

Preserving Your Family History Records Digitally

Gary T. Wright

October 2010

While gathering family history records over the years, you've probably been preserving them physically. So why consider preserving them digitally now?

This paper discusses the benefits and challenges of using digital preservation to both augment and enhance the preservation of your family history records. It also explores solutions to the challenges, identifies what types of family history records are suitable for digital preservation, and summarizes what is required to get started archiving digital records.

WHY DIGITAL RECORDS?

There are compelling reasons for embracing digital records. To begin with, once a treasured but very fragile historical document has been captured digitally, you'll no longer worry about damaging or soiling it from handling, nor will you ever grimace again over a photograph's fading colors and diminishing contrast. You'll also find that the 999th digital copy you make is just as bright and sharp as the original. And printing records will be simpler than ever before.

Furthermore, you'll realize how much easier it is to organize, find, and use records by letting a computer do most of the clerical work for you. And you'll be delighted to see how easy it is to share your family history records with others. A digital record can be emailed anywhere in the world in just seconds. If the computer file that contains the record is too large to email, you can get it to your intended destination almost as fast by using a free Internet service.

Digital record storage has other advantages. For example, it doesn't require much physical space, so you won't need to consume your precious real estate storing bulky boxes and costly containers. And if you use the archival storage media recommended in this paper, you won't have to worry about factors like mold and humidity destroying your historical records.

Accessing digital records is extremely easy, and doesn't require climbing a rickety step ladder to hunt for dusty boxes in the attic or hobbling downstairs to a musty, dark basement and brushing away cobwebs to search for something you *hope* is there. In fact, all your family history records can literally be just a few clicks away—at any time of the day.

Perhaps more importantly, you're probably facing the need to preserve digital family history records anyway, since most photographs and documents are now being created digitally. "Born-digital" is a term coined for records that originate in a digital form, such as a personal history written with a computer or a family photograph taken with a digital camera.

While you can print a born-digital record in order to preserve it physically, doing so will require extra work, may be error prone, may not retain the look and feel of the born-digital original, and may not include the descriptive information that makes organizing, finding, and using digital records so much easier. Also, preserving physical versions of born-digital records will not allow you to realize the other significant benefits described previously. In fact, very few organizations attempt to preserve born-digital records by printing them and then archiving the physical printouts.

DIGITAL RECORD REQUIREMENTS

Creating and preserving digital records requires (i) a computer with appropriate software, (ii) the means to digitize physical records (i.e., convert them from paper or analog to digital), and (iii) the ability to preserve the digitized records for posterity and extended family. If you plan to share your digital records with others, you will also need access to the Internet.

Virtually all physical record types can be digitized—genealogies written in family Bibles, copies of vital records (birth, death, marriage), copies of census records, photographs, journals, written and typed documents, oral histories, newspaper clippings, family videos, handwritten or printed music, artwork, books, maps, etc. And all can be preserved digitally.

WHAT IS DIGITAL PRESERVATION?

It is NOT merely backing up your data! Rather, it is a process that involves storing digital records (i) with descriptive information (ii) for a very long time (iii) in multiple locations (iv) at the highest resolution you can afford; (v) periodically migrating the records to new storage media in order to prevent data loss or the inability to read the data; also to take advantage of new storage technology; (vi) changing file formats before they become obsolete; and (vii) providing access to your digital collection now and in the future.

While these steps may seem overwhelming at first glance, this paper will discuss tools and techniques that can help to make digital preservation a straightforward and enjoyable activity.

DIGITAL PRESERVATION CHALLENGES

Just as preservation of physical records has challenges, so does preservation of digital records, although the challenges are different.

You cannot feel or see digital information, which is written to a storage medium that a computer can interpret. Digital information may be characterized as a string of 1's and 0's where each digit is called a "bit." Digital bits are written in a predefined file format that computer software interprets so a digital record can be rendered on a computer screen or a printer.

Unfortunately, computer storage media decompose over time, causing bit loss—which in turn causes data loss. Depending on the type of record, loss of a single bit could result in the

corruption of a word (in a personal history, for example), or it may only cause a tiny flaw in an image that is not discernable to the human eye (in a photograph, for example).

The random, unpredictable nature of bit loss caused by storage media degradation can complicate management of a digital archive if not addressed.

Another challenge of digital preservation is obsolescence—both of file formats and storage technologies. Historically, new or enhanced storage technologies and file formats are introduced periodically to improve functionality and/or lower costs. As these enhancements are embraced by the computer industry, their adoption can cause issues for digital preservation. The reason is that vendor support is eventually withdrawn for storage technologies and file formats that become obsolete as newer formats and technologies gain popularity.

As with bit loss caused by storage media degradation, obsolescence can also complicate management of a digital archive if not addressed.

ARCHIVAL STORAGE MEDIA CHALLENGES AND SOLUTIONS

Industry experience with storage media used for digital preservation purposes has shown that hard drives and commodity optical discs (writable CDs and DVDs) have an archival life as short as three years—meaning that digital bit loss can occur in just three years!

Unfortunately, no such information is available yet for USB flash drives, which many people use to back up their digital family history records today. Although the archival life of flash drives is currently unknown, it is not expected to be significantly longer than hard drives, CDs, or DVDs. Also, be aware that USB flash drives can only support a certain number of write/erase cycles before their useful life ends.

An improvement to CD and DVD archival life is provided with archive-grade optical discs. One manufacturer of such discs is MAM-A.¹

The MAM-A archive-grade CD features a 24 karat gold reflective layer plus ultraviolet light protection to extend data life. Writing to archive-grade optical discs can be accomplished with a standard CD/DVD drive attached to your computer, and written discs can be read with virtually any standard drive.

Other manufacturers of archive-grade optical discs include Delkin, Mitsubishi, and Verbatim. These vendors claim that their products have an archival life measured in decades. However, the actual archival life you realize depends on several factors.

To ensure maximum archival life of archive-grade gold optical discs, you should write at half the top rated speed of your drive for the disc type, and you should never write more than 4.2 GB (4200 MB) to a DVD (since a DVD is susceptible to read problems on the outer tracks that are caused by handling).

Never store the discs where they are exposed to light. Also, you should never write anything on the top of your discs or put adhesive labels on them. The discs should be kept in jewel cases to protect against scratches, and the cases should be stored in a vertical position to avoid warping.

These guidelines exist because archive-grade optical discs contain an organic dye material that changes hue as a drive's laser writes individual bits (i.e., the disc is "burned"). Writing at slower speeds results in darker bits. Over time, the darker bit spots fade, making them more difficult to be read by a computer, and the fading accelerates with exposure to light. Therefore, both the speed of writing bits and disc storage conditions have a direct impact on actual archival life realized.

For more information about maximizing the archival life of optical discs, see reference [2].

The longest DVD archival life achievable today is provided with the new M•DISC—a revolutionary optical disc technology developed by Dr. Barry Lunt and Dr. Matt Linford of Brigham Young University and manufactured by Utah-based Millenniata, Inc.³ In effect, the M•DISC offers a synthetic stone-like material

instead of organic dye to record your family history records as digital bits. A high powered laser in the Millenniata-developed M•WRITER etches the bits into this synthetic stone, and they can be read by virtually any standard DVD drive.

Unlike the archive-grade optical discs discussed previously, M•DISCs have no special storage requirements to maximize archival data life. However, it only makes sense to store them vertically in jewel cases.

In 2009, the U.S. Naval Air Warfare Center Weapons Division at China Lake, California tested four different brands of archive-grade DVDs and one commodity DVD along with the M•DISC. The project was an accelerated aging test that evaluated disc stability after being exposed to elevated levels of light, heat, and humidity. A report of the testing results stated, "None of the Millenniata media suffered any data degradation at all. Every other brand tested showed large increases in data errors after the stress period. Many of the discs were so damaged that they could not be recognized as DVDs by the disc analyzer."⁴

Based on this and other internal testing, Millenniata claims that the M•DISC has an archival life measured in centuries (i.e., a millennium!). Clearly, the M•DISC represents a breakthrough in archival storage media.

With the archival storage media solutions presented here, you can now preserve your digital family history records for many years without worrying about random, unpredictable bit loss and data loss that complicate the management of a digital archive.

However, the prospect of CD/DVD reader obsolescence remains a potential issue that is addressed below.

ADDRESSING THE CHALLENGE OF OBSOLETE STORAGE TECHNOLOGIES

Reflecting on computer technology history, you might assume that the day will eventually come when CD and DVD drives are out of production. What will happen to your family history records written on archive-grade optical discs then?

Fortunately, the digital preservation industry has outlined two different approaches to address this situation. Let us consider them here.

The first is *device archiving*, which requires that you not only preserve your digital records, you must also preserve a means to both read and render the records in the future.

Taking the case of an M•DISC as an example, the device archiving approach requires that you purchase one or more DVD drives (which could be M•WRITERS) and archive them in a safe place for future use—i.e., after your current M•WRITER stops working and you can no longer acquire a DVD drive. The expectation is that you or your posterity will be able to continue to read and render your digital records with an archived drive long after the industry moves away from DVD technology.

The device archiving approach has some serious risks which may be obvious to you. First, if the archived DVD drive fails to work properly when finally needed, you won't have a solution after all! Second, what if your future computer doesn't support the archived DVD drive when you power it up? The only way to solve this problem is to also archive DVD device software that can be loaded and run on your future computer.

For most people, this is not a viable solution. Hence, device archiving is probably too risky to become your primary method for assuring future readability of your family history records.

The second approach to address obsolete storage technologies is referred to as *migration* in the digital preservation industry. Migration involves copying your family history digital records to a newer storage medium that is about to replace the technology you are currently using. This approach, also called *media refresh*, typically requires more work than the device archiving approach, but has fewer risks.

The most obvious risk is procrastination—i.e., putting off the migration (copying) work until you can no longer read the records stored on the older technology (for whatever reason).

Assuming the migration work is completed in a timely manner, media refresh provides a viable solution for providing access to your digital record collection in the future.

However, the migration work may have to be performed by your posterity or your extended family—since you may not outlive the ability to read the archive-grade storage media discussed in this paper. Therefore, it behooves you to prepare your posterity and extended family for such migrations.

There are two software tools available that can help with migrations. One is TeraCopy,⁵ a high speed data copier. The other, Unstoppable Copier,⁶ can help recover data from scratched CDs and DVDs. Both can be downloaded over the Internet free of charge for home use.

Media refresh also provides a means to prevent bit loss if you choose to store your digital family history records on a storage medium that is not archive-grade (assuming you complete the migration before any bit loss takes place).

Some words of encouragement and direction are in order here. First, the Blu-ray Disc Association has recommended that Blu-ray disc drives be capable of reading standard DVDs. While this recommendation is not compulsory, most (if not all) Blu-ray disc drives sold today provide DVD compatibility. It is reasonable to expect that Blu-ray drives will support the DVD format for years to come.

Perhaps more importantly, you have the opportunity to change the course of history regarding archival storage media readability!

Consider the chronicle of the long-playing (LP) phonograph record. It was introduced in 1931, gained tremendous popularity in the third quarter of the Twentieth Century, was superseded by the compact disc in 1982, and yet you can still buy needles and turntables today—nearly eighty years after its introduction. Why? Because a significant market for playing LP records exists today. To illustrate, a recent Google search on “LP record turntable” provided about 142,000 results!

As more and more people like you invest in archive-grade storage media, a market is being created for archive-grade storage readers that can and will buck the historical computer technology trend. This is why you have the opportunity to change the course of history regarding archival storage media readability. *Carpe diem!*

FILE FORMAT CHALLENGES AND SOLUTIONS

Another challenge of digital preservation has to do with file formats. Obsolescence is of particular concern.

To illustrate, in the early and mid-1980s, WordStar was the most popular DOS word processing software in the world.⁷ Today it is effectively “abandonware” (i.e., no longer developed or maintained). Anyone attempting to preserve a WordStar document in 1985 would undoubtedly have a difficult time getting his or her personal computer to render it today—even if the storage medium used at the time were still readable!

File size can also be of concern.

When preserving digital family history records, you should always preserve them at the highest resolution you can afford. The reason is that a digital record’s resolution quality cannot be improved once the digital record is created. And since you don’t know how a record will be used in the future (either by you or your posterity/extended family), resolution can become problematic. For example, if the record is to be printed, print quality will reflect resolution quality of the digital record when you archived it.

For photographs, TIFF (Tagged Image File Format) provides very high resolution, but it also creates large files that consume considerable amounts of archival storage capacity. Converting a TIFF file to the JPEG format will reduce the size of the file, but the reduction will come at the expense of resolution. That’s because JPEG processing does lossy compression of the digital bits, which means that many of them are discarded (as many as 90% or more) in order to

achieve a significant reduction in file size. Such JPEG images may be suitable for viewing on a website, but they may disappoint if you try to print them. (Note: JPEG stands for Joint Photographic Experts Group—originators of the JPEG standard.)

Fortunately, there are file formats that can help overcome the challenges described above.

The first is the PDF/A format (Portable Document Format for Archiving). Recognizing the impact of file format obsolescence on digital preservation, the International Organization for Standardization (ISO) defined in 2005 an “electronic document file format for long term preservation.” Based on the Adobe PDF 1.4 format, PDF/A provides a self-contained, self-describing format that is independent of external sources. For example, it embeds relevant fonts and color information with the content data so that future computer software will be able to render the document exactly as it can be rendered today. In effect, PDF/A uses a software archiving approach to digital preservation similar to device archiving described previously for obsolete storage technologies.

Combined with the archive-grade storage media described in this paper, PDF/A provides a breakthrough in personal archiving!

PDF/A can be used for most record types. Audio and video are exceptions. Also, PDF/A does not allow encryption and requires the use of standards-based metadata (i.e., descriptive information). Since fonts used in the document must be embedded with the content data, the resulting file size will be larger than a corresponding (regular) PDF file.

Nevertheless, PDF/A offers the promise of renderability well into the future.

For more information about PDF/A, see the REFERENCES section. Also note that a PDF/A file has the same file extension as a non-archival PDF file (i.e., .pdf)—therefore you cannot detect a PDF/A file without examining the metadata that describe it.

A partial list of PDF/A software is provided here—

- Adobe Acrobat
- soft Xpansion Perfect PDF Master (free for personal use)
- Nuance PDF Converter
- Solid PDF Creator

Another file format worth noting is JPEG 2000. Like PDF/A, it is also an ISO standard, although it applies strictly to images. As an improvement to the 1992 JPEG standard, JPEG 2000 provides both lossy and lossless compression. Lossless compression allows the exact original data to be reconstructed from the compressed data. And yet, lossless compression typically achieves 50% to 60% reduction in file size compared with source files—yet no quality of resolution is sacrificed in the conversion. For this reason, among others, JPEG 2000 is becoming popular in the digital preservation industry. File extensions for JPEG 2000 files are .jp2 and .j2k.

Combined with the archive-grade storage media described in this paper, JPEG 2000 provides a breakthrough in personal archiving of images by simultaneously delivering the benefits of high resolution and reasonable file size!

JPEG 2000 software is identified in the following incomplete list of products—

- Adobe Acrobat and Adobe Photoshop
- FastStone Image Viewer (free for personal use)
- XnView (free for personal use)
- ACDSee Photo Editor
- Corel PaintShop Photo Pro

There is a stream version of JPEG 2000 that is used for digital video. If your software offers this version, be sure to select the image version for still images.

The author successfully tested most of the software identified above for PDF/A and JPEG 2000 (the Adobe products were not tested). The following test results are worth noting—

- Using a 14 megabyte TIFF image as a source file, all tested JPEG 2000 products provided a lossless compression benefit of 61%.
- Solid PDF Converter Plus creates a PDF/A with JPEG 2000 lossless compression, thus combining the best of both formats. However, the author encountered a software bug with one test and reported it to Solid Documents. A commitment was received to fix the problem, but no time frame was given.
- soft Xpansion Perfect PDF Master (which is free for personal use) does not allow for the addition of descriptive information (metadata) when creating PDF/A files. You must purchase the business version of this product from soft Xpansion in order to get this capability, which is discussed below.

To preserve digital audio files, the Waveform Audio File Format (WAV) is recommended. Compatible with both Windows and Mac OS operating systems, WAV software is plentiful (search on “free WAV software”) and is expected to be used for many years to come. MP3 and SP2 should be avoided for preservation.

Likewise, digital video should be preserved using QuickTime or the Audio Video Interleave (AVI) format. QuickTime runs with both Windows and Mac OS operating systems, but AVI does not run with Mac OS. Flash, MPEG-2, and MPEG-4 should be avoided when preserving digital video.

ADDRESSING THE CHALLENGE OF OBSOLETE FILE FORMATS

The file format recommendations provided in this paper are intended to maximize the renderable life of your digital family history records. In the event that any of the file formats you use appear to be losing vendor support, you should promptly migrate the affected records by converting them to replacement file formats and writing the transformed files to archive-grade storage media. This type of migration is called a *transformation* in the digital preservation industry.

To illustrate, if a new JPEG format is introduced in the future that enhances the JPEG 2000 format, vendors will undoubtedly provide software to convert JPEG 2000 files to the new format. This software will be available for a number of years as customers gradually transition to the new format, providing a window of opportunity for you to migrate your affected images.

Since the converted images must be rewritten, such a migration might also provide a needed media refresh—which is an example of reducing overall preservation workload by combining or overlapping tasks.

In order to ensure that you can migrate digital records before their file formats become obsolete, it is critical that you stay abreast of digital preservation technology. If you don't, you and your posterity may have some unpleasant surprises in store.

Once again, this kind of migration work, as well as the technology monitoring it entails, may need to be performed by your posterity or extended family—since you may not outlive the file formats you choose to preserve your digital records. Therefore, it behooves you to prepare them for such transformation migrations and ongoing technology watching.

GETTING STARTED WITH DIGITAL PRESERVATION

Assuming you already have a personal computer, you can get started with digital preservation by purchasing a digitizing device, an archival storage device, and obtaining the software you want to create archival file formats such as PDF/A and JPEG 2000.

The digitizing device can be a scanner or a digital camera. Perhaps you will want both. Scanners are easier to use, but not as versatile as digital cameras.

When using a scanner, always scan at the highest resolution for which you can afford the archival storage capacity. At the very least, you should scan at 300 dpi (dots per inch) if you

never intend to print larger than the resulting digital record size. 1200 or greater dpi is recommended if you think you will ever want to print a larger version of the record. The scanning device you purchase will have software that allows you to set the desired dpi.

When using a digital camera to digitize a physical record, make sure you have natural, flat, uniform lighting so you can avoid shadows and reflections. Using a tripod is recommended, especially to keep the lens parallel to the record being photographed (camera lenses magnify skew if they are not parallel).

Most digital cameras allow you to choose a dpi setting, so always choose the highest setting available. Then, when you load your digital pictures to your computer, they will have maximum resolution when functioning as source files for the archival versions you create. Although these camera pictures will require significant capacity on your computer's hard disc when you load them, you can delete them once you have created archival files and written them to archive-grade storage media.

If you have analog audio recordings that you want to preserve digitally, you can purchase an audio digitizer. These USB devices can digitize virtually any type of analog audio signal so the recording can be archived in the WAV format.

To digitize your analog family movies, a professional service is recommended to minimize cost and maximize quality. The same applies to 35mm slides, although slide scanning attachments may be available for the scanner you purchase (but they may be pricey). For more information on scanning slides, see reference [8].

If using a service, be sure to specify which archival file format you want for the output (either AVI (.avi) or QuickTime (.mov) for digital video, and lossless JPEG 2000 for digitized slides). If JPEG 2000 is unavailable, you might ask for TIFF and then convert the files to JPEG 2000 yourself.

As explained in this paper, archive-grade optical discs are recommended for personal archiving. If your computer already has a DVD drive, you can purchase archive-grade, dye-based recordable discs from several manufacturers. If your computer does not have such a drive, any externally-attached name brand drive will do. And if you would rather archive on M•DISCs, contact Millenniata for your nearest distributor. To help pay for the M•WRITER, you might consider spreading the cost with family, friends, an organization, etc. so the M•WRITER can be shared. Remember that virtually any DVD drive can read an M•DISC.

ADDING DESCRIPTIVE INFORMATION TO RECORDS

Once you have the necessary software, digitizing equipment, and archival storage in place, you are ready to get started with digital preservation. However, before you attempt to preserve any records, it is important that you develop a plan to add descriptive information (called metadata in the digital preservation industry) to the digital records you are planning to preserve.

At a minimum, descriptive information should include both contextual and historical information. Contextual information describes what the record is—for example, a copy of someone’s death certificate, a photograph of a named person, etc. Contextual information also relates the record to its environment, throwing more light on the person(s) to whom the record applies. The more complete and descriptive contextual information is that you add to a digital record, the more valuable, interesting, and endeared the record will become—to you, your posterity, and your extended family.

Historical information provides the source of the record (for example, the county, city, town, or church archive from which a copy of a birth certificate was obtained). It should also identify the creator of the record, if such information can

be determined. This is important for copyright reasons, which are discussed below.

Your plan to add descriptive information to digital records should begin with file names. A file name can contain both contextual and historical information. For example, when the author scanned a photograph of a distant relative, the scanning software gave the output file a generated name of—

110237489853.tif

One would never know from this file name what the record actually is (other than a TIFF image). But by changing the file name to—

Photo of Esther Elizabeth Knight on her
wedding day 8 May 1917.tif

anyone looking at the file name will immediately know exactly what the record is. When searching the contents of an archival disc, having this much information for all the file names listed will certainly help you zero in on the object of your search very quickly!

A caution is in order here. Current personal computer operating systems have a limit of 256 characters to identify the location of a file on the computer’s hard disc (called the file path). These 256 characters include the file name as well as the names of all folders that must be opened to navigate to the file. Folder names may also be descriptive. Therefore, the more nesting of folders you use, the fewer characters will be left for the file name; and hence the fewer characters will be available for descriptive information in the file name.

In general, it is best to rename files with descriptive information when you first create or load them—otherwise, you may never get around to doing it.

In order to create a full set of descriptive information, you should also add reference information (or tags—another type of metadata) to files when you create them. Reference information allows search software to assist you in locating and accessing records.

When the author scanned file 110237489853.tif as explained above, he also added the following tags by clicking on the appropriate software option buttons—

Title: Esther Elizabeth Knight on her wedding day 8 May 1917
Subject: Esther Elizabeth Knight wedding photo
Author: in the public domain
Keywords: Esther, Elizabeth, Knight, wedding, 1917, bride, photo, public domain

Keyword tags are especially useful when searching for records using PerfectSearch (more on this below).

If tags are to be used effectively, both file creation software and search software must support such tags. It has already been pointed out that soft Xpansion Perfect PDF Master (which is free for personal use) does not allow the addition of tags when creating PDF/A files—you must purchase soft Xpansion’s business version of this product to get this capability.

COPYRIGHT CONSIDERATIONS

Any time you deal with records, make sure you adhere to copyright law in regards to copying, printing, and distribution. This applies whether you are working with digital records or physical records.

To avoid violations, track down the source or owner of each record (if possible), then apply applicable copyright law. A wonderfully clear and concise summary of copyright law as it pertains to genealogy has been written by Michael Patrick Goad.⁹ Please take time to study his short, well written article. Some key points from it are reproduced here—

- If an original work of authorship was created after 1977, it’s copyrighted and it’s going to be for a very long time. The earliest that any work created after that will lose its copyright will be about 2049 – that’s assuming that the author died right after he authored the work.

- If it was created before 1923, there is no copyright on it anymore, so long as it was published. If it wasn’t published, it may still be protected by copyright.
- Works published before March 1, 1989 without proper copyright notice are almost always in the public domain because, under the law that existed before that, a proper copyright notice was required for copyright protection.
- Works published from 1923 to 1963 had to be renewed after an initial copyright term for protection to continue. The U.S. Copyright Office estimates that over 90% of works eligible for renewal were never renewed.

ARCHIVING RECORDS

Before writing any records to an archive-grade optical disc, you will want to organize them so as to be as efficient in writing as possible. Archive-grade optical discs are designed to be permanent; therefore you cannot change anything after it is written. You can write the entire disc at one time, or you may write just a portion of it and add files later. In general, writing one record at a time is not practical.

The number of records (files) you can store on an optical disc depends on the disc type and the average size of the records you want to write, as shown here.

Storage Media Type	Number of 2.5 MB records that can be written	Number of 1 MB records that can be written
CD	260	650
DVD	1680	4200
Blu-ray	18,800	47,000

(MB means megabyte)

Please note that there are no archive-grade Blu-ray discs available currently.

To simplify writing, it is recommended that you first copy the target files to a temporary folder and monitor the size of the folder as you proceed. For Windows, this can be done by floating your cursor over the folder name—a pop up will display the total capacity of the folder. In general, you should not exceed a folder size of 650 MB if writing to a CD, or 4200 MB (4.2 GB) if writing to a DVD.

Once the temporary folder is populated with the target files, you can start the writing (i.e., burning or etching) process. If the folder size exceeds the disc's capacity, writing will stop when the disc is full, leaving all remaining files unwritten. Of course, maximizing the number of files written to each disc minimizes the number of discs required.

An important preservation principal developed at Stanford University is LOCKSS (Lots Of Copies Keep Stuff Safe). The basic concept is this—the more copies you archive in different locations, the safer your records will be.

To apply LOCKSS to your archive, you should write a minimum of two discs per set of files and store them in two different locations as far apart as practical. Writing three discs and storing them in three different locations is even better. Perhaps you can exchange archival discs with friends and/or family to enhance the safety of your archived data.

It's a good idea to periodically test your archival storage media by opening files randomly and examining the contents to detect errors. This should be done at least annually.

If errors are found on a disc, retrieve a copy of the disc (which is why you need to apply LOCKSS!) and determine if it is error free. If so, then you can replicate the copy and dispose of the flawed disc. If the copy is also flawed and you have no more copies to examine, then you have no choice but to test each file and copy the error-free files to new archival storage media. For those files with errors, you can recreate them if you still have the original physical records.

Of course, applying LOCKSS to your archive requires that you get organized and develop a process to track (i) locations of the archival storage media, (ii) media age, (iii) when the media should be tested next, and (iv) when a media refresh migration should be performed. Fortunately, there is an abundance of software available to help you do this, such as Microsoft Access or Intuit QuickBase.

SHARING YOUR DIGITAL RECORDS

As mentioned in the beginning of this paper, sharing a digital record with others is fast and easy—as long as you have an Internet connection and email services. The author uses Yahoo email (mail.yahoo.com) because it is free and offers unlimited storage capacity. Also, it allows you to attach a file as large as 20 MB to an email. However, whether or not someone can receive such a large file depends on their email capabilities.

Should you want to send someone a file that is larger than their email software will accept, you can use a free transfer service instead. TransferBigFiles.com and YouSendIt.com are two websites that allow you to transfer large files over the Internet at no charge. Once you upload a file that you want to transfer, a link is provided which you can then email to your intended recipient. That person need only click on the link in your email to download the file to his or her computer.

A side benefit of an email service that provides unlimited storage capacity is that it provides a means to extend the LOCKSS principle for your personal archive. By sending yourself emails with attached preservation files, you can create a collection of such emails that will be stored on the email service provider's computer infrastructure. In effect, you can backup your archive on this infrastructure.

You should never rely on this approach to be your primary or even secondary archive, however, since the email service provider could start limiting storage capacity at any time or could even go out of business. And organizing so many emails to function as your primary archive might be difficult. Also, you may have difficulty accessing your email inbox when you urgently need to retrieve a record from your digital archive.

AS TIME GOES BY . . .

. . . it is important that you, your posterity, and your extended family monitor technology changes and take appropriate actions as needed. These actions, which comprise the *ongoing* aspects of digital preservation, include—

- Transforming file formats that are becoming obsolete to their replacement formats.
- Copying files to newer archival storage media to prevent data loss (depending on the type of archival storage media used).
- Migrating files to newer storage media so they can continue to be read if existing storage technology is becoming obsolete.

Clearly, digital preservation is not a one-time activity, nor is it a single-generation project. Your responsibility in the digital preservation chain is to gather, digitize, and preserve records the very best you can, then pass them on to the next generation of your posterity and/or extended family that has been prepared to carry on the work.

In many respects, digital preservation is like a relay race—you carry the baton for a period of time and then pass it on to the next runner. To prevent the baton from being dropped during the handoff, you and the next runner must work together in perfect synchronization. This means preparing and motivating the next runner to carry on the race without missing a step.

As this process is carried on from one generation to the next, your digital family history records can be preserved in perpetuity.

Yes, it takes work—but the payback cannot be measured.

INTRODUCING A PERSONAL ARCHIVING INTERNET SERVICE

Olson DataMax Archiving,¹⁰ a distributor of Millenniata products, is scheduled to launch an Internet service by year end 2010 that has the potential to streamline and augment certain

digital archiving tasks. The service operates as follows—

Once you have prepared a set of digital files for preservation (with descriptive information included), you can upload them via the Internet at www.LegacyDox.com. This website is user friendly and allows you to browse and select the files you want to upload.

Olson DataMax employees will download the files and write them to an M•DISC. A set of files up to 4.2 GB (4200 MB) in aggregate capacity may be written to a single M•DISC. As the files are written, they will be indexed using PerfectSearch,¹¹ a full-text search tool. The index will then be written to the M•DISC along with PerfectSearch software. Finally, the disc will be sent to you for easy archiving and for easy access to your family history records. You will be able to specify how many copies of the M•DISC you want for LOCKSS purposes.

When you load the indexed M•DISC, a search user interface will automatically pop up on your computer screen, allowing you to immediately start searching for records stored on the M•DISC. Searches can be made by keyword, file name, content type, etc. Searching is where descriptive information takes on new importance—the more descriptive information included with the archived files, the more granular your searches can be. For example, searching on the keywords “Esther Knight bride photo” will promptly display the TIFF file discussed previously.

The author provided Olson DataMax with a collection of personal files that was 4.2 GB in aggregate capacity for the initial beta test of this service. Upon receiving the indexed disc, the author began to search its contents and quickly found that it contained 1849 images, two audio WAV files, 108 Microsoft Office files, and sixteen PDF files. Searching on keywords “exquisite bride” produced five hits (one image and four Word documents). After reviewing the results of the beta test, the author was delighted with the deliverables of the service.

PUTTING IT ALL TOGETHER

A lot of information has been presented in this paper. It may be helpful to summarize and supplement it here in a step-by-step outline.

Step 1- Once you have decided to start using digital preservation to augment and enhance physical preservation of your family history records, you should develop a scheme to organize your digital records.

Personal computers store digital records as computer files that must be named and placed in computer folders. Computer folders may be nested (i.e., cascaded). To retrieve a specific file or digital record, the computer must know the file path. A file path starts with the highest level folder and works its way down to the lowest level folder and finally to the file name. Here is an example of a Windows operating system file path—

\Personal Archiving\Copyright Fundamentals for Genealogy by Mike Goad.pdf

Personal Archiving is the folder name the author used when creating this folder, and *Copyright Fundamentals for Genealogy by Mike Goad* is the file name chosen by the author when downloading this file from the Internet. *.pdf* is the file extension that identifies the file's format (PDF or Portable Document Format in this case).

As pointed out previously, the length of a file path is limited to 256 characters. Part of your organization scheme should include a descriptive information scheme. Decide a scheme for naming both folders and files with descriptive information that will not exceed the 256 character limit. Also keep in mind that you can always rename a folder or file by right-clicking on the folder or file name and selecting "Rename."

There is no right or wrong organization scheme. Yours should be a scheme that is logical to you, but is also easily explained to and understandable by others who will carry on the preservation work after you.

Step 2- If you don't already have a personal computer, get one that has a DVD drive. You can use either a Windows-based computer or a

Mac. Make sure you are trained on your computer and its operating system so you know how to use it effectively. Discussion of operating systems is beyond the scope of this paper.

Step 3- It is likely that you will need an Internet connection for your computer. This should be available from your local phone service or cable service provider. Make sure you know how to use the Internet and do searches before you get started with digital preservation. Some popular search engines are Google, Yahoo, and Bing.

Step 4- Acquire the digitizing equipment you will need. You may already have a digital camera. If not, you may want to get one anyway because digital cameras are so much more versatile than film cameras (now might be the time to make the jump!).

Most of your digitizing will probably be done best with a scanner—make sure you get one that will support at least 1200 dpi and has a platen (i.e., flatbed) that will accommodate the maximum size of records you will be digitizing. You might want to hold off digitizing audio and video until you have some experience digitizing paper records.

All digitizing equipment has supporting software that must be installed on your computer. Usually a CD is provided that will automatically install the software with very little interaction from you.

Before attempting to use any digitizing equipment, set the resolution level of each device to a setting that will meet your needs, as discussed previously.

Step 5- Decide if you want to write your own archive-grade storage media or use a service instead, such as the Olson DataMax Personal Archiving Internet service mentioned previously.

If you want to use gold optical discs, the DVD drive in your computer should be able to write them—you need only purchase some gold discs to get started. The website for MAM-A is provided in reference [1].

To use M•DISCs that you write yourself, you can purchase a Millenniata Starting Kit that contains an M•WRITER, supporting software, and a number of M•DISCs. The websites for Millenniata and Olson DataMax are provided in references [3] and [10].

If you use a service, you may not need to purchase anything for writing optical discs (but you'll need to read them).

Step 6- Decide which file formats you will use for preservation. PDF/A should be used for many (if not most) records because of its ISO designation as an “electronic document file format for long term preservation.”

JPEG 2000 is attractive for still images because it simultaneously delivers high resolution and reasonable file size, and should be an industry standard for many years to come.

Unfortunately, some digitizing devices will only produce lossy JPEG. If this is the case with any of your digitizing devices (such as a digital camera), do not convert the JPEG to a JPEG 2000—doing so will actually reduce resolution quality! You may consider transforming the JPEG to a high resolution PDF/A instead.

If your device will produce TIFF, use TIFF files as your sources for JPEG 2000 creation.

You might hold off deciding on audio and video formats until you have some experience with PDF/A and (possibly) JPEG 2000.

Once you have decided on the file formats you want to use, you should then acquire the necessary software. Names of products and vendors were provided previously. Some software can be downloaded over the Internet after being charged to a credit card. Other software can be ordered directly from the vendor. Most products feature a 30 day free trial.

In general, the maxim “You get what you pay for” seems to apply when acquiring file format software, so beware of software that is free for personal use (although this type of software can help you get acquainted with a file format and a particular product before making a financial commitment).

Step 7- If you have any digital records already, such as digital family photographs, they will make excellent source files from which to try your first preservation actions. If you don't have any, take a few pictures with your digital camera, load them to your computer, and you will have a test collection ready to go.

First, review copyright considerations for the test records and eliminate from the collection any that will violate copyright law if you copy them for preservation reasons.

Next, set up the folder organization that you developed in *Step 1* using the descriptive information scheme that you also developed.

After that, select a file from the test collection that does not violate copyright law and copy it to the appropriate preservation folder.

If you drag the file to this folder rather than copy it, you will move the file by changing its file path. You may or may not want to do this. Sometimes it is good to keep the original file in its original folder so you can go back to it if anything goes wrong. If you drag the file to the new folder and have a problem, you cannot go back to the original. But the downside of copying is that it takes more disc space. As you get more experience working with digital files, you'll discover which approach works better for you.

After you have you copied the test file to its target preservation folder, decide if you want to change the file format to a preservation format (assuming this is possible). PDF/A software will allow you to transform a file such as a Word document or a JPEG image to the PDF/A format. Likewise, JPEG 2000 software will allow you to convert a TIFF image to the lossless JPEG 2000 format. Performing such transformations is where you get experience with preservation formats and your new software. Don't be afraid to try something—as long as you keep the original file around, you can always start over and delete the previous unsuccessful attempt. And remember that you can always rename a folder or file at any time.

Whether or not you transform your test file to a preservation format, the next step is to enrich the file with descriptive information. As explained previously, this involves renaming and tagging the test file with contextual, historical, and reference metadata.

Step 8- After you have created a number of test files, you can test writing them to the archive-grade storage medium of your choice.

Using a service such as the Olson DataMax service mentioned previously involves merely uploading the files through the Internet and paying for the service.

Writing your own discs involves a bit more work. In this case, it is recommended that you copy (not drag or move) target preservation folders and/or files to a temporary folder while monitoring the size of that folder. When you have filled the temporary folder with the folders and/or files you want to test, launch your DVD writer software and point it to the temporary folder. When the writing process is complete, you will have your first preservation disc!

Test this disc by reading it and studying what you have just created. You may or may not want to make changes as a result of your testing.

Step 9- Once you feel that you have mastered the preservation activities discussed so far, you are ready to start digital preservation in earnest. At this point it is advisable to develop a plan for completing the work for all the physical records that you want to preserve digitally. Then it is a matter of following this plan to completion.

One aspect of this plan should address LOCKSS. As stated previously, you should make at least two copies of every preservation disc and store them in different locations. Creating three discs and storing them in three different locations is recommended. You may want to exchange archival discs with a friend or relative to maximize the safety of your newly archived data. As Steve Olson¹⁰ put it, *“Don’t play Russian Roulette with your memories!”*

Step 10- When your plan is completed, you are not finished with digital preservation. As discussed previously, you should also develop a process to track (i) locations of your archival storage media, (ii) media age, (iii) when the media should be tested next, and (iv) when a media refresh migration should be performed. The process should include a method for reminding yourself when these actions need to be performed.

You must also stay abreast of digital preservation technology so you can react appropriately as discussed in this paper.

Step 11- Since digital preservation is neither a one-time project nor a single-generation responsibility, your final task will be to prepare willing family members and/or willing posterity to carry on the work you have started. This means preparing and motivating these people to take the baton of responsibility from you when the time comes without missing a step.

FINAL THOUGHTS

Digital preservation should not replace physical preservation of your family history records. Rather, it should be used to augment and enhance physical preservation. This paper has described numerous ways how digital preservation can do this.

Preserving your family history records digitally can be both enjoyable and satisfying. Get started now—don’t put it off. Encourage your friends and family to get started too. As you do the labor of love required to preserve these records digitally, you can expect to have an uplifting spiritual experience as you connect with your memories and your ancestors. And best of all, you’ll be able to sleep better at night!

REFERENCES

Recommended PDF/A links—

www.aiim.org/documents/standards/19005-1_FAQ.PDF (*frequently asked questions*)

www.adobe.com/enterprise/pdfs/pdfarchiving.pdf (*white paper*)

www.pdfa.org/doku.php?id=pdfa:en (*PDF/A Competence Center*)

¹ www.mam-a.com/longevity_research

² *Care and Handling of CDs and DVDs: A Guide for Librarians and Archivists* (by Fred R. Byers Oct 2003)

www.clir.org/pubs/reports/pub121/contents.html

³ www.millenniata.com

⁴ www.millenniata.com/orphan/china_lake_summary

⁵ www.codesector.com/teracopy.php

⁶ download.cnet.com/Roadkil-s-Unstoppable-Copier/3000-2242_4-10580640.html

⁷ *Wikipedia, the Free Encyclopedia*

⁸ www.flatbed-scanner-review.org/35mm_slide_film_scanners/scanning_35mm_slides.html

Also see—

www.abstractconcreteworks.com/essays/scanning/Backlighter.html

⁹ *Copyright Fundamentals for Genealogy* (by Michael Patrick Goad 29 July 2003)

www.pddoc.com/copyright/genealogy_copyright_fundamentals.htm

Also see— www.pddoc.com/copyright

¹⁰ www.olsondatamax.com

¹¹ www.perfectsearchcorp.com/index.html