

Geneva Historical Society Digital Preservation Plan

Maya Reid, MLIS Candidate

University of Maryland, College Park

Table of Contents

Executive Summary	3
Survey and Report.....	4
Introduction	4
Digital Content	4
Content Management	5
NDSA Levels of Digital Preservation.....	5
Further Collecting and Resources	6
Conclusion.....	6
Next Steps.....	7
Introduction	7
Metadata	7
Storage and Geographic Location.....	8
File Fixity and Data Integrity	8
Information Security	9
File Formats.....	9
Conclusion.....	10
Policy.....	11
Purpose	11
Scope	11
Standards	11
Metadata	11
Storage Management.....	12
File Fixity	12
File Formats.....	13
Information Security	13
Review.....	13
Related Documents	13

Executive Summary

The Geneva Historical Society is located in Geneva, New York. Committed to serving the community and preserving the memory of Geneva, the Historical Society operates with a small staff and limited resources. During the Fall 2018 semester, the Geneva Historical Society participated in a consultation project for a Digital Preservation course at the University of Maryland, College Park. The purpose of this project was to advise the Society on ways they could improve their digital preservation practices.

The Society's digital holdings include their archival collection, educational and promotional materials, and business-related digital objects. The main concerns for the Society were a lack of inventory for their digital content, duplication of scanned materials, and the use of personal filing systems. This lack of standardization led to inefficiency as the staff found it difficult to locate their digital content. This adversely affected both the staff and the public.

To advise the Society, I formed recommendations based on the National Digital Stewardship Alliance's Levels of Digital Preservation. These recommendations are to:

- Create an inventory in Microsoft Excel with permission-based access. This inventory will be backed up daily to minimize any accidental loss of data. To make sure the inventory is always up-to-date, employees will record object metadata in the inventory immediately after scanning.
- Write a file plan to standardize the documentation of digital objects. This file plan will cover the file name and location standards. These standards will make it easier for the Society to locate their digital content.
- Have several copies of their complete collection. These copies should be updated annually.
- Run annual fixity checks to ensure digital materials have not been corrupted or tampered with.
- Use a limited amount of known and open file formats. Formats at risk of becoming obsolete or inaccessible should be migrated.

With these practices in place, the Society's digital content will be safer and easier to manage and navigate.

Survey and Report

Introduction

The Geneva Historical Society is located in Geneva, New York. The Historical Society is committed to preserving Geneva history and using digital and analog materials to educate the community and sustain community interest in Geneva's memory. The Society is made up of a small group of employees who wear several hats. The Society's greatest digital preservation concern comes from too many duplicates, a lack of a content management system, and no inventory of digitized items.

Digital Content

The Society's digital content includes historical materials that are part of their archival collection, promotional and educational materials, and digital items generated through business-related actions. The historical collection includes materials available on the Society's website, on NewYorkHeritage.org, and on the Rochester Regional Library Council's NYS Historic Newspapers website. There are photographs, videos, and audio available for public access on the Society's websites. The items available on NewYorkHeritage.org includes post cards, papers, and photographs. On NYSHistoricNewspapers.org are the Society's microfilmed newspapers. The Society no longer has the original newspapers and now rely on this website for research, both internal and for the public. The Society also has StoryMaps, which are hosted by Hobart and William Smith Colleges.

The Director of Education and Public Information uses digital content for promotional and educational purposes. The Educator uses photographs and maps from the Society's collection, public domain materials from websites such as Library of Congress and Wikimedia, and photographs taken at the Society's events. The Educator uses Illustrator, Indesign, and Photoshop files for publicity and programming. The Curator uses Publisher files for exhibits.

The Society's digital holdings also include image collections on the website media gallery through WordPress, a MailChimp media gallery, content on Google Photos and Dropbox, and other institution-based materials.

Content Management

Much of the Society's current struggle with their digital content comes from the lack of management for these materials. The staff at the Society practice independent digitization. When a staff member needs something digitized, they do it themselves. Employees are unaware of what items others have already scanned. These digitized items are kept in folders on the employee's computer, using their own filing system. This process has led to much duplication.

This duplication continues on the shared computer located in the Photo Archive. Using an Epson V300 scanner, staff scan items and store them on this computer. The materials saved on this computer are accessible to all of the staff through their own computers. This leads to duplication of the staff's individual files.

The employees do not all have the same computer or operating system. They use Windows but not everyone has Windows 10. The digital formats used by the Society are generally jpg, tiff, and pdf. The digital copies have varying resolutions and file sizes and there is no standardized filing system, so finding objects that other employees have saved is difficult and time-consuming. If the Educator is looking for a certain image and knows it has already been digitized, she will ask the Archivist or Curator if they know the location. If the image isn't digitized or easily accessible, the Educator will use the Epson V300 scanner and save the image to her desktop.

The staff agree that improving the state of their digital content is a priority. The staff believe that the digital content is too unorganized and that time is wasted looking for items and scanning materials already digitized. Getting their digital content under control is critical to the Society's mission, as both staff and members of the public are negatively impacted by the disorganization. Organization is needed for more efficient access and use.

NDSA Levels of Digital Preservation

In terms of the NDSA Levels of Digital Preservation, the Society is at Level 1 for the five areas. Although the Society has many copies, these duplicates are unintentional and uncontrolled and many have different factors (like resolution and size). The Society uses a Synology NAS system. Each employee's desktop is backed up daily using Syncback Free to a Synology NAS over P2P network. The Society currently rests at Level 1 for the Storage section. When it comes to file fixity, there does not seem to be a process in place for checking fixity for digital objects. For the Information Security section, employees do not know what files other staff members are using as

they digitize items independently, use different software, and have their own filing systems. There is no inventory for the Society's digital content, so the Metadata section is also at Level 1. The Society does seem to use a limited set of formats with jpg, tiff, and pdf, but without an inventory of the digital objects, a complete list of the formats in use is unknown.

Further Collecting and Resources

There are more digital objects that the Society would like to collect. There are 50,000 photographs and business and family records that still need to be digitized. The archivist would actually like to digitize the entire archive but does not have the resources. To improve the state of the digital content, staff members could devote extra hours to digital preservation and the Society could accept community volunteers. The Society has previously worked with a yearly budget of \$2000 for office equipment, including computers and software licenses.

Conclusion

The Society needs to get the content management of their digital materials under control, especially since they intend to add thousands of more objects into their holdings. Uncoordinated filing systems and a lack of an inventory contribute to their mass duplication and current struggle with inefficiency. Fortunately, the Society's resources and staff commitment make the prospect of an improved digital preservation process likely.

Next Steps

Introduction

The digital holdings of the Geneva Historical Society, located in New York, are spread out over various websites and institutions. The Society has partnerships with the Hobart and William Smith Colleges, NewYorkHeritage.org, and the NYS Historic Newspapers website. The Society's digital holdings can be found on these websites, their own Synology NAS storage system, and on Google Photos and Dropbox. These holdings include materials such as photographs, post cards, and microfilmed newspapers.

One of the largest problems the Society faces with their digital items stems from a lack of organization. When an employee digitizes an item, that item is saved and stored according to that employee's own personal naming and filing system. Staff regularly do not know what other employees have scanned, leading to much duplication and difficulty in finding digitized materials.

Before writing a preservation policy for the Society, I will first review the Society's current standing in the NDSA Levels of Digital Preservation (LoDP) and then recommend next steps that the Society can take to improve their standing. The NDSA levels can be found [here](#).

Metadata

Although Metadata is ranked as the fourth priority on the NDSA LoDP, it is a critical concern for the Society. To address their duplication and disorganization issues, it is pertinent for the Society to create an inventory of their digital holdings. This inventory can be created on a Microsoft Excel spreadsheet and should have headers for information such as directory location, file format, and file size. Resolution would also be a beneficial header so that, in the future, staff members can know if the digitized image fits the proper resolution that they are looking for. There should be a backup of this inventory in case the original file is accidentally deleted or corrupted. Simple ways to backup the inventory include using Dropbox or Google Sheets.

Other than just creating the inventory, the Society would also benefit from standardizing the documentation of digital materials. This standardization should include a file naming policy so that items are always named in predictable and understandable ways, increasing findability. While creating the inventory, files should be renamed in order to fit the standard. Another way to

standardize the documentation of the digital materials is to restructure the directories so that everyone may know the locations of items, instead of them being restricted to a single employee's desktop. One way to structure the directories can be to base them on the physical locations of the materials. This way, if an employee knows where to find the physical object, they can use that same process to find the digitized version. Thus, the inventorying process would also involve relocating digital objects to fit into the new directory structures.

So that every employee will know how to name and locate items, the Society should create a "file plan" document. The purpose of this document will be to explain the standards for naming and storing items and why the rules exist. For future digitization, staff members should record file information in the inventory right after scanning.

Creating the inventory will be a high resource requirement. Due to the amount of time that creating the inventory will take, the Society can use community volunteers and dedicate staff hours to the project. The Society can also have a "file clean-up" day or time every so often so that every staff member can devote time to inventorying and cleaning up the collection.

Storage and Geographic Location

The Society currently uses a Synology NAS system. To satisfy the storage requirements of the NDSA LoDP, an organization needs to have multiple copies of their collection in different geographic locations. To meet the level 1 requirement of LoDP, the Society needs to create a copy of their digital collection. The Society could use an application such as Preservica to store their digital materials. To reach the second level, the Society will need to document the storage systems they are using and what they need to use them (such as access information). The Society will also need to create and find a storage location for a third copy of their digital collection. This third copy can potentially be given to one of the Society's partners.

Fulfilling these steps will demand a medium resource requirement from the Society. Improving the Society's standing in this area could cost both time and money (should the Society decide to use an application like Preservica).

File Fixity and Data Integrity

File fixity refers to the digital object remaining as intended through time and transfer, meaning that details such as content and file size do not change. The purpose of fixity checks are to ensure

the digital objects have not changed. There are no fixity checks currently being done at the Society. To reach level 1 of LoDP, the Society needs to create fixity information upon creating and acquiring a new digital object. To do this, I recommend that the Society record the file size and file count. This information can be recorded in the inventory. To reach level 2 of LoDP, the Society needs to check file fixity for all acquired digital materials. The Society can do this using AVP's free application Fixity. I would also recommend that the Society practice annual fixity checks to make sure none of their content has corrupted.

For materials already in the Society's collection, the Society can record their file size and count while creating the inventory for future reference. Working on file fixity will be a low resource requirement for the Society. Documenting fixity information for existing files can be done while working on the inventory and the procedure for adding fixity information for new digital objects can be added to the file plan. Checking file fixity can be done using free applications, so there does not need to be a financial cost to the Society.

Information Security

Some information security is implemented at the Society in the Photo Archive. Although the public may use the computer, they cannot access all the files that staff can. These access restrictions can be further implemented to protect the digital collection. To reach level 1 of LoDP's information security section, the Society needs to identify who has read, write, move, and delete authorization to individual files. This should be done after the inventory has been created and existing materials have been renamed and relocated. The benefit of having these restrictions is that unauthorized people will not be able to tamper with or misplace files, maintaining the order and efficiency of the organization. The Society should also document who has the authorizations and access restrictions for the content.

Implementing this level of information security will be a medium resource requirement for the Society. The Society will have to decide and document who will be able to access, alter, move, and delete which collection materials, which may require a lot of time.

File Formats

Although the employees at the Society already know that they use formats such as pdf, tiff, and jpg, it is still important to document all of the file formats in use. This can be done once the inventory is created. After documenting the formats for each item in the collection, the Society

will be able to see each the formats being used. The Society should limit their formats to “known open formats,” meaning formats that are widely used and non-proprietary. This lessens the risk of the format becoming obsolete. To reach levels 3 and 4 of LoDP, the Society will need to monitor these file formats to see if they are becoming obsolete and migrate at-risk formats when needed.

Working on the file format section of LoDP will be a low resource requirement from the Society, as it should be relatively easy to do once the inventory has been created.

Conclusion

The most pressing need for the Society is the creation of an inventory. Once the inventory is created, the Society will be able to advance in the other areas of LoDP with lower resource requirements.

Policy

Purpose

The purpose of this policy is to set rules and guidelines for the management and preservation of digital objects in the Geneva Historical Society. This policy aims to address the following concerns and risks:

1. Duplication of digital items
2. Difficulty locating digital items
3. Deletion and corruption of files

The procedures discussed in this policy intended to mitigate these risks and concerns as much as possible.

Scope

The scope of this policy covers born-digital items and digitized objects in the Society's collection.

Born-digital refers to items originally created in digital form, such as podcasts or the Society's promotional or educational materials.

Digitized objects refer to the scanned copies the Society makes of their physical materials.

Standards

This policy is written to conform to the National Digital Stewardship Alliance's Levels of Digital Preservation. Due to this, increasing the quality of metadata, storage maintenance, file fixity procedures, file formats, and information security are the focus of this policy.

Metadata

An inventory of the collection will be created and maintained in a Microsoft Excel workbook. Headers for the spreadsheet may include directory location, file format, file size, resolution, physical location of the item if digitized, or any other headings the Society finds useful. A back-up inventory will be created and used on an external website such as Dropbox or Google. The back-up will be updated daily. Having a regularly updated back-up copy of the inventory will minimize damages in case the workbook is accidentally deleted or tampered with. To ensure this

inventory contains every item in the collection, employees will add object information immediately after digitization.

A file plan document will be created. This file plan will assert file naming and directory location policies. The file naming policy will be dictated by the Society. General guidelines for the naming policy include using underscores or hyphens instead of spaces and only using lowercase letters. Legacy holdings must be renamed in order to fit the naming policy. Legacy holdings refer to digitized and born-digital objects created before the implementation of this policy. A tutorial for batch renaming files in Windows 10 is included in the Related Documents section.

Digital objects should be placed on a shared drive in the proper location following the Society's directory location policy. Directories will be based on physical locations of objects, unless the Society finds a more suitable method. Legacy items should be relocated to fit the proper place in the directory.

Employees will refer to the file plan when naming and storing digital objects.

For quality control, the Society will practice "file clean-up days" twice a year, so that every employee can dedicate time to ensuring the metadata protocol is followed.

Storage Management

The Society currently uses a Synology NAS system. This server will be located in a cool temperature-controlled environment to prevent heat damage. The server should require administrative access so that a limited number of people are able to alter files.

Two additional copies of the Society's collection will be created to meet the standards of NDSA's Levels of Digital Preservation. For the first copy, the Society will use Preservica to store their collection. A second copy will be given to a partnering institution. The Society will document all storage systems in use and the information needed to access and use them.

Changes to a file must be replicated on all storage copies. The copies will be updated annually.

File Fixity

Using AVP's Fixity application, the Society will perform fixity checks annually. These fixity checks must be performed on all three copies of the Society's collection in order to be considered

complete. Fixity information, such as hashtags, should be recorded in a spreadsheet with a high level of authorization.

File fixity refers to digital objects remaining as intended through time and transfer, meaning that details such as content and file size do not change.

File Formats

All the file formats in use will be documented in the Excel inventory. The Society will limit the formats in use, so that only known, open, and widely-used formats are chosen for preservation. At risk-formats will be migrated to safer formats when needed.

At-risk formats refer to formats that are becoming obsolete or inaccessible.

Information Security

The Society will identify who has read, write, move, and delete authorization to individual files. These authorizations and access restrictions will be documented.

To protect the data recorded in the Excel inventory and reduce risk of deletion, only one person at a time will have permission to edit the workbook. When the inventory is open and in-use, others will be able read it but not edit. Since the workbook will also be backed up daily, this will reduce any damage caused by accidental deletion.

If possible, unauthorized attempts to access files will be logged and reviewed daily.

Every computer will be upgraded to the same operating system. The Society will schedule days to manually update the operating systems and run virus scans. It is important to manually update operating systems in case an automatic update does not occur.

Review

This digital preservation policy will be reviewed annually to check on the effectiveness of the policy and whether it meets the needs of the Society.

Related Documents

[NDSA Levels of Digital Preservation](#)

[How to batch rename files in Windows 10](#)

[Digital Formats: Factors for Sustainability, Functionality, and Quality](#)