to steward their collections, and seeking out and accepting those collections when they become available to us. We must also reach out to the vendors and software developers that create the tools used across multiple communities, encouraging them to use more preservable format standards and build export functionality that allow more portable, personal control over personal digital content. In this way, we are encouraging better personal stewardship and improved preservation of what will certainly become our future collections.

There is a clear mandate that libraries, archives and museums must assume responsibility for educating the public about strategies for personal digital archiving and personal curation, and for exploring new approaches to processing, preservation, and access. Personal digital items initially have value to the individuals who generated them, but once those items are transferred to a collecting institution they will have a collective value to society. We also have a less well-established mandate to work with those citizen archivists who are taking the initiative to save large collections of digital ephemera and the at-risk output of underserved communities. Large amounts or small, all digital materials have the potential to be vital to cultural history studies in the future.

6. REFERENCES

- [1] Bowman, S. and Willis, C. 2003. *We Media: How Audiences are Shaping the Future of News and Information*. The Media Center at the American Press Institute, Reston, Virginia. <http://www.hypergene.net/wemedia/weblog.php>
- [2] Cox, Richard J. 2009. Digital Curation and the Citizen Archivist. *Digital Curation: Practice, Promises & Prospects*. pp. 102-109.
- [3] Gemmell, Jim, Gordon Bell and Roger Lueder. 2006. MyLifeBits: A Personal Database for Everything. *Communications of the ACM*, 49, 1 (January 2006), 88–95.
- [4] Gordon Bell and Jim Gemmell. 2007. A Digital Life, *Scientific American*, 296, 3 (March 2007), 58-65. http://www.scienceandsociety.org/web/Library_files/A.DigitalLife.pdf
- [5] Wilkinson, Alec. 2007. Remember This? A Project to Remember Everything We Do in Life. *The New Yorker*, (May 28, 2007). http://www.newyorker.com/reporting/2007/05/28/070528fa_fact_wilkinson
- [6] Bush, Vannevar. 1945. As We May Think. *The Atlantic Monthly*, 176, 1 (July 1945), 101-108. <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/>
- [7] Marshall, C.C., McCown, F., and Nelson, Michael L. 2007. Evaluating Personal Archiving Strategies for Internet-based Information. *Proceedings of Archiving 2007*. (Arlington, Virginia, May 21-24, 2007), Society for Imaging Science and Technology, 151-156. <http://research.microsoft.com/apps/pubs/default.aspx?id=63770>
- [8] Jones, William. 2004. Finders, keepers? The present and future perfect in support of personal information management". *First Monday*, 9 3 (March 1, 2004) <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1123/1043>
- [9] Teevan, Jaime, William Jones, and Benjamin B. Bederson. 2006. Personal information management. *Communications of the ACM*, 49, 1 (January 2006), 40-43.
- [10] Jones, W. 2007. Personal Information Management. *Annual Review of Information Science and Technology*, 41, 1 (2007), 453–504.
- [11] Jones, W. 2007. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. San Francisco, CA: Morgan Kaufmann.
- [12] Jones, W., & Teevan, J. 2007. *Personal Information Management*. Seattle, WA: University of Washington Press.
- [13] Marshall, C.C. Rethinking Personal Digital Archiving, Part 1: Four Challenges from the Field. *DLib Magazine*, 14, 3/4 (March/April 2008). <http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html>
- [14] Marshall, C.C. Rethinking Personal Digital Archiving, Part 2: Implications for Services, Applications, and Institutions. *DLib Magazine*, 14, 3/4 (March/April 2008). <http://www.dlib.org/dlib/march08/marshall/03marshall-pt2.html>
- [15] Beagrie, Neil. 2005. Plenty of Room at the Bottom? Personal Digital Libraries and Collections. *D-Lib Magazine*, 11, 6 (June 2005) <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>
- [16] Williams, Pete, Katrina Dean, Ian Rowlands and Jeremy Leighton John. 2008. Digital Lives: Report of Interviews

- with the Creators of Personal Digital Collections. *Ariadne* Issue 55 (30-April-2008).
<http://www.ariadne.ac.uk/issue55/williams-et-al/>
- [17] John, Jeremy Leighton. Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools. Paper presented at iPRES 2008: The Fifth International Conference on Preservation of Digital Objects, London, UK, September 29-30, 2008.
http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf
- [18] Paradigm project, *Workbook on Digital Private Papers*, 2005-7. <http://www.paradigm.ac.uk/workbook/>.
- [19] Kirschenbaum, Matthew G., Erika Farr, Kari M. Kraus, Naomi L. Nelson, Catherine Stollar Peters, Gabriela Redwine, and Doug Reside. 2009. *Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use*. White Paper. Washington, DC: National Endowment for the Humanities.
<http://www.neh.gov/ODH/Default.aspx?tabid=111&id=37>
- [20] John, Jeremy Leighton, Ian Rowlands, Peter Williams, and Katrina Dean. 2010. *Digital Lives: Personal Digital Archives for the 21st Century: An Initial Synthesis*. Digital Lives Research Project (3 March, 2010).
<http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf>
- [21] Matthew G. Kirschenbaum, Matthew G., Richard Ovenden, Gabriela Redwine, and Rachel Donahue. 2010. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington D.C.: Council on Library and Information Resources.
- [22] McDonough, Jerome P., et al. 2010. *Preserving Virtual Worlds: Final Report*.
<https://www.ideals.illinois.edu/handle/2142/17097>
- [23] Hafner, Katie. 2004. Even Digital Memories Can Fade. *New York Times* (November 10, 2004).
<http://www.nytimes.com/2004/11/10/technology/10archive.htm>
- [24] Hesseldahl, Arik. 2005. How To Preserve Photos For 500 Years. *Forbes* (April 14, 2005).
http://www.forbes.com/2005/04/14/cx_ah_0414photo.html
- [25] Cohen, Patricia. 2010. Fending Off Digital Decay, Bit by Bit. *The New York Times*. (March 15, 2010)
<http://www.nytimes.com/2010/03/16/books/16archive.html>
- [26] Fox, Stuart. 2010. Digital Age Presents New Problems for Historians. *Tech News Daily* (July 16, 2010).
<http://www.technewsdaily.com/information-overload-digital-age-presents-new-problems-for-historians-0867/>
- [27] Paul-Choudhury, Sumit. 2011. Digital legacy: Respecting the digital dead. *New Scientist* (May 6, 2011).
<http://www.newscientist.com/article/dn20445-digital-legacy-respecting-the-digital-dead.html>
- [28] Carroll, Evan and John Romano. 2010. *Your Digital Afterlife: When Facebook, Flickr and Twitter Are Your Estate, What's Your Legacy*. Berkeley, CA: New Riders Press. <http://www.yourdigitalafterlife.com/>
- [29] Walker, Rob. 2011. Cyberspace When You're Dead. *New York Times Magazine* (January 9, 2011).
<http://www.nytimes.com/2011/01/09/magazine/09Immortality-t.html>
- [30] Bell, Gordon. 2000. Dear Appy, how committed are you? Signed, lost and forgotten data. *Ubiquity* (February 1 - February 28, 2000).
<http://ubiquity.acm.org/article.cfm?id=334401>
- [31] Wright, Gary T. 2010. Preserving Your Family History Records Digitally. *FamilySearch*. (October 2010).
https://wiki.familysearch.org/en/White_Paper:_Preserving_Your_Family_History_Records_Digitally