

Heather Campbell

IS 586AO Digital Preservation (Robert DeCandido)

24 January 2019

Personal Digital Archives: Strategies for Educating the Public

In the first decade of the 21st century, information scientist Catherine C. Marshall carried out a series of studies regarding the preservation of personal digital belongings, a process known as personal digital archiving (PDA). In a two-part article based on her research, Marshall provides an in-depth consideration of the challenges involved in PDA, and she translates these challenges into implications for developers and information professionals, including librarians (“Part 1”; “Part 2”). By 2010, the Library of Congress (LC) had developed a simple strategy for encouraging the public to begin preserving their digital materials (“Notice”). Mike Ashenfelder, a writer and video producer for the LC, promoted the strategy in blog posts and in a keynote address at a 2012 conference on PDA, where he called on information professionals to collaborate and host educational events in order to further public awareness (“Signal”; *Communicating*). Although the LC’s strategy addresses several of the issues revealed by Marshall, its minimalist approach may be insufficient for long-term preservation. This essay attempts to synthesize the concepts outlined by the LC, Ashenfelder, and Marshall in order to determine whether the LC’s approach should be revised.

1. The Library of Congress’s Four-Step Plan

During 2000–2016, the Library of Congress advanced several initiatives to promote research, develop best practices, and educate the public about personal digital preservation (“Notice”). One of the results of these efforts was a short video, *Why Digital Preservation Is Important for You*, which encourages the general public to act now to prevent the loss of their

digital possessions, including personal photos, papers, music, and videos. In the video, the LC warns that these digital items are fragile, dependent on potentially obsolescent software and hardware, and at risk of becoming inaccessible. The video also provides practical instructions for taking action—a simple, four-step plan for managing a personal digital archive: “1) Identify what you want to save; 2) Decide what is most important to you; 3) Organize the content; and 4) Save copies in different places” (*Why* 00:07:07). The video describes an additional task—migrating the archive to a new storage location (preferably an external hard drive) every five to seven years—but does not list it in the four-step procedure (*Why* 00:05:00).

In digital preservation terms, the underlying strategies of the LC’s plan are *bitstream copying* (making exact duplicates of digital items and, typically, storing them separately from the originals) and *refreshing* (copying digital items “from one long-term storage medium to another of the same type, with no change whatsoever in the bitstream”) (Cornell University Library). Bitstream copying and refreshing are considered to be necessary parts of a digital preservation strategy; however, they are not by themselves “a complete program” (Cornell University Library). The LC’s video uses “migration” as a synonym for refreshing, but migration is actually a more robust preservation strategy, as it involves the conversion of data to a new technology—hardware or software (Cornell University Library). To be fair to the LC, such a strategy likely requires more instruction than is possible in a short video aimed at the general public. Moreover, migration has its own limitations, as it often does not result in exact copies (Digital Preservation Coalition).

2. Ashenfelder’s Keynote Address

Ashenfelder elaborates on the LC’s simple procedure in *Communicating Personal Digital Archiving to the General Public*, his keynote address at the 2012 Personal Digital Archiving

conference. Ashenfelder contends that institutions, while building a complex body of knowledge about digital preservation, have neglected to educate the general public about its importance. He argues that institutions must publicize the consequences of failing to preserve personal digital information: on a personal level, owners may have difficulty finding it or lose access to it entirely; on a societal level, some personal digital collections serve a social purpose or contribute to shared cultural heritage (00:04:40). In addition, Ashenfelder insists that institutions must offer simple, practical guidelines on how to create and maintain a personal digital archive—without overwhelming people with information. He cites the LC’s four-step plan as a good start but notes that it is still evolving; he also alludes to improvements in software and cloud storage that might automate and facilitate the digital preservation process (00:17:50). In the meantime, however, Ashenfelder calls on libraries and other cultural institutions to collaborate in public outreach activities that empower people to manage their digital belongings. In particular, he refers to the Personal Digital Archiving Day Kit designed by the LC, which augments the four-step procedure with guidelines for different media types, scanning tips, and information about the lifespan of storage media (Library of Congress, *Day Kit*).

3. Marshall’s Challenges and Implications

The LC’s simple, urgent digital preservation strategy, outlined by *Why Digital Preservation Is Important for You* and advocated by Ashenfelder, echoes the admonition of digital preservationist David S. H. Rosenthal to act now, even as technology is evolving: “Everyone—just go collect the bits” (00:56:30). However, Marshall raises concerns about this limited approach to digital archiving, an attitude that she summarizes thus:

If we make a minimal effort to maintain the digital assets that we have, the archiving problem will take care of itself. As long as storage capacities keep growing, we will just

take what we have with us each time we upgrade to new hardware. Moreover, even arcane file formats . . . are not cause for concern. Even if it is no longer in active use, a format as ubiquitous as JPEG will continue to be decodable into perpetuity. . . . We will have the bits and be able to locate decoders. What more will we need? (“Part 1”)

In the first part of “Rethinking Personal Digital Archiving,” Marshall identifies four problems with this minimalist approach: 1) digital stewardship, or curation,¹ is difficult, and people do not spend much time on it; 2) digital belongings often are “distributed among different stores,” making it hard to keep track of them; 3) the value of digital items is difficult to determine, especially as they continuously accumulate; and 4) retrieval from distributed, long-term storage requires special search strategies, particularly when several “not-quite-duplicate” copies exist (“Part 1”).

In the second part of her article, subtitled “Implications for Services, Applications, and Institutions,” Marshall clarifies that her aim is not to derail efforts to standardize file formats and develop robust repositories. Rather, based on the challenges she identifies in Part 1, she argues that “it is not enough to stop there” (“Part 2”). To “refocus the energies” of developers and technologists, she translates those challenges into four basic questions to be answered when considering personal digital archiving: “What should we keep?”; “Where should we put it?”; “How should we maintain it?”; and “How will we find it again?” (“Part 2”). In explaining the implications of these questions, Marshall suggests ideas for archiving services and applications that could be developed to address them. She also advocates partnerships among technology companies, publishers, libraries, and other institutions in order to foster communal maintenance of digital collections.

¹ Digital curation and digital stewardship are synonyms that refer to “the activities of making decisions to delete or preserve digital content” (Sinn et al. 223).

4. A Synthesized Approach

While their audiences are different, the Library of Congress (including Ashenfelder) and Marshall share an overarching goal: to promote and facilitate the preservation of individuals' digital materials. They also agree on the basic reasons that personal digital archiving is important: people want to save their digital belongings, and they want to be able to access them in the future. Moreover, although Marshall finds the centralization and periodic refreshing of personal digital files to be an inadequate preservation strategy, her four basic questions have parallels in the LC's public education strategy. Marshall's ideas could be used to expand the LC's recommendations, while maintaining the basic four-step structure.

Inventory. First, Marshall and the LC concur that individuals need to know where they have stored their personal digital items. The LC recommends the creation of a single "my archive" folder, to be copied and stored in multiple locations (*Why* 00:03:45). By contrast, Marshall views centralization as unrealistic; instead, she proposes that individuals maintain a catalog of digital items and their storage locations ("Part 2"). The LC provides somewhat similar advice in a handout designed for its Personal Digital Archiving Day Kit, in which it advises people to create a summary or inventory document that describes their archive and where items are stored (*Day Kit*).

Selection. Second, Marshall and the LC agree that selecting what to keep is a primary consideration. The LC simply advises individuals to "decide what is most important to you" (*Why* 00:02:33), but Marshall acknowledges the difficulty of this task—the "cognitive load of assessing value is enormous" ("Part 2"). Studies of public attitudes about personal digital curation have confirmed that individuals find it challenging and time-consuming to determine the value of digital files (Sinn et al. 229), and that they find it important to have a system of criteria

for the selection of files to be preserved (Mičunović et al. 117). Marshall proposes that digital archiving applications could make use of “value indicators”—intrinsic metadata about a digital item derived from its source, user actions, and disposition—to assess its relative value to a user (“Part 2”). Until such a tool is available, however, Sinn et al. suggest that information professionals provide appraisal and selection guidelines to the public (235).

Organization. Third, Marshall and the LC acknowledge that preserved digital items must be findable. As part of its “organize” step, the LC recommends creating subfolders to group items by media type and using consistent, descriptive file names (*Why* 00:03:51). Yet Marshall contends that “ordinary desktop search and file browsing will be insufficient to support long term access to personal materials” (“Part 2”). As a solution, she proposes various new modes of access, including “stable digital places, digital geographies, venues for re-encounter, visualization tools, tools for finding and choosing among duplicates, and application-independent viewers” (“Part 2”).

Tools supporting some of those access modes are in various stages of development. Stuff I’ve Seen (SIS) is a personal information retrieval tool that uses contextual cues—including time, author, and previews—to help people find items they have viewed in the past, but the prototype is not yet publicly available (Dumais et al.). Photos saved in the Apple Photos application are automatically presented in “curated collections called Memories,” based on “people, places, holidays, and more” (Apple). However, that functionality is available only within the Photos application; once images are exported, users need to know how to access the photo metadata that contains the original date and location information. At present, as recommended by the LC, most individuals must rely on descriptive file naming to promote discovery.

Maintenance. Fourth, Marshall and the LC agree that personal digital archives need to be maintained. The LC recommends periodic transfer of a personal archive to new storage technology, along with the creation of a backup copy (*Why* 00:05:35). Marshall concurs that refreshing of storage media is necessary, but she also mentions other necessary maintenance, such as malware checks and standardization of file formats (“Part 2”). In addition, because personal digital collections continuously grow and change, Marshall highlights the necessity of more complex maintenance activities that anticipate future use and provide an appropriate level of security (“Part 2”).

Later research has supported the soundness of these digital preservation strategies. A 2015 report on personal digital archiving, produced by the UK-based Digital Preservation Coalition (DPC), offers recommendations that parallel the four activities discussed above: determine the contents of your personal digital archive (*inventory*); prioritize what you want to preserve (*selection*); take good care of your archive (*organization*); and address key threats to your personal digital files (*maintenance*) (Redwine and Beagrie 15, 25, 17, 14). For each recommendation, the report gives practical suggestions in three tiers: “Quick Wins,” “More Effort,” and “Maximum Effort” (Redwine and Beagrie). Grouping suggestions by the effort they require is especially useful for information professionals tasked with providing instruction in PDA, as it helps them to tailor information for users with different abilities and needs.

5. Next Steps

After more than a decade spent advancing digital preservation initiatives, the Library of Congress announced that it would be retiring its *Digital Preservation* website as of 1 January 2019 and passing responsibility for its digital preservation programs to other organizations, including the Pratt Institute (“Notice”). This transition offers opportunities for information

professionals to update the LC's original message that the threat of information loss is the primary reason for individuals to begin personal digital archiving. Several studies have shown that most people do think that preserving their personal digital information is important, but their actual preservation practices do not always align with their attitudes (Marshall, "Part 1"; Sinn et al.; Mičunović et al.). Therefore, while the fear of loss will continue to convince some individuals to take action, other rhetorical appeals might be effective as well.

The DC Public Library (DCPL) Memory Lab lists daily organization, security, and legacy as reasons to care about personal digital archiving ("The Memory Lab"). In an era of data breaches and reports of the improper use of personal data—events that have prompted news articles on how to export personal data from social media sites (Chen)—security and control may be particularly convincing reasons for many people to begin PDA. Digital legacy is another area that requires increased awareness. In one study, the majority of participants had not considered what would happen to their digital records after they died (M. Mičunović et al. 125). The DPC's 2015 report also advises that individuals embarking on PDA should consider whether they wish to bequeath digital records or have certain files destroyed; its recommendations include "make long-range plans for your archive" (Redwine and Beagrie 13, 16). Comprehensive public outreach about digital preservation should at least mention digital estate planning, which a contributor to the LC blog *The Signal* deemed "an essential part of any personal archiving effort" (Bailey).

In addition to refining their message about the importance of personal digital preservation, information professionals must continue to expand practical educational offerings to the public. Building on the four-step strategy discussed above—inventory, selection, organization, and maintenance—they can update and tailor their services, depending on the

needs of their primary audience. The DCPL's Memory Lab is one successful example. Established in 2016 as part of a National Digital Stewardship Residency project, the lab offers digital preservation workshops, digitization information and equipment, and online resources that reflect the four steps, such as an inventory template and guidelines on metadata and storage ("The Memory Lab"; "Taking Care of Your Digital Stuff"). Patrons have provided positive feedback: "For many, digital preservation is a new issue and they are shocked by its value. Some patrons have wanted to attend multiple program sessions" (Mears). The success of the DCPL project led to the creation of the Memory Lab Network, a system of partner libraries that aims to "lower the barriers for libraries to learn about and consider adding digital preservation tools and programs to their own public services" (DC Public Library, "Memory Lab Network").

Although the Memory Lab Network is geared toward public libraries, it is important to note that public librarians are not the only informational professionals carrying out PDA education. Academic librarians have developed personal digital archiving workshops geared toward faculty (Brown) and even refocused their jobs toward helping users "effectively build, search, and organize their own personal and scholarly information collections" (Cahoy). Collaboratively produced resources, like the DPC's 2015 report, could be used to inform education programs at *any* institution (Redwine and Beagrie). As technology evolves and digital belongings continue to accumulate, all information professionals must be prepared to collaborate, as Ashenfelder recommended, in the design and delivery of educational programs that tackle what the DCPL deems "the complex and paralyzing problems of personal digital archiving" ("The Memory Lab").

6. Discussion Questions

1. Imagine that you are instructing the public about personal digital archiving. For each of the four preservation activities discussed in this essay (inventory, selection, organization, maintenance), what specific information would you provide? Would you emphasize one activity over others? Why? (For example, when describing the inventory step, how would you explain what digital belongings are? Would you add an explanation of copyright?)
2. When designing and delivering public education programs on personal digital archiving, how can information professionals account for different ages, backgrounds, motivations, and skill levels in their audiences?
3. In his keynote address, Ashenfelder said that the LC's PDA strategy needed a mnemonic device to help people remember the steps. What are some ideas for a memorable abbreviation or slogan, either for the LC's original steps (identify, decide, organize, save) or for the activities proposed in this essay (inventory, selection, organization, maintenance)?

Works Cited

- Apple. "Enjoy Your Memories in Photos." *Apple Support*, 16 Oct. 2018,
<https://support.apple.com/en-us/HT207023>.
- Ashenfelder, Mike. *Communicating Personal Digital Archiving to the General Public*.
<https://www.youtube.com/watch?v=5ySHVTivrQw>. Personal Digital Archiving 2012.
- Bailey, Jefferson. "The Big Digital Sleep." *The Signal*, 30 Jan. 2012,
<https://blogs.loc.gov/thesignal/2012/01/the-big-digital-sleep/>.
- Brown, Nathan. "Helping Members of the Community Manage Their Digital Lives: Developing a Personal Digital Archiving Workshop." *D-Lib Magazine*, vol. 21, no. 5/6, May 2015,
doi:10.1045/may2015-brown.
- Cahoy, Ellysa Stern. "The Challenge of Teaching Personal Archiving." *The Signal*, 9 Apr. 2012,
<https://blogs.loc.gov/thesignal/2012/04/the-challenge-of-teaching-personal-archiving/>.
- Chen, Brian X. "How to Delete Facebook and Instagram From Your Life Forever." *The New York Times*, 10 Oct. 2018,
<https://www.nytimes.com/2018/10/10/technology/personaltech/how-to-delete-facebook-instagram-account.html>.
- Cornell University Library. "Digital Preservation Strategies." *Digital Preservation Management*, 2003–2007, <http://dpworkshop.org/dpm-eng/terminology/strategies.html>.
- DC Public Library. "Memory Lab Network." *Memory Lab*,
<https://www.dclibrary.org/memorylabnetwork>. Accessed 23 Jan. 2019.
- . "Taking Care of Your Digital Stuff." *Memory Lab*,
<http://libguides.dclibrary.org/memorylab/resources/digital-stuff>. Accessed 20 Jan. 2019.

---. "The Memory Lab." *Memory Lab*,

<https://www.dclibrary.org/labs/memorylab#What%20is%20the%20Memory%20Lab?>

Accessed 20 Jan. 2019.

Digital Preservation Coalition. "Glossary." *Digital Preservation Handbook*, 2019,

<https://dpconline.org/handbook/glossary>.

Dumais, Susan, et al. "Stuff I've Seen: A System for Personal Information Retrieval and Re-

Use." *ACM SIGIR Forum*, vol. 49, no. 2, Jan. 2016, pp. 28–35,

doi:10.1145/2888422.2888425.

Library of Congress. "Notice; Background." *Digital Preservation*, <http://digitalpreservation.gov>.

Accessed 21 Jan. 2019.

---. *Personal Digital Archiving Day Kit*.

<http://digitalpreservation.gov/personalarchiving/padKit/index.html>. Accessed 20 Jan.

2019.

---. *Why Digital Preservation Is Important for You*. 2010,

<https://www.youtube.com/watch?v=VWkPufGDA6o>.

Marshall, Catherine C. "Rethinking Personal Digital Archiving, Part 1: Four Challenges from the

Field." *D-Lib Magazine*, vol. 14, no. 3/4, Mar. 2008, doi:10.1045/march2008-marshall-

pt1.

---. "Rethinking Personal Digital Archiving, Part 2: Implications for Services, Applications, and

Institutions." *D-Lib Magazine*, vol. 14, no. 3/4, Mar. 2008, doi:10.1045/march2008-

marshall-pt2.

Mears, Jaime. *DC Public Library National Digital Stewardship Residency Project Final Report:*

The Memory Lab. June 2016, <https://www.dclibrary.org/memorylabnetwork>.

Mičunović, Milijana, et al. "Data Organization and Preservation in the Context of Digital and Networked Media: Public's Attitudes, Habits and Practices in Relation to Digital Curation of Personal Digital Data." *Libellarium: Journal for the Research of Writing, Books, and Cultural Heritage Institutions*, vol. 9, no. 2, Mar. 2017, pp. 109–29, doi:10.15291/libellarium.v9i2.258.

"Mike Ashenfelder." *The Signal*, <https://blogs.loc.gov/thesignal/author/mashe/>. Accessed 23 Jan. 2019.

Redwine, Gabriela, and Neil Beagrie. *Personal Digital Archiving*. Digital Preservation Coalition, 14 Dec. 2015, doi:10.7207/twr15-01.

Rosenthal, David S. H. *CNI: How Are We Ensuring the Longevity of Digital Documents?* https://www.youtube.com/watch?v=h53DMtBUxsk&time_continue=42&app=desktop. Coalition for Networked Information (CNI) Spring 2009 Task Force Meeting, Minneapolis, MN.

Sinn, Donghee, et al. "Personal Digital Archiving: Influencing Factors and Challenges to Practices." *Library Hi Tech*, vol. 35, no. 2, June 2017, pp. 222–39, doi:10.1108/LHT-09-2016-0103.