

Personal Digital Archiving: Train-the-Trainer Workshop
Massachusetts State Historical Records Advisory Board
2019-06

Slide 1:

Hello and welcome to Personal Digital Archiving, a train-the trainer workshop

Slide 2:

The primary purpose of this course is to provide librarians, archivists and other cultural heritage professionals with the knowledge and tools to take this presentation back to their institution and hold their own Personal Digital Archiving workshops for their community – whether that is a workshop at a public library, a presentation to students at a university, or working with content creators at an archives.

Therefore, we will also take this opportunity to address some of the issues covered in this workshop a little more deeply than you might with a general audience. As we go through this slide deck, some additional content is added, but a streamlined version of the presentation slides are available from the Massachusetts State Archives at SHRAB@sec.state.ma.us.

Slide 3:

This presentation was originally created by The Society of Georgia Archivists, The Georgia Library Association, and the Atlanta Chapter of ARMA

Slide 4:

Originally it was presented as an online webinar by Oscar Gittemeier, Wendy Hagenmaier, and Michelle Kirk. They have made this available licensed under a [Creative Commons Attribution 4.0 International License](#). As long as you credit the creator, please *feel free to share, adapt, and reuse the materials*.

Slide 5:

This version of the workshop is brought to you by the Massachusetts State Historical Records Advisory Board through support from the National Historical Publications and Records Commission.

Slide 6:

The first topic that we will be covering is “The What and Why of Personal Digital Archiving.”

Slide 7:

We are going to start with a few definitions. Many of you may be affiliated with an institution that has a retention schedule. For the general public, retention is just “how long do I keep this

stuff?” Most people usually think about retention in terms of their taxes or other financial documentation, but really we can think about retention for aspects of our digital life. Whether it’s “Administrative Use” – keep it as long as you need it, a retention period set by legal or business requirements or Permanent retention – usually for sentimental or historical reasons we all deal with retention, we just don’t frame it that formally.

Now archiving is interesting – it’s a term that has been taken over by computer engineers. Our email asks us periodically if we want to Archive our mail; we may have an “archive” feature on an enterprise software program at work. What is meant in those situations is really just moving the files into a slower, less frequently accessed storage tier. It doesn’t mean that the system is doing all the things that we as archivists mean if we were archiving something – monitoring file types, creating and checking checksums and cataloging for continuing access.

There are myriad of reasons why should we care about archiving our personal records. We are documenting our lives and what is important to us and ensuring that some sort of legacy is passed on after we are gone. Our records show our place in the world and where we fit, how our lives intersect with the lives of others.

Slide 8:

What qualifies as a Personal Record?

Records are things constituting pieces of evidence about the past, especially an account of an act or occurrence kept in writing or some other permanent form (*Google Dictionary*)

When we look at records we want to distinguish between Personal vs. Business Records

An organization owns all its records that are created as evidence of its business transactions

Likewise, individuals own records of their personal business transacted, and any other records created for historical purposes

Slide 9:

This slide comes very much from a records management point of view. With my archives hat on, I would say that everything that is created documents something – an activity or an event or an occurrence. From a more records management point of view, there are some documents that are created that are so insubstantial or transitory in nature that they don’t rise to the level of being a record. Basically they are your administrative use documents. We can’t save everything and it is helpful to narrow down those things we need to hold for a specific amount of time and those things we want to keep permanently.

- Questions to ask yourself when determining if something is a record:
 - Might I need this to substantiate a claim?
 - Is there a legal or financial/tax reason why I should preserve this?
 - Does this have intrinsic or historical value which makes me want to keep it indefinitely?
- Make a list of the records
 - Group into records and non-records

- Determine how long to keep the records

Slide 10:

The privacy rights clearinghouse is a great resource for helping you determine how long to retain different types of records. Most of your financial records are likely to fall in the 7 year range while some legal documents you will want to keep indefinitely. Medical records may be stored in a system at your doctor's office or health system, but if you change health systems you should take your records with you.

Slide 11:

The records are deal with each day are increasingly born digital and most records can be preserved digitally instead of in paper format. However, it is important to understand that unlike paper records, long term digital records need special attention and monitoring. Paper records, kept in a dry, cool environment are stable while digital records are like perpetual toddlers, you need to keep an eye on them constantly. The closer to the point of creation that we can identify the retention of a born-digital records the better – knowing the disposition can influence the choices you make for the record.

Slide 12:

Next let's look a little closer at the landscape of personal digital records

Slide 13:

Digital records do not exist on their own. There are

- interactions between hardware and software
- interactions between software and files
- interactions among record creator, record steward, and record user
- So, it's important to understand the ecosystem of personal digital records

Slide 14:

The Personal Digital Records Ecosystem can exist locally, or increasingly offsite through cloud storage and web based applications. There is the hardware of our personal computers and phones, each running a different operating system that may or may not talk with a different device we use - all managing information and images that we create deliberately or even that are created by the system itself.

Slide 15:

A record can also be represented in many ways.

Slide 16:

For example, a cat video could be represented by an icon representing a moving image file, it is composed of an audio and visual track that are brought together in a what most people think of as a video in this case played through a web based program all of which is at it's core a binary series of zeros and ones.

Slide 17:

Next up we have Best Practices for creating Personal Digital Records

Slide 18:

Generally, we create records without considering how and when we will need them later. If you can't find or open your records, they are of no use to you.

Slide 19:

There are ways to increase the findability of your records. Providing additional information about the files, called metadata, can assist in locating the correct files later. If you are creating a textual file, utilizing a file format the supports full text searching can be helpful as well. A standard file naming structure, as well as an organizational structure, can help locate and identify where a file is and when it was created.

Slide 20:

Will you be able to open and use the record five years from now? How about 10 or 20 years from now? There are several factors the impact your ability to access your records in the future: the file formats they are created in, the media or storage environment and location they are kept in.

Slide 21:

At this point we are starting to get into some of the more technical aspects of digital preservation that you may not want to include in a presentation to the general public.

Digital records are created using some sort of computer, either on your desktop, in your phone, or in some other device. Most records are created in a desktop computer environment running an operating system. An operating system is a piece of software that tells your hardware how to interact with other software. Most likely the operating system you are working with is Microsoft Windows, but you might also be working with Mac OS or Linux. When working with digital records, it is important to know what operating system the files were created in because they might not work in a different system.

Another important piece of hardware in the digital environment is a server. Servers are specialized computers that carry out specific functions depending on the software that is installed on them. Some servers are configured to store records, some are configured to run database programs, and others support a variety of applications.

We may also use other physical devices such as external hard drives or portable media to store information, or we may utilize portable media to help transfer files from one system to another.

Slide 22:

Software is just a set of instructions that tells a computer what to do. Software organizes data and saves it in specific file formats. Software and file formats change over time, so it is important to understand what format your files are in and if they are accessible.

Slide 23:

If you are bringing in digital records into your institution, you may want to set up a policy for what types of files formats that will accept and what preservation activities you will engage in.

You may not be able to provide public access to all file formats that you receive or there may be conversation processes required to provide access to some file formats. This chart is focused on word processing, spreadsheet, and image files with a “catch-all” statement for other, more complex file types.

Slide 24:

Let’s take a different look at the digital environment. This is the Open Archival Information System reference model, more commonly known as the OAIS model. This model lays out what is going on with digital preservation. We work with content creators and producers to bring together Submission Information Package, the SIP, which is moved into a storage environment and repackaged as an Archival Information Package, AIP, which is finally presented to the user or consumer as a Dissemination Information Package, a DIP. There are lots of training courses and classes that provide information on the main activities of preservation, so I would like to focus on a few components of the model the don’t get quite as much attention.

Slide 25:

Specifically, we are going to briefly look at those components the encompass the whole digital preservation process: the administration and management for your program and preservation planning. We often get so caught up in the tools and the processes of doing stuff to digital objects that it is easy to forget that there are a lot of other factors that go into having a robust, sustainable digital preservation program. And we really are looking at a program – something that continues over time and brings together a wide range of people, processes and technology.

Slide 26:

Thankfully, there are an increasing number of models we can use to assess our current programs as well as systematically build for the future and enhance our capabilities. One such model is the Digital Preservation Capability Maturity Model (DPCMM) that was developed by Charles Dollar and Lori Ashley. As you can see from the diagram, the basic producer to consumer workflow continues through the model, but it is bounded by 15 components that help build your program.

Slide 27:

Additionally they have worked closely with the Council of State Archivists' State Electronic Records Initiative to translate these components into a framework. As you can see from the elements, there are several that don't deal directly with the digital objects at all. Even if you don't have a lot of technical expertise, there are areas like policy and strategy that you can work on to help grow your digital preservation program.

Slide 28:

On the CoSA website there are definitions of different levels of compliance or proficiency for each of the elements. For example, for Policy you can be at Level 0 where you do not have a written digital preservation policy all the way to Level 4 in which you have a digital preservation policy that is reviewed and updated at least every two years.

Slide 29:

Additionally, there are tips for how to move from one level to the next and links to resources that can help.

Slide 30:

In addition their work with CoSA, Dollar and Ashley have set up a free online interface for using the framework. You can log in and assess your program from 0-4 in the 15 component areas.

Slide 31:

Another tool for assessing your digital preservation program is the National Digital Stewardship Alliance levels of preservation. This model is focused more closely on the digital objects themselves, but again, it provides an approach where you can move your way up from one level to the next. I've yet to see an organization that can do everything at once, so it's important that we acknowledge that our programs aren't perfect, but we can take small steps to continuously make them better.

Slide 32:

You may also have heard of the Trustworthy Repositories Audit and Certification (TRAC) document and checklist. Here I want to highlight again, that section A has very little to do with

digital objects directly. We are looking at your whole institution and the development of a program.

Slide 33:

The components of TRAC have been more formally codified into ISO 16363: A Standard for Trusted Digital Repositories which again has a section looking at your organization as a whole, not just the digital objects and technology.

Slide 34:

Finally, I want to mention the work of Anne Kenney and Nancy McGovern which really focuses on thinking of digital preservation as more than just managing digital objects. The dpworkshop website provides some great context to this approach as well.

Slide 35:

Now let's transition back to more general topics such as ownership and copyright concerns of personal digital records that apply to a broader audience.

Slide 36:

Ownership and copyright don't always go hand-in-hand. Ownership includes owning the bits you make while copyright includes having the right to reproduce, distribute, perform, display or create derivatives of a digital records and the right to authorize others to do the same.

Slide 37:

Just because you own a digital file, it doesn't mean that you also own the copyright.

Slide 38:

When you are the creator of a digital file, be proactive about your rights. For example, consider applying a Creative Commons license to a file you are sharing. Also, the Digital Public Library of America has worked with other national library associations to provide the Rightsstatements.org website.

Slide 39:

If you are utilizing a third party platform to share or store your digital files, be sure to pay close attention to the terms of service and license agreements. Are you giving away rights or ownership to your content? This infographic it from 2014, so the specifics of this image will have changed, but it still serves as a reminder that we often give away ownership of our records without realizing it.

<https://www.news.gatech.edu/2014/04/18/do-you-read-terms-service-maybe-you-should>

Slide 40:

Some considerations include the software you use to create the file and if you are using software as a service for storing and sharing the file.

Slide 41:

We all encounter service agreements, and we all ignore them. In a study conducted in 2016, Jonathan A. Obar and Anne Oeldorf-Hirsch found the for following:

Slide 42:

- Average Reading time for Privacy Policies: 73 seconds (compared to expected time of 30 minutes for readers of average speed)
- Average Reading time for Terms of Service: 51 seconds (compared to expected time of 16 minutes for readers of average speed.
- 98% missed clauses about sharing data with the National Security Agency and employers and about providing a first-born child as payment for services

Slide 43:

Due to very high-profile cases of misuse of data and security breaches, privacy and security of personal digital records is an increasingly important issue for many people.

Slide 44:

Personal records have always contained personally identifying information, but the sheer quantity of digital records can make it harder to stay on top of where your information is and who has access to it.

Slide 45:

One way to protect your data is encryption where a piece of software converts the file to a format that is harder to read. The advantages include enhanced privacy and control, but you do insert another layer of software requirement. If you lose the encryption key or the software needed to decrypt the file becomes unavailable, you do run the risk of not being able to access the file in the future.

Slide 46:

Far more common than encryption are passwords. Research has shown the length can be a bigger deterrent than complexity. A Pass Phrase or a sentence can be more useful and easier to remember than a single word with lots of capital letters and special characters. And we all know we are supposed to use different passwords for each of our accounts, but it can be difficult to implement this practice. Using a password manager can be really helpful where you just need to remember one password and the software system remembers the rest.

Slide 47:

Always be aware of the environment you are in. You can have the most complex and secure password in the world, but if you are fall for a phishing or social engineering attack and give out your password, you are still breached.

Slide 48:

Now let's transition to best practices for storing personal digital records.

Slide 49:

In addition to storage that may be part of your computer or other device, storage options include cloud storage, external storage media and personal servers.

Slide 50:

Cloud storage is much more common than it was even a few years ago. Cloud storage is just storage on servers that someone else manages. If you are utilizing a music streaming service or uploading your files to Google Drive or iCloud, you are utilizing cloud storage.

Slide 51:

Many common cloud storage vendors provide a baseline of free storage with additional storage available for purchase.

Slide 52:

Some of the pros of cloud storage in that you can sync data across multiple devices and access data from any device and the data is being backed-up. It is also relatively inexpensive if you are dealing with a small amount of data. It does mean that you are giving up some control over your data. It is subject to security and privacy breaches. Additionally, there may be restrictions on the types and sizes of files that can be uploaded. Most importantly, companies and services can change, and services agreements can change as well. Make sure that you understand how to get your data out of a system before you put the data there to ensure that you don't lose access down the road.

Slide 53:

External storage is storing data outside your device.

Slide 54:

Examples include external hard drives and portable media devices like flash drives, DVDs, and CDs. Be careful with portable media – they can be lost, damaged, or sometimes are harder to monitor for preservation. Also, if you utilize external hard drives or portable media be sure you are storing it in a location different than your main storage. It doesn't help if you back up your data to an external hard drive and then have it sit next to your computer; if a disaster like a

house fire hits, you will lose both copies. Instead, bring it in to work with you or store it at a friend or family member's house.

Slide 55:

The pros of external storage devices is that you can continue to add on over time and they can serve as a backup for your data if your device is lost or damaged, and they may be cheaper than cloud storage if you are dealing with a lot of data. However, they can easily be lost or damaged and if you aren't checking them regularly it is possible that the data can rot or decay without your realizing the loss. Also, you need to make sure that you continue to have the hardware available to access the portal media. Some of you may be old enough to have a drawer full of floppy disks that you no longer have a drive for and even CD/DVD drives are not necessarily standard on a computer these days.

Slide 56:

Personal servers take the idea of external media to the next level. Personal servers are hardware and software that provide networked service and centralized access to data.

Slide 57:

Examples include a dedicated onsite server and virtual private server.

Slide 58:

Pros for a personal server include privacy, security and control of your data which you can centralize and access files remotely; cons include upfront costs, as well as there is the responsibility of managing software and hardware over time.

Slide 59:

Regardless of how and where you store your files, a concept that can help ensure that your files remain accessible over time is LOCKSS – Lots of Copies Keeps Stuff Safe. You can utilize a combination of the types of storage that we discussed above. There is additional complexity of managing and keeping track of multiple copies and keeping them in sync. But you have the security of replacing a file if one version will be corrupted and there is the convenience of having a file available locally while having geographic distribution of the copies.

Slide 60:

Our next section takes a look at best practices for access and ongoing management of personal digital records.

Slide 61:

When we look at ongoing management of digital files, we are looking at a records management model. There is the initial creation of the file and its original use. Ideally, we would then appraise the file, but most of us don't do that. We take tons of digital photos and don't bother

to weed out the bad one or we don't clean out old versions of documents. If we do select files that we want to save, a major consideration for continuing to be able to access these files is the file formats the files are saved in.

Slide 62:

Some file formats are more stable and easier to preserve. Generally, they are formats that are known as open formats that aren't owned by a particular company that keeps the underlying code closed. Open formats can also be opened with multiple software programs and across different operating systems. The chart here presents some examples a types of media, and what types of formats they are usually created in and what formats are easier to preserve, as well as the tool that can be used to create the preservation format. The Library of Congress also publishes a lot of guidance on what formats are better for long term digital preservation.

Slide 63:

Once you have selected your files to save, make sure that you are checking back in on your files. Digital files are perpetual toddlers – you need to constantly keep an eye on them. Your file formats may need to be transformed into newer formats in the future. Your hardware will definitely need to be upgraded over time. You need to be sure you can continue to find and access the files over time.

Slide 64:

Our final section deals with best practices for your digital afterlife.

Slide 65:

In the same way that we want to pass along our physical artifacts and mementos, we are going to want to pass along our digital photos, movies, and files. The Library of Congress has a brochure of recommendations for personal digital records. Here we see highlighted some of the topics that we have already covered like keeping track of your files and being thoughtful about your file naming convention. They also recommend multiple copies stored in different places and encourage renewing your media every 5 years, at the latest.

Slide 66:

There are an increasing number of resources available to help you plan your digital legacy. In addition to platform specific advice from services like Google and Facebook, there are associations and non-profits that are compiling advice and resource lists to help you think about managing your online presence after your passing.

Slide 67:

If you think your person records may have historical significance beyond your own family, the Society of American Archivists and the Council on Library and Information Resources provide guidance on preparing your records for an archival repository.

Slide 68:

Thank you for joining me for this review of Personal Digital Archiving best practices. Please feel free to reach out to the Massachusetts State Historical Records Advisory Board at SHRAB@sec.state.ma.us if you have any follow up questions. And thanks to the National Historical Publications and Records Commission for their support of the MA SHRAB and this presentation. Again, thank you and best of luck with your personal digital archiving.

Slide 69:

Photo credits