CHAPTER 3

The Library of Congress and Personal Digital Archiving

Mike Ashenfelder
Library of Congress

Digital preservation is a familiar issue among the world's leading cultural institutions. But despite over a decade of institutional success in preserving digital files and collections, most of the general public—the largest group of digital-file stakeholders in the world—are unaware of what digital preservation or personal digital archiving is or why they should care.

Today, most people, young and old, have some sort of digital files to preserve; digital photographs are the most common. They need to know that their digital stuff is at risk of being lost unless they do something about it. The Library of Congress is trying to remedy that by

reaching out to partner with public libraries and other local institutions to teach the public about personal digital archives.

That is a tremendous task. Why should the Library be involved?

The Library is firmly rooted in digital preservation and has been building up its expertise since the mid-1990s with the American Memory Project. But the Library's true digital preservation work began in 2000 when it helped found the National Digital Information Infrastructure and Preservation Program. Its goal was and is to foster digital preservation research, collaboration, and standardization among government agencies, cultural institutions, and other stakeholders.

About 5 years ago, the Library recognized that many people were accumulating digital files (primarily photos, due to the proliferation of cameras on cellphones) but were not aware of the proper way to save these files. Even worse, most people were not aware of the potential threats to digital files: getting stuck on obsolete media, getting lost, or accidentally being deleted. To correct the situation, the Library developed information resources to help raise a general awareness about preserving personal digital material.

On the surface, it seemed like a daunting task to try to communicate everything the Library knows about digital preservation. But it became apparent that its institutional digital preservation knowledge could be distilled and simplified into a few steps and that information would be sufficient to help people get started. "Simplification" is a crucial strategy: Simplifying our information is essential in order to avoid losing an audience due to long-winded explanations with unnecessary technical detail. Too much information can bury a message.

We began narrowing down institutional digital preservation practices to find commonalities with basic consumer practices. Essentially, institutions and individuals use similar equipment (computers and storage media) and have similar stakes (loss of files). Institutions and individuals acquire content (acquisition), and they should organize it and name it in a way that will help them search and find it later (metadata). Finally, the content should be backed up.

Most people already know that they should back up their files. Like flossing or exercising, they know it is a good practice. But despite their good intentions, they may or may not actually do anything about it. It is usually a task put off until "later." The best thing that libraries can do is explain digital archiving, explain how easy it is to safely archive personal digital material, and stress the consequences of inaction and the threat of potential loss of valuable digital possessions.

Potential Loss

Our general explanation of digital loss goes something like this: We can "store and ignore" physical items such as books, paper photos, and documents under optimized conditions for years and expect that we can access them any time. The key word is *expect*. But "store and ignore" does not work with digital files such as audio, video, photos, and email because they are dependent on hardware and software to make them work. If either hardware or software is ignored for a significant length of time, it becomes obsolete, and the digital file will become difficult to access. It essentially becomes trapped.

Software makes the files accessible, and the hardware (the storage medium) is the container in which the files reside. Each storage medium has vulnerabilities and a limited lifespan. The coating on a CD can flake off. Floppy disks require computers built with disk drives that can read them. External hard drives can be dropped and damaged. In general, the life of storage media is cut short by at least three factors:

- Lack of durability
- Obsolescence
- Usage and handling (the more often a storage device is handled, the greater the possibility it will fail)

The whole arrangement is fragile. Not only are files placed at risk just by being transferred around, but the storage device itself needs to be kept in a stable environment with moderate temperature and humidity, and it needs to be protected from harmful elements. Paper

books and photos are also fragile; they should not be stored in a damp cellar, for instance, because of the potential damage from mildew and mold. Technology needs similar considerations regarding its own particular vulnerabilities.

Obsolescence is a fact of life in the digital age. As storage technology improves, previous generations of storage containers become obsolete. Even up-to-date computers may not have:

- Drives that can read older media
- Hardware connections that can attach to older media (or media drives)
- Device drivers that can recognize older media hardware
- Software that can read older files on media

And when stuff gets trapped on obsolete media, it is costly to get an expert to rescue it.

Lack of organization can also make finding files difficult. It is easy to lose track of files if they are scattered among websites, floppy disks, thumb drives, and CDs. Leaving content online—whether in the cloud, attached to emails, on social media sites, or on other online services—is risky because commercial services can and do go out of business, at which point the material stored on their sites could get deleted.

Despite these risks, the solution to mitigating the threat of digital loss is deceptively simple: Organize and back up your files. The Library broadened the solution to four best practices for personal digital archiving.

Personal Digital Archiving

The digital preservation steps that the Library delineated for the general public are scaled-down versions of the steps the Library takes with its own digital collections.

For instance, when we take in digital collections, those collections are assigned identifiers and metadata, moved to storage, and backed

up. Copies of the backups are kept in geographically different locations. The storage media is replaced every 5 years or so, and the data is transferred (migrated) to new storage media. Files are verified periodically, and their integrity is checked to make sure they are still intact.

So the approach to personal digital archiving that we recommend (and that is described in the sections that follow) is generally the same as for institutional collections: Locate everything to be saved, decide what to keep, organize the files, and save copies in different places and manage the collections.

Locate Everything to Be Saved

Files must be found before they can be backed up. Transfer them off of CDs, old floppy disks and storage media, the camera, and wherever else they might be. Download files from emails and social media sites. Then, gather them all into one place, such as a folder on a computer.

Why transfer files off of cameras? Given a digital camera's large storage capacity, it is common to let photos accumulate on the camera for months. What if the camera gets lost or stolen? And photos can be accidentally deleted from the camera. Get them off the camera or phone and save them with the other digital files to be backed up.

Decide What to Keep

Select the nicest ones, the ones worth keeping, and delete the rest. Does anyone really need 50 photos of clouds or 200 photos of autumn leaves? Blurry, unrecognizable photos? Delete them. Homework from 10 years ago? Delete it. Be decisive and thorough. Whittle the mass of photos down to the best, the "keepers." Toss out document drafts. Clear the clutter.

Organize the Files

There is no set system for organization; it just needs to be consistent and predictable to make it easy to find materials later. Descriptive folder and file names help. The descriptive names could be file types, with photos in a "Photo" folder and documents in a "Document"

folder, or by dates, with files named by month or year. Really, any system of organization that makes sense to you is fine.

Save Copies in Different Places and Manage the Collections

Most institutions replicate their digital collections in a separate geographic location far away from the source collection. In the event of a disaster, the distant, replicated collection will be safe, intact, and accessible, backed up on tape or spinning disk drives.

Similarly, personal files should be backed up in separate locations on at least two different types of storage devices. For example, save a copy on a backup drive and a copy on CD or on a flash drive or in online storage (the cloud). Diversity in storage formats is important because no storage device is 100 percent reliable.

Professional photographers who rely on digital files for their income have what they call the "3 - 2 - 1 rule":

- Make 3 copies.
- Save at least 2 onto different types of storage media.
- Save 1 in a different location from where you live.

And there is no set frequency of how often to back up—the more often the better, though.

As time passes and your storage device grows obsolete, the files that reside on that device become difficult to access. To help ensure ongoing access, move your collection to a new storage device every 5–7 years. When you buy a new computer, it is a good opportunity to buy a few new backup devices as well.

Taking an active role with personal digital collections will keep them safely preserved and accessible; ignore them and the collections may become inaccessible. It is that simple. Technology should, in time, automate the archiving process for us and make backing up less tedious. But until the backup and archive process is entirely automated, we need to care for it ourselves.

Another aspect of managing personal digital collections is estate planning. It is important to let a loved one know where important

documents reside and supply them with URLs and passwords, if needed. (See Chapter 5 by Evan Carroll for much more on this topic.)

Digital Photos Make Us All Stakeholders

Digital photos rank highest among materials most people want to save. Most cameras come with software and instructions for uploading photos to the computer, but many cellphones do not. Not every cellphone owner is aware of the issues associated with digital photo files, so the Library has created a few different resources and has given several presentations about aspects of preserving digital photographs. Although we try to avoid complicating the explanation with too much technical information, offering a little technical background enables people to better understand the potential threats to their photos and the consequences of their actions.

Beginning at the moment when someone takes a digital photo, three things happen almost immediately:

1. The camera saves the photo as a digital file.
2. The camera assigns a name to the file.
3. The camera inserts some technical information into the photo file about the conditions of the shot.

Most consumer cameras and camera phones save the file in a JPEG format, but the average person should not be concerned about formats unless he or she has an interest in working with a special format. JPEG is the widely accepted standard format, and most websites to which photos can be uploaded only accept JPEGs. When in doubt, we recommend leaving the file format as it is.

Renaming files with descriptive names can help you quickly identify photos later without having to display the contents. Renaming files will not affect the contents, the format, or state of the photo files.

Some photo editing software enables the addition of a description into the photo file, just as you would write a description on the back of a paper photo. This is standard practice for professional photographers

who add their copyright and contact information to each of their digital photos. This information can be displayed with the proper software. For a list of metadata resources, visit photometadata.org/META-Resources-Metadata-Links-and-Resources-Guide or search online for *embed description into digital photo*.

TIP

Warning! Some photo editing and photo management software programs store added descriptions into databases separate from the actual photos. Even though the descriptions are associated with individual photos, they do not actually get inserted into the photo files. This is important to know if you want your digital photos to carry extra identifying information when they are moved or viewed with other software. Also, if descriptions are added to photo files, the process of uploading those photos to social media sites and photo sites and as email attachments may strip out the descriptions. Be aware of the risks when you transfer photos. The Photo Metadata Project did a comparative test of several different social media resources (www.embeddedmetadata.org/social-media-test-procedure.php). For more information about the issues surrounding metadata stripping, see "Social Media Networks Stripping Data from Your Digital Photos" (blogs.loc.gov/digitalpreservation/2013/04/social-media-networks-stripping-data-from-your-digital-photos).

None of these steps so far has any effect on the state or quality of the photo itself. What if you want to modify a photo? Sharpen it or brighten it or enhance it? The first thing to remember is *never work with the original; always work with a copy*.

When you are ready to modify a copy of the photo, it helps to understand a little about the nature of digital photo files and what effect a modification might have on them. Digital photos are made up of millions of colored and shaded dots or squares called *pixels*. If an

image comprises 10 pixels on each horizontal line, and if the image measured 1″ wide, the measurement would be stated as 10 dots per inch (dpi) or 10 pixels per inch (ppi).

Resolution is directly related to the density of dpi/ppi. The more dots or pixels per inch a photo has, the higher the potential resolution of the photo, and the fewer the dots or pixels per inch, the lower the potential resolution of the photo. A larger palette of colors or shades of gray generates more pixels per inch on average, so images can be smoother, sharper, and delicately nuanced. A 50,000 dpi/ppi photo will look far nicer, with more subtle gradations, than a 10 dpi/ppi photo.

But the more pixels in a digital photo, the larger the file size will be, just because it has more information. And the larger the file size, the fewer photo files a camera card can hold. A camera menu may offer a choice between saving photos in a high quality (amounting to fewer photos) or lesser quality (amounting to more photos). That is a decision each individual needs to make.

Again, always preserve originals and work only with copies.

TIP

When uploading photos to certain websites, such as photo sharing or social media websites, the photos may get *compressed* to reduce the size of the file and to enable a quicker upload. This compression may be *lossy*, which means that some pixels are discarded in order to reduce the file size. Compression can also lower the quality of the photo. Once a photo is compressed and pixels are discarded, the pixels can never be restored or the original quality regained. However, compression is common, and it is not a bad thing if done right. Often a compressed image looks good enough for most webpages.

Scanning

This chapter so far has been about the challenges of archiving so-called born-digital photos. Scanning paper photographs into digital copies can also be a challenge, mainly due to unclear user manuals and word-of-mouth misinformation.

For pixel density, photography experts at the Library recommend scanning 4″ x 6″, 5″ x 7″, and 8″ x 10″ photos at 300 dpi/ppi. If the scan settings are larger than 300 dpi/ppi for average-size photos, the eye will probably not detect a significant increase in detail.

Photographic slides require a different dpi/ppi setting, due to their physically small size. Referring back to the earlier dpi/ppi explanation, a photographic slide has more photographic information packed into an inch than, say, a 5″ x 7″ paper photo of the same image. Therefore, it is best to scan slides at a higher dpi/ppi setting—about 1,800 dpi/ppi—to capture all of the image "information." The same is true for very small photos.

If the scanning software offers a choice of file formats for saving the scanned photo or slide, the best choices are TIFF or JPEG. TIFF is a lossless format, which means that the scanned image file will be uncompressed and data rich. It will also be a large file size. JPEG is a lossy, compressed format, though if the scanner settings offer a choice of quality (rather than specific dpi/ppi settings), selecting "best" or "highest" will retain the maximum amount of image data. JPEG files, being compressed, will be smaller than TIFFs.

Once the photos are scanned and digitized, treat the files the same as any other digital photos: File names can be changed, descriptions added, originals backed up, and copies modified.

Outreach

To get this digital preservation information out to local libraries and the general public, the Library created a personal digital archiving section on the digitalpreservation.gov website and populated it with instructional videos, downloadable brochures, and topic-oriented

pages. We wrote about how to archive the most common digital possessions: audio, video, photographs, email, documents, and websites. (In fact, most of them can be treated the same way: Organize them, back them up, and upgrade to new media about every 5–7 years.)

Some concepts are tricky to explain, so we created a primer to address topics such as scanning, how long storage media will last, how to transfer photos from your camera to your computer (not an uncommon question), and how to archive email.

Our videos (digitalpreservation.gov/multimedia/videos.html) are brief, and they clearly explain the basics of digital preservation and adding descriptions to photos. Podcast interviews (www.loc.gov/podcasts/digitalpreservation) with leading information technology experts help answer questions like how to get video off of (copyright-free) DVDs. (The Library, home of the U.S. Copyright Office, understandably defends copyright.)

And we learned that to communicate more effectively, we needed to make our terminology less complex and to avoid words and terms like *photometadata*, *digital objects*, *file formats*, *data*, and *records*. These are terminology speed bumps for the general public that only disrupt the flow of explanation. For example, instead of using a word like *metadata*, the word *description* is used instead.

In addition to publishing on our website, we also post digital preservation information to Facebook (www.facebook.com/digitalpreservation), Twitter (@ndiipp), and our blog (blogs.loc.gov/digitalpreservation). The videos and podcasts are added to the Library of Congress channel on iTunes and YouTube (bit.ly/11T9ONf).

Each year, during the American Library Association's (ALA) Preservation Week (usually held in April), we hold a Personal Archiving Day at the Library, a free event to which people can come and talk to the Library staff about preserving digital and physical items. We were also invited to speak about preserving digital photos at several events hosted by the Smithsonian Institute's National Museum of African American History and Culture. We also annually host panels at South by Southwest and staff tables at the National Book Festival.

After Library staff had worked at a few public events for personal archiving, we realized that no matter how much we publish on the web, there is no guarantee that the general public sees or understands any of it. The most effective means of communication is just what we do at the events: answer questions on the spot and directly educate the public.

By far the most effective event, in terms of reaching the most people, is the National Book Festival. For 2 days, we answer questions from a nonstop river of people, and we learn from our interactions with people exactly what we communicate effectively and what we need to clarify. Then we work on the information, create a new brochure, update a webpage or refine our text, and publish it again.

Something else we discovered at our tables at the National Book Festival is that obsolete media are powerful teaching aids. Each time we work the event, we load our table with an array of storage devices, current and obsolete, from punch cards to floppy disks, data tapes to laserdiscs. These artifacts make an impact on everyone who passes by. Older people are gleeful to see something from early in their careers, such as punch cards and zip drives, and parents explain to their fascinated little children what an 8″ floppy disk is. But in handling all of the "museum pieces," the point is driven home: All storage media becomes obsolete. That gives us an opening to talk about good stewardship practices for personal digital collections.

During Preservation Week 2012, staff from the Library partnered with public libraries to give presentations on personal digital archiving. These public library visits were not only outreach events, they were also aimed at establishing an ongoing relationship to provide and improve information resources for the public.

At each public library presentation, the audiences were grateful for the information, and the public library staff was welcoming and pleased to have the Library's involvement. Some of the public library staff came up with ideas of their own, such as asking people to donate copies of their personal files to a community repository at the library;

in that way, the public library could participate in the community's local history efforts.

Based on this success, we expanded our strategy to help local libraries with their outreach efforts by developing Personal Digital Archiving Day Kit (digitalpreservation.gov/personalarchiving), which includes a variety of resources to help users stage their own events.

Reaching a Vast Audience

Each April, the ALA hosts Preservation Week, a week of events—held online around the country—designed to spread awareness about the preservation of personal and shared collections, both physical and digital. Our colleagues asked us to do two webinars for Preservation Week in 2013, and so, on March 20, 2013, we presented a webinar titled, "Hosting a Personal Digital Archiving Day Event," which we designed mainly for library staff, and on April 24, 2013, we presented a webinar titled, "Personal Digital Archiving," which we designed for a general audience.

Both webinars were well-attended, and afterward, the webinar hosts sent us a list of questions the attendees had submitted. We sorted the questions by topic and used the questions and answers to help refine the focus of our information resources. Two-thirds of the questions fell into two topic areas: digital photos and storage media.

After Preservation Week, at the suggestion of the Public Library Association (PLA), we began writing a biweekly blog post for Publiclibrariesonline.org about personal digital archiving. Response to the posts has been overwhelming.

Our goal at the Library is to reach as many people as possible and to make sure that the concepts of digital preservation and personal digital archiving become common knowledge and part of everyday life. We need to communicate on a vast scale if we want to be effective in helping the basic ideas seep into the public consciousness.

Libraries still seem like the optimum place for community outreach. They exist to make information freely available to their communities,

and they will be around for many generations to come. Public libraries are increasingly involved in spreading digital literacy—facility with the internet and developing technology skills—into the communities they serve, and the knowledge of digital preservation and personal digital archiving is certainly part of digital literacy.

Until recently, our focus had been for our small team to encourage local libraries to start teaching about personal digital archiving and keep regular educational sessions going on their own, but we realized that was not a practical or realistic approach; at best, we might reach only several dozen people per year.

So, instead of working from the bottom up (i.e., beginning with our local public libraries), we decided to work from the top down. We approached the ALA to explore opportunities for collaboration. Within the ALA, the PLA and ALCTS were excited about collaboration. They were brimming with ideas.

We shared the goal of spreading awareness to local librarians about personal digital archiving and of its importance to individuals and the community in which they live. We brainstormed on ways to work with local public libraries.

There was a suggestion that a library might invite a local archivist or professional photographer to talk about digital photo preservation. They might invite a local oral historian or town historian to talk about personal contributions to local history and offer suggestions on how communities might add their own digital resources. During these events, librarians or library staff could include a PowerPoint presentation from our Library resources, or they could play one of our videos to lay the groundwork.

ALA, with support from the Institute of Museum and Library Services, has formed a partnership with StoryCorps, the organization that drives a mobile recording studio around the country, gathering oral histories from whoever wants to record them. The ALA news release (americanlibrariesmagazine.org/news/ala/ala-storycorps-receive-imls-grant-develop-library-outreach-program) stated: "Pilot libraries will receive equipment, training, promotional materials and other

resources to help them implement community documentation projects using the popular StoryCorps interview model. Local libraries will retain copies of all interviews, but preservation copies will also be deposited with the Library. The project team will produce freely shareable training materials to help public libraries better understand strategies for sustaining local oral history programs." This is yet another example of public libraries serving their communities in innovative ways.

We continue to explore collaboration with public libraries on ways to help spread information. We will keep the drumbeat steady and consistent to help ensure that the idea of personal archiving becomes an unremarkable part of life, second nature. We are firm in our conviction that people should have a basic knowledge of how to take care of their digital stuff.