Managing Microfilm Collections in the Age of Digital Imaging Technology

Wayne J. Metcalfe
Elizabeth Hallam Smith

Managing Microfilm Collections in the Age of Digital Imaging

Presented By
The Genealogical Society of Utah

- A Brief History

The Church of Jesus Christ of Latter Day Saints established the Genealogical Society of Utah in 1894. The early Articles of Association listed among other things the benevolent goal of establishing and maintaining a genealogical library for the benefit of the society members and others. An educational goal was envisioned of sharing information that related to genealogical matters. The original library contained 3 printed volumes. From these humble beginnings the collection of the Genealogical Society of Utah has grown to one of the largest genealogical collections in the world with over 300,000 titles in its collection.

In 1938 the GSU began the newly commercialized process of microfilming original records that contained genealogical information. As an early pioneer of microfilm, the GSU helped establish this technology as a means of disseminating genealogical information to the masses. Today the GSU has over 2,300,000 rolls of microfilm and 740,000 microfiche in its collection. Microfilming is currently being conducted in over 200 projects worldwide.

The extensive collection of microfilm includes records from 110 countries in the world. These microfilm collections include records of genealogical significance from government, religious and private collections.

Several years ago the GSU began developing digital imaging systems to image and disseminate genealogical information. The GSU is actively converting many of its microfilm rolls to digital images. Today the GSU is pursuing digital imaging projects around the world. The GSU has over 15 active digital projects and many more are planned for the remainder of this year and the future. By 2006 the GSU should have over 80 digital cameras imaging documents around the world.

- Microfilm vs. Digital Images for End Users

As digital imaging matures, many archivists will be faced with the dilemma of how to address the coming digital age. Decisions must be made whether to maintain a proven microfilm technology that is accepted by the archival community or to embrace a new imaging technology that appears to be inevitable but which has no proven preservation methodology. Let's take a few minutes to examine the advantages of microfilm technology vs. digital technology.

Because microfilm is an actual photograph of the original record, camera capture is relatively straightforward and low-tech. This type of analog capture allows for easy storage and retrieval. Simple magnification of the human-viewable microfilm image is all that is necessary to retrieve the original content. No sophisticated equipment is necessary to view the microfilm. Silver-based microfilm, properly exposed, developed and stored is a long-term preservation medium. Life expectancies of 500 years are generally accepted as feasible. Most archivists accept microfilm as an excellent medium for preservation. It provides a relatively inexpensive means of preserving precious documents.

Inherent with any technology are disadvantages when compared to other technologies. Microfilm has several drawbacks. The analog nature of microfilm makes indexing of individual frames difficult. Even if the microfilm is marked so that each image is identified, the correct roll or fiche usually must be retrieved by hand and loaded onto a machine for viewing. It becomes difficult to coordinate the index search with the retrieval of the individual image without requiring manual searches of rolls and fiche.

Microfilm can only be searched by one individual, on one machine, at one time. High demand microfilm must either be duplicated so that several copies exist or only one person will be able to use the microfilmed documents at a time. Multiple copies of microfilm require additional storage space.

Because patrons have direct access to the microfilm, damage can occur to the microfilm. Fingerprints and dirt can damage the microfilm and reduce image quality. Improperly maintained equipment can scratch, tear, wrinkle or scorch the microfilm. For every roll damaged a new copy must be made from the master microfilm roll. Each duplication further damages and degrades the master copy's quality.

Microfilm has a limited tonal range that can render some documents unreadable. One of the characteristics of microfilming is to increase the contrast of documents that are filmed. Contrast build-up can render some documents nearly unreadable. This contrast build-up can be particularly difficult for documents that have a limited contrast range between text and background densities.

There can be some concern about the environmental impact of microfilm developing and reproduction processes. Some governments carefully monitor the effluence discharged from the microfilm developing process. Special environmental considerations may need to be addressed in order to develop microfilm in certain areas. The chemicals used to develop the microfilm, though generally considered to having minor environmental impact, have raised some concerns.

With the advent of digital imaging several distinct advantages to conventional microfilm have been identified. Starting with the capture of the digital images, the camera operator is allowed to view each image after capture. This evaluation allows the operator to make adjustments to create the best image capture possible. Exposure, light balance, reduction and all other aspects of the image capture can be modified to render the best image possible.

Digital images can be manipulated in many different ways. The file formats can be changed. The files can be compressed to save storage space. The images can be enhanced to improve readability. Security watermarks and encoding can be added to the images. Images of good quality can be converted from color or grayscale to bitonal images. These changes can be done individually or by batch processing. Digital images that are in color can be changed to grayscale or bitonal images. This conversion is a method of saving space when a document of less bit depth may still convey the information that the higher bit-depth document contained.

The digital images can be stored on a single computer and made available to a single user or the image files can reside on a computer network. The computer network can deliver the images via a local area network, an intranet or the Internet. Using the security features of most computer operating systems access to the images can be broadly given or restricted as needed by the archive.

With the use of metadata, data about data, digital images can be accessed through integrated databases. These databases can allow direct access to the images. The metadata can provide information about the images, their content, technical information about the image and its capture, cataloging information and other administrative information as needed.

Digital images can be encoded with security measures that will allow detection if the image has been modified. Digital watermarks can be inserted on the image so that the image is readable but easily identified as an image that belongs to a specific collection or carries a certain restriction. Because digital images are delivered via electronic means, the user can be prevented from modifying or damaging the original image.

Digital images whether viewed one time or hundreds of times, maintain the same quality. Software can even be provided that allows the user to view and manipulate the image to match their viewing preferences. Such manipulation does not change the original file.

Digital technology also provides for access in remote locations. It will no longer be required that a user visit a specific facility to view images, but they will be able to view images in the comfort of their own home or from their work place.

Despite these advantages, digital imaging has several disadvantages compared to microfilming. New technologies always have higher costs than those technologies that are more mature. Digital imaging equipment tends to be more expensive than microfilming equipment. The digital files must be refreshed to new media regularly. As computer hardware and software change, the images must be migrated to new formats. The constant vigilance requires more financial dedication to digital preservation than microfilm requires. The storage and preservation costs of digital can be ten times or more the cost of microfilm storage and preservation.

The physical media that the digital images are stored on requires constant monitoring. The media that digital images are stored on must be regularly checked for errors and failure rates. The images must be migrated to new storage media as media changes. As file formats change and new standard file formats are devised, the images may need to be converted to the newer formats. Every technology change must be reviewed and its impact on the image collection considered. Missing one generation of migration can spell disaster for preserving access to the digital collection in the future.

Digital images are not human-viewable by themselves. Digital images require computer equipment to translate the files into human-viewable form. This whole process of conversion is technology based. As time progresses so does the technology. Technology advances require that digital images must be managed to account for the changes in technology.

- Microfilm Conversion Methodology

Many archives have collections of microfilm. For many years microfilm has been one of the preferred mediums used to preserve and distribute documents. However, digital images have been shown to be advantageous in delivering information to the user. Converting microfilms to digital images provides the archivists and patrons with several distinct advantages over the delivery of the images via microfilm.

Digital images can be delivered to the user without the damage that accompanies the use of microfilm when delivering the same images. Patrons cannot damage a digital image. The digital images can be viewed on computer workstations that can also serve other archival functions. This dual-purpose use of equipment can further justify the expenditure of money to convert to digital imaging. Microfilm viewing equipment is generally used for only one single purpose. With a properly configured computer system, multiple users can access the same material at the same time. Those collections that contain high usage materials can more easily be delivered to multiple users simultaneously.

Indices to the documents can be created that allow for direct access to the image that relates to the indexed entries. These indices allow for greater penetration into the document collections. Microfilm that may have previously presented a challenge for user access and use may now be more accessible by digital conversion. With appropriate software the patron can enhance or manipulate the image to increase readability without affecting the original image file. Such enhancement improves the user experience with the documents. Security measures can be implemented to prevent unauthorized access and copying of the image. Watermarks can be applied to the images and copies to show a restricted use of the document.

Microfilm can act as the preservation copy and the scanned digital image can be used as the distribution copy of the images. The microfilm if properly exposed, processed and stored could act as the preservation copy of the images. The scanned digital images could be converted to a distribution file format and used as the copy that would be distributed to users. Extraordinary efforts to preserve the digital images could be ignored. If catastrophic loss of the digital images occurred, the microfilm could act as the preservation copy of the images and the microfilm could be rescanned. This method would allow for the best benefits of microfilm and digital images. The microfilms superior preservation costs and longevity and the access and flexibility of digital image delivery would both be realized.

While microfilm conversion to digital has many virtues there are some drawbacks that must be considered. Digital images created from microfilm will have all of the imperfections of the microfilm they are scanned from. Some image enhancement may be possible but poor microfilm will produce poor digital images. Additionally, files sizes for digital images scanned from microfilm can be quite large as the quality of the digital image increases. Files sizes in the range of 20 – 40 MB for TIFF uncompressed images are common.

Several considerations must be addressed when implementing a microfilm to digital conversion plan. A carefully planned microfilm to digital conversion can spell success for an archive and its patrons.

Microfilm scanning equipment is available for purchase from several companies around the world to facilitate the transfer of microfilm images to digital. Most companies that manufacture microfilm equipment also manufacture microfilm scanners. This equipment can cost from the low end of USD $5,000 to over USD

$75,000. The scanning equipment generally will need to be connected to a computer system. This computer equipment must be able to handle the large amounts of data that will flow from the scanner. Slower computer equipment will mean slower production time. Computer disk storage space for the digital files must also be considered. Average uncompressed TIFF 8-bit grayscale files scanned at 300 DPI on the document can result in file sizes from 20 – 40 MB in size. For a 35 mm roll of film with 1000 images that can translate to 20 – 40 GB of disc storage space.

Purchasing microfilm-scanning equipment can be an expensive proposition. For those archives with smaller collections a less expensive alternative exists. Service bureaus from around the world can provide the service of scanning and converting microfilm into digital images. These service providers can provide images on a single roll of microfilm or thousands of rolls of microfilm. Some providers will require the microfilm be sent to their site for scanning. Other providers at an increased cost may come to the archives to scan the microfilm on site. The service bureau alternative can allow an archive to scale their costs as they begin the conversion to digital imaging. As the archive gains experience with the digital imaging, decisions can be made to continue vendor provided scanning or to acquire equipment and train archives staff to do the conversion.

Maintaining a facility that images documents on microfilm and stores the microfilm in an archival manner requires trained personnel that understand the archival process. Capturing and preserving digital images requires specialized knowledge. However, this knowledge is no more complex than the knowledge required for microfilm imaging and preservation. Dedication to training personnel in the digital process will allow for wise decisions in establishing a digital program. Mentoring by established archives or institutions can assist in gathering this knowledge and training personnel. The GSU and other institutions can provide knowledge about the general processes and assist in establishing a digital program.

If the archive decides to establish its own in-house microfilm conversion or digital imaging program, dedicated personnel will most likely be necessary to run and maintain the equipment, process images, establish preservation and migration plans and implement them. If the archives only desires to house digital images, then personnel already assigned to preserve other objects may be trained to plan and implement a digital preservation and migration plan.

Choosing a proper file format for the digital images can be an important decision. Microfilm can be scanned to digital files of varying file formats. Most primary scanned images are in bit mapped or raster graphics file format. These images are composed of rows and columns of dots or pixels that represent the scanned image. The number of rows or columns per inch of document represents the dots per inch or DPI resolution of the image. As the DPI increase so does the file size of the document increase. A 300 DPI resolution document will be two to three times as large as a 200 resolution DPI document.

Images can be stored in compressed or uncompressed file formats. Compressed formats squeeze the file size down and allow for more images to be stored in the same amount of storage space. Two types of compression are used, lossless and lossy compression. Lossless compression allows the image to be restored to its original size and quality without loss of information or degradation to the file. The other version of compression is lossy compression. This compression scheme allows the image to be reduced to very small sizes relative to the original file size. The compromise is that quality of the image is sacrificed. Once compressed the quality of the original image cannot be returned after decompression. This compression scheme is used with the JPEG image format. The amount of compression in a lossy file format generally corresponds to the quality of the image delivered. Less lossy compressed file sizes tend to provide better quality images. Higher lossy compressed file sizes cost less to store and can be delivered more quickly over network and Internet connections. A careful balance of image quality, speed of delivery and cost of storage must be considered when determining the amount of lossy compression to use on a set of digital images.

Several different image formats exist. TIFF is considered by many to be the de facto standard image file format for digital preservation. This raster-based format can allow for uncompressed, lossless compressed and lossy compressed images. Many institutions store preservation images in TIFF uncompressed formats. JPEG is another common image file format. JPEG is a lossy compression format. This format can deliver readable images at greatly reduced file size. This is one of the standard image file formats used on the Internet. Several other public domain and proprietary file formats exist.

Careful consideration must be given to the selection of a file format. Formats should be selected that are generally available throughout the digital imaging and computer world. Selecting a file format that is not

standard or widely accepted can mean expensive conversion costs later on if the file format is discontinued or no longer supported. Proprietary file formats that could be discontinued or under-supported can be a concern. Those popular file formats in the public domain generally tend to be more stable and allow for longer digital image conversion cycles.

Digital files can be stored in several different ways. Popular optical media such as CD's and DVDs can be used to store the digital images. These media types can have life expectancies from as little as 1 year to 100 years depending on the quality of the original disk. CDs and DVDs containing gold tend to have the longest life. Magnetic disc storage that is not refreshed can have a life expectancy of 10 years. Magnetic tape has a life expectancy of 5 – 15 years. With most of these media, the technology used towrite the digital image on the media will be outdated before the media fails. A regular program of media testing and refreshing the data should be undertaken.

- Digital Recapture

On occasion microfilm conversion to digital imaging may not be practical. When microfilm is of poor quality, the scanned image cannot improve the quality. The best that be hoped for is an accurately scanned poor quality image. In such cases where the original documents are still available digital recapture of the documents should be considered. This method can offer several advantages in addition to those listed for microfilm / digital conversion:

Better tonal ranges can be obtained from the original record when the document is scanned to grayscale or color digital images. Microfilm tends to increase contrast and distort the actual tonal representation of gray areas of the document. Digital images of ink-smeared or soiled areas may exhibit text visible to the naked eye but obliterated on the microfilm copy of the image. Grayscale or color digital images would show more detail in these low contrast areas.

At the time of digital imaging the image resolution, masking, exposure, and other camera settings can be set to best suit the particular documents being imaged. This allows the operator to manipulate the camera setting to achieve the ideal settings to render the documents most favorably.

There is at least one major factor that must be considered when evaluating digital imaging of original records instead of scanning existing microfilm. Many original records are very fragile. These records may not be able to withstand the rigors of the physical handling necessary to digitally image them. In some cases the microfilm images may actually be in better condition than the original documents. In other circumstances, the original documents may not be available to digitize. In such circumstances, microfilm conversion would be the wisest or perhaps only choice.

Digital imaging equipment comes in several styles and price ranges. From flatbed scanners to planetary digital cameras, the imaging equipment can cost USD $100.00 to in excess of USD $75,000.00. Most digital camera equipment will generally require a computer to operate software that controls the digital imaging device. Computers must be of sufficient power to handle large graphics files and quickly process images as they are scanned. As with microfilm scanning, attention must be given to computer disk storage space for the files created in the scanning process.

Service bureaus can also be found worldwide to image documents in digital formats. Digital imaging on-site or off-site can be arranged. The prices of digital imaging services range widely based on the number of documents to be imaged and difficulty in imaging those records.

File format and image size considerations are identical for both microfilm digital conversion and original digital imaging. Again, a study of the documents will help determine a balance between quality images with larger file sizes and cost savings and speed of delivery associated with smaller file sizes.

- Metadata

Digital images are discrete individual objects that by external observation reveal no information about their content. Metadata is data used to describe digital images in a manner that allows for organized access and information about the images. Metadata can provide information about the content description about the images, the technical data about the creation of the images and the intellectual property rights about the images. Indexing information linked to the images can be stored in the metadata. Administrative information necessary to administer the images within the organization can be added to the metadata.

This metadata must also be stored in electronic form. As the images are migrated and converted in the preservation process the metadata must also be migrated and converted. Careful checks must be made to assert that the metadata is coordinated and synchronized to the images as the images move through the preservation process.

When implementing any kind of system for delivery of digital images, consideration must be given to the methods of delivery of the distribution images. Distribution copies of digital images can be delivered in several different ways. Archives may desire to keep distribution costs low or to limit access to the images. A stand-alone computer dedicated to viewing the images may be a viable choice. This computer could use CD's or DVD's of images. The images could be viewed in much the same way as microfilm would be viewed on a microfilm viewer. While this system might be practical for a small collection, a larger collection may require a more sophisticated approach.

Larger collections of distribution images may be loaded onto computer disc arrays that would allow for large numbers of images to be stored. These disc arrays can be configured through a network to provide access to the images by a network of computers. The images from the disc array could also be delivered via a server to either an intranet or an Internet site. The server system can be scaled from a single server / disc array serving a local intranet to enterprise-sized computer systems designed to deliver millions of images per day.

Charging a fee for access to the images can be used to offset the cost of image delivery. The success of charging fees for access is dependant upon the perceived value of the images by the user. Value can be added to the images by providing indices to the images. Content that is desirable to the public and enhanced access to that content can expand the user base and enhance their ability to use the images.

The usage rights to images has been a topic of concern. This has been accentuated by the ease of delivery of digital images. Careful research must be undertaken to assert the archive's rights to digitize microfilm or original documents in their possession. Purchased microfilm in many cases may not allow for conversion to digital images. Published volumes may contain active copyrights that will not allow for digital conversion. Manuscripts may be protected from copy and distribution. Permissions must be sought for such digital conversion. An understanding of a particular country's copyright laws is paramount. Careful documentation should be maintained as a regular part of a digital imaging program. Contracts, letters of release, and other types of documentation related to imaging rights should be retained. Because digital images allow for greater access by the public, increased exposure to legal action can be anticipated unless assertive action is taken to determine and guard the rights of the archive to the digitized material. Legal consultation may be necessary to determine the archive's rights and duties.

- Conclusion

Microfilm collections provide an excellent preservation method for valuable collections. In today's world of computers and electronic technology, digital images provide unprecedented access to the information that archives hold. The process of digital imaging and microfilm conversion must be carefully considered. If the process is properly implemented the archive will be rewarded with a system that meets the needs of the archive and its patrons. By capturing original documents with digital cameras or converting microfilm to digital images, archives can provide the access desired by users. Creating and linking indices to the images will allow users to access records once difficult to search and enhance their research experience. Increased access to records can mean more patron use of the valuable documents held by the archives. The costs of implementing digital images from microfilm are greater than delivering microfilm, but the rewards are worth the effort. Allowing more patrons to use the records of the archives is the ultimate reward of a carefully crafted digital imaging program.