General Service Office Archives DIGITIZING ARCHIVAL MATERIALS GUIDELINES

INTRODUCTION

We are living in a world of evolving technology. As such, archives and libraries must constantly adapt to new standards when establishing and maintaining digital repositories. Creating a digital repository can be a challenge, partly because it involves a variety of divergent subjects and issues. The methodology behind creating digital repositories, an awareness of the costs involved, and the ability to make decisions on what to digitize are all areas you might need to address when thinking about building a digital library. Moreover, you will need to understand the technical criteria for the different materials in your collection, such as handwritten documents, photographs, magazines, books, audio, film, and more. The security of the repository, questions of access and use, as well as the issues of servers and back-up computer systems are also key points to consider.

The General Service Office Archives regularly receives questions from local A.A. Archives on digitizing collections and these guidelines are intended to be informative; they do not include everything there is to know about managing digital repositories, but they present a good start. No repository, no matter how large or small, will be able to digitize their entire collection. Digitization can be costly, time-consuming and some materials may not be fit for digitization or even worth the effort. Not all area or district archives have the resources for digitization projects and may instead choose to focus on other forms of preservation.

WHAT IS DIGITIZATION?

Digitization is the process of converting information, such as text, photographs, audio and video files, into digital format. Digitization allows for the preservation of the content of the material by creating an accessible copy, putting less strain on the original. Digital objects are then maintained in a digital repository that offers a convenient way to store, manage, access and preserve these copies.

It is important to note that, while often conflated, "digitization" and "digital preservation" are two separate

concepts. Digital preservation encompasses a variety of activities that aim to provide reliable, indefinite access to digital resources. Digitization is often one of those activities, however merely digitizing a record does not ensure that it will be usable forever. Other sections in these guidelines provide more information about how to properly preserve and maintain digitized records.

There are three basic digital repository functions.

- 1. The acquisition, or capture, of digital content.
- 2. The storage and management of digital content.
- 3. The retrieval of digital content and creation of deliverables.

Hardware and software are needed for these functions. Their cost can range from zero to expensive based on the choice of hardware and software.

BASIC CONSIDERATIONS IN PLANNING YOUR PROJECT

Before establishing a digital repository consider some of the following questions: What are your reasons for developing a digital repository? Which reasons are most significant to your archives' mission statement? For example, some of the reasons for digitizing might include increased access and use, security, preservation, management, and authenticity. At the General Service Office Archives our digitization efforts have aided in improved access, as well as improved preservation of the materials by reducing handling of the originals. You will want to take time to decide upon which digitization equipment would be best for your repository. Basic research should be done prior to purchasing hardware and software. Look into possible threats to your repository, including media, hardware, software and/or power failures.

MAINTAINING PRIVACY AND ANONYMITY

Protection of the privacy and anonymity of A.A. members is a critical issue that must be examined when maintaining digital files. The development of procedures seems to be a matter of local policy decision by the archives committees, but the necessity



for protecting the confidentiality of correspondence and the anonymity of the correspondents is, without question, an important consideration and a trust that falls upon all A.A. archivists and archives committees. You may want to consider implementing a policy for using the digitized material and include guidelines for posting material on the internet and social media. For more information on anonymity online see GSO's A.A. Guidelines on Internet (MG-18).

SECURITY

Creating and sharing digital files online may present a host of unique challenges relating to security. For example, if your archive stores or shares files through a cloud storage provider (think Google Docs, Dropbox, etc.) you will want to carefully read any user agreements to determine how and to what degree that provider can view or use your data. Additionally, the archivist(s) should discuss security measures around access to online materials. You may consider utilizing password protection, user permissions, or other features to ensure that records containing personal information remain secure.

As digital redaction tools have proliferated, so have conversations around data security. Many websites allow you to upload documents, edit or redact them through your web browser, and download the new version. However, this opens up the possibility of the provider retaining your original files and/or embedding unwanted and potentially dangerous programs in your downloaded file. Remember that A.A. archives often contain sensitive information and choose your providers accordingly.

SOFTWARE REQUIREMENTS

First, decide on the type of scanning application that would best suit your needs. For small, local collections, simple Windows or Mac-based applications may be used. For a significant collection containing at least 5,000 or more digital files, consider using content management (and file management) software.

It is important to have good software tools for scanning, file editing and PDF file functionality. Many scanner vendors bundle software with the driver software needed for the scanning device. There is an abundance of open-source software that can be downloaded from the Internet. Many software and hardware vendors offer reduced prices to nonprofit entities.

If a program is open-source, its source code is freely available to its users. Thus, a user has the ability to take this source code, modify it and redistribute it. Opensource software rarely comes with technical support and users rely on an online community for guidance and support. Technical expertise may be needed to use open-source software. On the other hand, proprietary or closed-source software consists of programs distributed by a brand for purchase. Proprietary software offers user support, which is typically an attractive quality for users without advanced technical skills. Regardless of which type you use, you will need some technical skills to get your system up and running. It is important to carefully review the Terms and Conditions of your software license prior to any signed agreement. One important consideration to address is to know what levels of control are maintained by you with regard to accessing and editing your data, should your software license expire.

HARDWARE REQUIREMENTS

Basic hardware needs will typically consist of the following:

- ✓ A personal computer (preferably one with a high processing speed)
- ✓ A high-quality monitor or laptop screen
- ✓ An external hard drive for storage (capacity based on the size of your collection)
- ✓ A laser color/black and white printer
- √ A flatbed scanner

Optional hardware may include the following:

- ✓ An internal or external CD-ROM drive (if you will be working with material on discs)
- √ A hand-held scanner

You should also verify that your computer has the necessary ports to accommodate any external drives you may use. Consider the types of material you plan to digitize (text, photographs, artwork, etc.) before choosing a scanner. Flat-bed scanners are recommended for most digitization projects, while a hand-held scanner is useful for oversized material, fragile material, and bound-volumes.

FILE STORAGE

Storage media for the repository's preservation files is probably best satisfied by hard-drive systems. You may choose to store "access" or "use" copies of files on discs or flash drives; however these media will not guarantee longevity and are not generally recommended for master files. If material is stored on discs, it is recommended to use high quality or "archival" quality CD-Rs and DVD-Rs and to store these discs in a temperature and humiditycontrolled environment. Rewritable discs (CD/DVD-RW) pose more risk for data loss and are not generally recommended for archives.

In order to ensure the longevity of files, archives and libraries often have both "preservation" and "access" copies of digitized assets. A preservation copy is usually saved in the highest-quality format possible and stored on a drive or in a folder where it will not be edited. An access copy may be of a slightly lower quality (to help cut down on storage size) and can be duplicated, shared, etc. This helps to ensure that if digital files are edited, the preservation copy still reflects the original content and it reinforces the concept of LOCKSS — an archival standard which tells us that "Lots of Copies Keeps Stuff Safe". Even storing the same version of a digital object in two different places can significantly aid in protecting against file loss or data corruption.

Software and hardware technologies require ongoing attention due to continuous and rapid advancements. Your digital records can be put at risk for loss due to lack of software support, hardware going out of production, or even the inability to find compatible ports or charging cords. This is called digital obsolescence. Oddly, today we can still read centuriesold original documents in their native language. Yet, in fairly short order, digital media technology has advanced from floppy disks, to diskettes, to CDs, to DVDs, to Blu-Ray disks and flash memory media. Remember that any storage medium can fail at any time. Archive your data on more than one medium and check your archives regularly for failures.



Some examples of storage media designed to store data, audio, or video over the years.

Cloud storage is often considered as an option for backing up and/or sharing files. It can be a useful tool in preventing data loss if your computer or external hardware backups were ever destroyed, lost, or became unusable. However, particularly when thinking about records that require maintenance of anonymity, it is important to be aware of how cloud storage functions. Cloud backup involves sending a copy of the data over the internet to an off-site server. The server is usually a third-party service provider, who charges a fee based on bandwidth, capacity, or number of users. Most of these companies assume some level of control over the data you have deposited into their custody. Before utilizing cloud storage, carefully examine the Terms of Service and consider these significant issues:

- <u>Security:</u> How safely is your data protected against malicious acts like hacking? How proactively does the provider work to prevent data loss?
- <u>Privacy:</u> What rights, if any, does the provider hold to data that you upload? What privacy controls are offered to monitor access to data?
- <u>Flexibility:</u> Does the provider store data in a proprietary format? How are you able to access or remove your data if choose to end your relationship with the provider or if they go out of business?

SELECTING ORIGINALS TO DIGITIZE

Many archivists are faced with the daunting question of what to digitize. Some learn about the digitization efforts of other repositories and think that they are just not doing the right thing. Rest assured that your efforts in preserving A.A. history are a good start. We cannot tell you specifically what to digitize because the scope of your collection is unique.

Be selective in the choice of planned digital content by focusing on quality not quantity. We also suggest that you focus on digitizing material that pertains to your area or district, such as minutes, fliers, correspondence, and other items generated by your local A.A. entities. Digitizing these items will provide you with an electronic duplicate, which is easy to access and use, yet sparing the original item from overuse. In addition, many scanning and editing programs allow you to perform OCR (Optical Character Recognition) on documents which will make them text searchable. While this works best with typed documents, some handwriting may be captured as well. OCR is a great tool if you need to search through a lengthy document for a piece of information and in many cases can drastically reduce research time.

Some materials are born digital, meaning they originate in a digital form (e.g. desk-top publishing, digital cameras, etc.). However, they should be imported into your digital repository and saved in a way that adheres to the file naming and formatting standards already agreed upon for digital objects. Textual documents can be digitized if no harm is done to the document during the process.

Documents that do not fit onto a standard flatbed scanner are called oversized documents. These documents, as well as bound books, should be digitized using a hand-held scanner, scanning digital camera, or a standard digital camera. It is not recommended to scan bound volumes on a flatbed scanner as this may cause permanent damage to the book spine and binding.

BUILDING A STRUCTURE FOR YOUR DATA

Now that you have adequately addressed these significant questions, formulate a plan on organizing the data in a sensible and useful way. This is also a time to decide on a file name convention for your folders and files. A file naming system should be agreed upon prior to digitizing in order to create consistency. Descriptive file names should relate to the item being digitized and may contain the name of the item and/or the date. For example, 2024_ProjectTitle_ItemTitle. If you have questions about file naming guidelines, feel free to contact GSO Archives.

An example of a potential folder structure:

- District Meeting Records (Parent Folder)
 - ► Meeting Minutes (subfolder)
 - ► 2024_05_01_District10_MeetingMinutes(file name)

METADATA

Any individual engaged in a scanning project will come across the term "metadata." Simply put, metadata is "information about information." It is information used to describe, locate and retrieve files stored in a digital library. It is the key to ensuring that electronic resources will continue to be easily accessible in the future. In Microsoft Windows and other software, metadata may also be called "properties." Most metadata, such as file size and date, is automatically generated when a file is created. Application software also generates metadata when a file is created, such as the file name and type.

Names and values should be simple and consistent. One way to achieve this is to use a controlled vocabulary. This is simply an organized list of words and phrases used to tag digital content and retrieve it through search. Use unambiguous and meaningful folder and file names and descriptive keywords. A file name is metadata and should offer a description of its content.

In the library environment there are several complex standards or schemas that exist for describing digital files for different types of material (books, photographs, audio, etc.), and each will have its own unique structure. However, as long as consistency is maintained, you can easily create your own simple metadata schema to fit your repository's resources and needs.

DIGITIZATION SPECIFICATIONS

A number of factors will affect the appearance or playback of digital objects, as well as their longevity. Many file specifications and formats exist to try to account for these factors. Due to the wide variety of object types and the unique characteristics of each item being digitized, we cannot provide exhaustive guidelines here. However, some general tips listed below provide a good starting point for digitizing the most common types of objects in a digital repository.

Document type	Resolution	Color mode
Documents with printed type (e.g. laser printed or typeset)	300 ррі	8-bit grayscale mode
Documents with poor legibility (handwritten, copied, etc)	400 ppi	8-bit grayscale mode
Photographs and color documents	400 ppi	24-bit RGB mode

Open file formats should be used as they are publicly obtainable and are recognized by multiple programs. The table below provides for the most common types of open file formats used in digitization projects.

File Type	Long-term preservation	Access/use copy
Images	TIFF	JP2, JPEG, PNG
Documents	PDF/A, PDF ^{**}	DOCX, RTF, TXT
Spreadsheets*	PDF/A, PDF ^{**} , CSV	XLSX, ODS
Presentations	PDF/A, PDF ^{**}	PPTX, ODP
Audio	BWF, FLAC, WAV	MP3, MP4
Video	AVI, MOV (Uncompressed)	MPEG4, WMV

*Saving these file types as PDFs may help to prevent future rendering issues. However, if your spreadsheets or presentations contain formulas, transitions, etc., you may want to preserve those functionalities by saving an MS Office compatible copy as well.

** If using PDF format, choose the highest quality settings available, including lossless compression, high resolution, and searchable text



The 74th General Service Conference Convenes: "Connecting with Love, Unity and Service"

The Twent Continues Startistic Conference was held April 4-20, 2544 in Sets tooks City, at the Startistic Handle as the Monolevin Redge Wab 154 Conference monters the "Anti-General Fertiles" Wab 154 Conferences monters the "Anti-General Fertiles Conferences provided city contraction of a smars worth of engineer ent and daster for an ingress actions the trends States and Context and April 10 and 10 are started by the action and the start and the approximation of engineer ent and daster. The policy of the action of the trends States and Context and Approximation of the trends States and Context and Approximation of the trends States and Context and Approximation of the Approximation of the Approximation and Approximation of the Approximation of the Approximation and Approximation of the A

"We are all providing to be part of Conference weeks and want to thank off of you has your dedicated service to the reduxing transfer our work. Work happy and gateful heads, we have fore apportantly so demunstrated asserts and an doughd A voting in the Assess before is. The free and transfer of the source of the non-kew days with the a second to the determination of the source of the assessment to the determination of the source we poole to the providence of the non-kew days with the source of the source of the source of the source we poole to the providence of the source we poole to be a providence of the source."



The 74th General Service Conference Convenes: "Connecting with Love, Unity and Service"

The 74th General Service Conference was held April 14-20, 2024, in New York City at the Marriott Hotel at the Brooklyn Bridge. With 134 Conference members the 74th General Service Conference provided the culmination of a year's worth of engagement and discussion on topics integral to A.A. members, groups, districts, areas and regions across the United States and Canada, whered in the beginning of another year of sharing. There as our fourther worth to be the "We are all privileged to be part of Conference week and want to thank all of your for your dedicated service to th Fellowship that saved our lives. With happy and gratefh hearts, we have the opportunity to demonstrate our gra itude by the focused attention we give agenda items an our thoughtful voting on the issues before us. The los and respect we show each other in the next few days wi be a testament to the effectiveness of our present.

Poor scanning resolution and color settings can result in low fidelity copies (left) that are inadequate for preservation.

QUALITY CONTROL

A quality review of the scanned images is a highly critical and significant aspect of the digitization process. Regular inspection of the scanner should also be performed to ensure that physical matter, such as dust or specks of paper, is not left on the flatbed.

Remember to always keep your work area clean, which includes daily maintenance of the scanning equipment. Many old documents or books tend to be dusty or even moldy and will leave potentially harmful substances on the flatbed. Visual checks should be done periodically to the digital image files to ensure quality is maintained in the following areas:

- The correct format (TIFF, JPEG, etc.) and resolution are used.
- The digital image has the correct orientation.

For multi-page documents, all pages are accounted for and in the correct order.

- The digital image is complete and not cropped.
- The image is not skewed.
- Digital artifacts (dust, specks of paper, etc.) do not appear on the image.
- Image quality is maintained for color, tone, sharpness, and contrast.
- Metadata related the digital files are named properly.

It's important to remember that simply digitizing files does not guarantee that you will have access to them forever. Before establishing your digital repository, take time to think about a regular schedule of file maintenance that will help to keep your assets organized and usable for the future. You may want to consider an annual review of your file structure to ensure that newly digitized items are being properly described and sorted. Check the reliability of any external storage media you're using and confirm that it is still in working order and files can be opened. Update instructions on how to access files (passwords, data locations, etc.) and keep them in a safe place. Also, consider staying up to date on file format standards — over time you may need to "migrate" or update files to new formats to ensure they can be opened by newer software programs.

PRESERVATION OF ORIGINAL MATERIALS AFTER SCANNING

The end result of a digitization project is a digital surrogate of the original. It is important to take proper care of the original material. Although you will have a digital copy, it is still vital to protect historical documents from deterioration by using archival quality material and optimum environmental conditions for storage.

If original materials are ever being removed from the collection for any reason, it is even more vital to ensure that you have multiple reliable copies of the digital surrogate stored in different locations. You may also want to keep metadata on why the originals were removed and where they went, if applicable, to provide context for future users.

See the GSO Archives' Preservation Guidelines for a more comprehensive and in-depth look into the preservation of archival material.

FREQUENTLY ASKED QUESTIONS

What is digital obsolescence and how can I protect my archives against it?

Digital obsolescence is a general term for the loss or degradation of digital assets due to changes in technology. For example, floppy disks were at one time a standard for digital file storage and most home computers had floppy disk drives. Now, even if you purchase an external drive to read a floppy disk, the files on them may have degraded, been corrupted, or simply be unreadable on modern software applications.

Avoiding digital obsolescence is an ongoing process. You can help protect your files by saving them in preservation-friendly formats, storing copies on different types of media (or in the cloud), and doing regular checks to see if files are still readable by newer systems. Many archivists also create preservation copies of files — these are rarely opened, never edited, and can be used to create or replace access copies that may have been corrupted, deleted, or lost.

What are some long-term challenges for digital archives?

Maintenance of files and continuity of access to data are two major long-term challenges facing smaller archives. It is important to remember that digitizing files is not the end of our work, but just another step toward preservation. Be sure to consider ongoing budgets for things like cloud storage, new hardware or software, etc. Try to implement sustainable workflows that let you digitize and maintain data over time to avoid getting far behind on existing projects. Have plans in place for moving digital or physical storage in the event of an emergency. Lastly, help to keep your files accessible by storing information and passwords for file storage locations in a safe place and, if applicable, sharing with trusted team members.

How do I decide what to digitize first?

This is often a tough choice, but there are many factors you can consider to help you along the way! First - what materials are most at risk of loss or damage? Items that are torn, faded, stored in poor environmental conditions, etc. may be good candidates for early digitization to ensure you have a reliable copy in the future. Next, how unique are the items in your collection? Some widely distributed resources, like the A.A. Grapevine or Box 459, are already available in digital format through the GSO. Try to focus on avoiding duplication so you can get the most out of your digitization project. Lastly, what items are most important to your community? Consider discussing your project with other A.A.'s who may also use the archive or be familiar with the history that is being documented. Together, you can determine what digital collections might provide the greatest benefit to the community.

What database/software should I use for my digitization project?

Because there are so many products available, and because the needs of each digital archive are unique, the GSO Archives cannot make specific recommendations for software, hardware, or archival supplies. Many reputable archival organizations have guides on smallscale digital archiving that give examples of workflows and list the products they use. Consider your information sources carefully and try to find multiple reviews or product descriptions if you're able.

Should I store copies of files on multiple types of media? If so, which ones?

There are no hard and fast rules, but one common method used in archives is LOCKSS — Lots Of Copies Keeps Stuff Safe! Any type of storage is subject to potential loss, failure, or destruction. Files stored in the cloud may become inaccessible because of server loss, lack of internet connection, or even a lost password. Physical files can be lost, stolen, or damaged in environmental disasters. Digital files stored on CDs, flash drives, or hard drives can be damaged, corrupted, or broken. You can help to protect your data by ensuring that you have adequate virus and malware protection on your computer, that you are sharing or receiving files from known sources, and that you are maintaining basic environmental controls for physical disks and drives.

How do I handle donations of digitized or born-digital materials?

"Born-digital" generally refers to materials that were created in an electronic format. A document written using Microsoft Word and saved on your computer is "born-digital", whereas a hand-written letter that you scanned and saved a copy of would be "digitized". In the course of your digitization project, you may find or receive more of either of these types of files that can be integrated into your collection. For donated materials, we suggest creating a donor agreement that acknowledges the transfer and documents what usage rights you have for the items. Try to keep clear metadata regarding where items originated from and whether they are being changed in any way so you can locate them if needed. You may want to rename files and/or save them in a more preservation-friendly format to ensure consistency across your collection.

What types of metadata should I include in my collection?

This will depend somewhat on the collection itself, but in general you should strive to ensure that your materials are well documented and discoverable. Dates are an easy example of this — whether you list it in the file name or in a spreadsheet/database, it is always helpful to know when an item was created. Location information is also similarly useful. Group names or relevant area/district information may also be helpful for future searching. If you identify individuals in your metadata, be sure that those records are secure and do not violate anonymity. You may also want to use a controlled vocabulary to create subject or topic listings that can apply to different items to help group them together. If your time and resources are limited, start with a basic description that captures the most vital information about the object, and then return later if you're able to add more detailed descriptions.

Can I share archival materials online? What about copyright?

These are both complicated questions which many archivists navigate repeatedly, especially since our collections are so unique. Sometimes we do not have documentation regarding copyright holders or sharing permission for materials, and we may not even know whether such documentation was ever created! In such cases, it is best to proceed with caution and assess potential risks based on the specifics of the collection and how material will be shared.

If you've determined that you do have sufficient permission to share material online, remember that A.A. archives require additional care in protecting privacy and anonymity. The "Security" section of this guide contains helpful information on how to keep digital files safe whether they are in the cloud or on a physical disk/drive.

PLEASE NOTE...

For answers to any specific questions, and lists of additional resources, feel free to contact the GSO Archives at archives@aa.org or 212-870-3400. Other helpful information is available on GSO's A.A. website, www.aa.org.

SOURCES FOR MORE INFORMATION:

Please note that the GSO Archives does not endorse nor affirm these websites, and simply provided them as helpful external resources.

DPLA Self-Guided Curriculum for Preservation https://dp.la/news/new-self-guided-curriculum-fordigitization/

Digital Preservation Coalition http://www.dpconline.org/advice/ preservationhandbook