

Digital Preservation Report for the National Trust for Historical Preservation

Conducted by Margaret Rose Hunt
Fall 2018

Table of Contents

Digital Preservation Survey - Completed October 4, 2018	4
Background on the National Trust for Historic Preservation	4
Scope of the Digital Collections	4
Current Organization of Digital Materials	5
Staff Response to Digital Preservation	5
Future Collections Impacting Digital Preservation	6
Human Resources and The Use of Technology	6
Preservation Plan	7
User Needs Gathering Phase	7
Education and Training Phase	7
Pilot Program Phase	8
Implementation	8
File Management	8
Storage and Geographic Location	9
Fixity and Data Integrity	10
Information Security	10
File Formats	10
File Management	10
Storage and Geographic Location	11
Information Security	11
File Formats	11
Aggressive Plan	12
File Management	12
Storage and Geographic Location	12
Fixity and Data Integrity	12
Digital Preservation Policy	14
Importance of Digital Preservation	14

Scope of Digital Preservation	14
Purposes of Digital Preservation	14
Digital Preservation Practices	14
Challenges of Digital Preservation	15
Digital Preservation Participants	15
Resources	16

Digital Preservation Survey - Completed October 4, 2018

Background on the National Trust for Historic Preservation

The National Trust for Historic Preservation is a non-profit organization that works to protect and advocate for historic buildings across the United States. There is a central headquarters for the operations in Washington, D.C., both on the business side and on the collections side, as well as the individual historic sites in multiple states. The collections staff works with the institutional documents for collections and with the staff at the sites that are found across the country. These sites fall into 3 categories: Stewardship sites, Co-Stewardship Sites, and Affiliated Sites. Stewardship Sites are owned and operated by the National Trust, while Co-Stewardship Sites are owned by the National Trust but are operated by a separate non-profit. Affiliated Sites are neither owned nor operated by the National Trust, but are affiliated with the organization for tech support and marketing. The National Trust has no involvement with the collections of Affiliated Sites. Of the 27 sites, 7 are Stewardship Sites, 14 are Co-Stewardship Sites and 6 are Affiliated sites. The national office has a role in overseeing the collections at the Stewardship and Co-Stewardship sites, but the collections are maintained and organized by the staff at the individual sites. This has led to a lack of standardization in the past when including collection documentation in the collections management system and when naming files.

Scope of the Digital Collections

The majority of the digital materials held by the National Trust and their sites are photographs of 3-dimensional collection objects, photographs of paper materials in the collections and scanned documents containing information about the collections. These photographs and scanned documents (as .pdf documents) are found in the National Trust's Re:Discovery Proficio Collections Management System. The version of Re:Discovery that the National Trust bought is a cloud-based version that can be viewed by anyone with login information. A large portion of institutional documents, not necessarily related to the collections, was scanned in 2013. These scanned documents are in a folder on the scanned drive. The file names are indicative of what is in the file, for the most part. It is usually an overview and therefore the collections staff is not sure what is exactly in the set of scanned documents. The collections documents, like deeds of gift, are held by the site or by the national office. Some of this information is included in their Re:Discovery Proficio system, either through the attachment of documents or through the entry of the information into the system fields. Most of this information is held in an analog form, rather than a digital one. Modern photographs of the sites are found in a separate digital asset management system housed in the National Trust's intranet. The day-to-day operations of the sites and the main office are now electronic and are stored either on the individual employee's hard drive or in a shared drive in a folder for the site. The scanned documents are also found on this shared drive. The sites and the central office have not scanned any visual materials or do not have any audio files in the collections at this time.

There is standardization among file formats because they only have certain types of files (.jpeg and .pdf files for the collections materials).

Current Organization of Digital Materials

The National Trust currently has three locations for their digital materials (collection photographs and documents): Re:Discovery Proficio, individual drives and a shared drive.

The materials in Re:Discovery can be accessed by the staff and volunteers at the sites, and by the staff at headquarters. Their Re:Discovery is divided between the Collections module and the Libraries and Archives Module. The Collections module contains the records for the 3 dimensional collections objects held and displayed in the sites. The Libraries and Archives Module contains the records for the paper collections. There is some overlap between the two modules with books, as those are either in the Collections Module or in the Libraries and Archives module. Each type of module allows for the attaching of photographs and documents.

The materials in the individual drives and on the shared drive are documents referring to the collections as well as the normal running of each site and sometimes photographs. These are organized into folders, usually by site, but there are folders that are more subject-specific. Within the site folders there may also be folders labeled by the year and the name of a current or former staff member with photos of a site. There appears to not be a larger organizational structure governing where certain types of files are kept and where backup copies are kept.

There is no inventory that provides an overview of all of the materials, what they are, and where they are located. A central problem, then, of maintaining these files over a long period of time is not knowing if there is anything missing. Due to having files spread over multiple locations, there is no standardized way of ensuring data security through file fixity or through editing permissions. There are different levels of access within Re:Discovery, which does provide security and back-up of the materials. There is also a log of when the last edits were made to the digital records within the system. There is also not a document that details who has which editing and removing permissions for digital materials or set way of determining where files are stored and which site-specific files are put in the shared drive.

Staff Response to Digital Preservation

The staff at the national headquarters are committed to find a way to centralize their digital resources but there are issues with implementation at the sites, especially the smaller sites. Their hope is to create a preservation plan that can be followed by a collections staff member (if there is one) at an individual site, but especially by interns and volunteers at the sites. Few of the sites have dedicated collections staff, so a primary goal and function of the preservation plan is that it can be implemented by interns and volunteers. To be effective, it should not require specialized technical knowledge, either with computers or with collections management. There have been large-scale standardization workshops with the site staff in the past (a previous focus was on the standardization of the finding aids created at the different sites) and the organization/centralization of digital materials could be the focus of a future workshop. There also is not a smaller organizational structure among the sites, except for the distinction between the Stewardship, Co-Stewardship, and Affiliated Sites mentioned above. The primary goal for the staff at the headquarters is to have everything in a centralized location, rather than trying to find something by looking in multiple places.

Future Collections Impacting Digital Preservation

The collections staff do not anticipate getting more materials in differing file formats in the future, other than if the site staff begin to create videos of events at their site. There is the expectation that as more materials are added to the collections at the sites, there will be more .pdf documents and .jpegs of collection items. These new materials would be integrated into the existing file structure and uploaded to Re:Discovery Proficio.

Human Resources and The Use of Technology

The national headquarters has collections staff, include a collections director and a fellow, as well as interns on a fairly regular basis. The sites have different staffing structures due to the differences in the size and budget of the sites. Not every site has a collections manager, so anyone who would be working on digital preservation at these sites would be a volunteer or intern. This necessitates an easily followed and clear digital preservation plan that would integrate smoothly into day-to-day operations. On the technical side, there is a precedence of the national organization helping smaller sites by buying a scanner for the site. This sort of support could be used to buy external hard drives or cloud-based storage.

Preservation Plan

User Needs Gathering Phase

Goal: This phase is meant for the national headquarters to understand the needs of the sites so that the preservation plan can integrate those needs with the needs of the national headquarters.

The staff at NTHP headquarters will conduct a short survey to determine what the sites know they need to preserve. This survey should include the files they feel need to be preserved, why they need to be preserved, what files they think headquarters needs to have access to, the number of staff people they have, and any resources for digital preservation that they have. This process needs to be a conversation between the individual sites and the headquarters so that both sides get the information and the support they need. It is possible that their needs will be at odds; there may be files that they do not want to share with the national organization (in the case of the Co-Stewardship sites), but that are necessary for legal reasons for headquarters to have access to. This process will also help the sites feel as though they have a role and a voice within the process, and will prevent the national organization from implementing unhelpful policies.

Education and Training Phase

Goal: Provide the individual sites with support so that they have as much information before and during implementation that will make the process go smoothly.

Survey results and detailed workflows will be distributed to the sites as they prepare to manage and preserve their digital material. Any questions that sites have will be answered before the implementation date so that misinformation does not result in an accidental mishandling of files. Answers to questions that may be relevant to multiple sites or were asked by multiple sites should be widely distributed among the sites. Any technical or specific processes should be described in detail, including screenshots if necessary.

Pilot Program Phase

Goal: This phase is meant to test the policies and steps deemed necessary by the national headquarters and the individual sites.

The plan steps defined in the Implementation phase should be taken on by a few sites to begin with to test their effectiveness and the sites' ability to implement them. The sites that are admitted to the Pilot Program should be a representative sample of Stewardship and Co-Stewardship sites, as well as sites with larger and smaller staffs. This will give headquarters an idea of where the plan may need to be adjusted or where extra support might be needed. During the Pilot Program and the Implementation Program, different scenarios for data loss, such as server loss, natural disaster or deletion, will be described as a way of showing the importance of these processes. This program will be conducted in conjunction with the Education and Training Phase. Education and Training will occur before sites implement the steps and will continue through the process as new pieces are added.

Implementation

Goal: The purpose of the Implementation Phase is to begin preserving and organizing the digital files held by the national headquarters and the sites.

After the pilot program has been completed, the training will be given to the rest of the Stewardship and Co-Stewardship sites. Feedback from the sites should be solicited at multiple points during the process to ensure that the sites are getting enough support. If a site cannot complete the minimal plan, then extra support should be allocating, including looking for grant funding.

Minimal Plan

Goal: To gather information on the files that currently exist and organize them so that the national headquarters and the individual sites know what they have.

Audience: These steps need be implemented both at Headquarters and at each of the individual sites.

File Management

1. Going forward, all files will be named and organized according to an established file plan. The names of the files and their folders will be standardized, as will the folder names. This will prevent any confusion in the future about what information is contained within a folder or file.
2. Files that have been scanned for researchers will be integrated into the file management system and into the file preservation plan as they are being scanned.

3. The file management plan will be a .pdf and will be pinned to the top of the shared file directory. This will allow the regulations to be in any easily accessible place so that it can be found and followed.
4. A File Migration workflow will be established to migrate important files from their locations on individuals' drives to the shared drive so that no important documents are lost. It will mirror any already existing plans for accession and preservation of paper documents. If a file would not be kept for long-term use as a paper document, it should not be kept long-term in a digital form.
5. On a pre-determined day once a year, a few hours should be devoted to catch any files that have entered into the system and were not arranged according to the file management plan. This can be implemented using a group reoccurring Google or Outlook Calendar event to remind staff members that the file management system needs to be maintained.

Storage and Geographic Location

1. The starting point for this section is an inventory of all files and their locations. This is a labor-intensive project but very necessary. Without an inventory it will be difficult to determine what needs to be migrated and even what all the main office has on their computers and may need from individual sites. The inventory will require at minimum the file name, the folder location, the file size and a controlled-vocabulary category list (ie. legal documents, donor records, etc.) The categories can then be grouped by risk (low, medium and high) which will determine which materials must be preserved in other steps and sections of the implementation plan.
2. Files that are not in the correct location need to be moved to the correct location and have that location updated within the inventory.
3. All files determined to have high level of need will be copied and saved to an external hard-drive. Any time static document with a high-risk is added to the collection it will be added to the external hard-drive. Any files that have a high priority and are being actively edited will not be added to the external hard-drive until they become static. A second copy in a separate location will prevent data loss in case of server damage. Storing files on an external drive takes less time than uploading documents and adding them to the individual collection spaces on Re:Discovery Proficio.
4. Documents on personal drives that do not have great institutional value will not be included in this inventory. These documents should be migrated into the inventory as they gain institutional relevancy. This is similar to retention schedules of paper documents, but would apply to digital materials. Individual employees can maintain their own inventories for their personal use.

Fixity and Data Integrity

The basic level of file fixity will be accomplished through the inventory described above. File fixity is essentially whether the file has stayed the same on a bit level, ie. that the file is identical to an earlier version and no data within the file has been corrupted or deleted. The inventory will standardize where files are kept and also give the organization an idea of the files that are in their system. This prevents files from being lost because no one knew they existed and is the first step in monitoring files.

Information Security

The minimum suggestion for this section is to create a document that lists the editing permissions for the different categories. Site collections managers should have control over their own files and should share permission with the relevant staff at headquarters. This will show headquarters and site staff what files they should have control over and what is not meant to be shared.

File Formats

Only accept and create files in .pdf (for documents) and .png (for images). A standard for files will help the larger organization track obsolescence and migrate file types if necessary. These file types are used frequently in standard practices and will not require a lot of migration. Documents within the inventory and in the external hard-drive will be static and will not be edited often, so the stability of a .pdf is desirable. The main migration will occur when moving documents from a .jpeg or a .doc file format. These are not as stable and are harder to preserve.

Moderate Plan

Goal: These steps build on the steps in the minimal plan to preserve the files at a higher level.

Audience: These steps should be implemented at Headquarters and are recommended for the individual sites.

File Management

1. Files that have not been used within a pre-determined amount of time and have been designated as not being institutionally relevant for long-term use will be deleted, the same as de-accessioning processes for paper documents. This will ensure that unnecessary files are not taking up valuable server space.
2. File management upkeep will happen twice a year for each staff member to catch files that may not have been integrated into the file management system mentioned in the minimal plan.

3. Historically-relevant photographs should be downloaded from the shared asset-management system twice a year to ensure that these images are preserved. The images should be downloaded with the highest resolution as a .tiff file and considered among the medium level of risk for a file.

Storage and Geographic Location

1. The next level would be to migrate the documents with the highest preservation priority to Re:Discovery Proficio into the each collection's individual record. This will give the documents a tertiary location and a third copy.
2. Documents that are static that have a medium level of need will be added to the external hard-drive and will be added of the list of files that will be checked for fixity.

File Fixity and Data Integrity

1. To make the documents outside of Re:Discovery more secure, yearly checks of file sizes will determine if a file has become corrupt. If the file size has changed, the file should be replaced with a copy either from Re:Discovery or the external hard-drive.
2. The editing history for each file (date, who edited it and the new file size) should be recorded in a spreadsheet to not mistakenly report that there is a data issue. This editing history applies only to the documents that have a high level of risk. This is implemented to track file changes to ensure that the files are the same files that were originally saved.

Information Security

The sites and national headquarters should create password protected folders for documents that have editing restrictions within the shared drive based on the editing permissions created in the Minimal Plan. This will protect the files from employees at other sites that should not have access to the other sites' information. Folders for individual sites should only be accessible by that site's managers and headquarters' staff. As staff leave, these restrictions need to be migrated to another staff member before the previous staff person leaves and then migrated again when a new person is hired. These steps should be included in any of the workflows for bringing in new employees or when other employees leave.

File Formats

The sites and the national headquarters should create and accept files in .tiff (still image) and .pdf/a (document) and migrate files that are .png or .pdf to the other standards. These file types are widely used and are more stable than a .png

or .pdf. The suggested migration ensures that there are not multiple file types existing on the drives that can make preservation more difficult because there is no one standard file.

Aggressive Plan

Goal: The goal of this plan is to preserve the files at a much higher technical level and to standardize legacy file names.

Audience: Suggested for Headquarters, advised for the sites. The File Management and Storage suggestions are of the highest priority.

File Management

1. Legacy Files (files created before the file management system was created) will be re-named according to the created file management process. This will make the files more easily organized and recognized within the shared and personal drives.
2. File Management audits will happen once a quarter so that there is not a large backlog of files that are not correctly managed.

Storage and Geographic Location

1. A fourth copy of the high-priority material will be copied to another external hard-drive and swapped with another site. Sites should swap their hard-drives yearly and Headquarters swaps their hard-drive a site. A much higher level of storage preservation is to have a copy of high-risk files in another geographic location in case of a natural disaster. An easy way of accomplishing this for the individual sites is to pair them during the Education and Training phase and have the two sites swap hard drives so that they both have data in a separate geographic location.
2. The medium level need documents will be added to Re:Discovery Proficio. Proficio gives the files a tertiary location and another copy, will also providing data integrity with little staff time needed.
3. Static low-level documents will be added to the external hard-drive. These files may not have as much long-term value as the high-risk documents, but they still may be institutionally relevant and should be preserved.

Fixity and Data Integrity

1. Fixity of high priority files will be monitored using AVP's Fixity tool. AVP's tool monitors data integrity at a much higher level than just looking at the file size. A file might become corrupted but the file size may not change and staff would not realize that there was something wrong. Fixity can determine whether each byte of the original file is still present and therefore if any of the data has been compromised.

2. Fixity of medium priority documents will be added by monitoring the file size of the documents. Monitoring the file size will help staff catch if data has accidentally been deleted or changed in the file.

Information Security

Create a detailed log of changes made to each file in the file fixity spreadsheet. This will help distinguish between intentional changes to files and changes that were the result of data loss. For example, if a file size changes because a paragraph was deleted because a restriction is no longer relevant, the change in the file size could make staff believe that data was accidentally deleted.

Among these recommendations, those under File Management and Storage and Geographic Location will provide much more file security and should be implemented before the suggestions for Fixity or Information Security.

Digital Preservation Policy

Importance of Digital Preservation

The digital materials held by the National Trust for Historic Preservation have historical and legal value. This value means that those materials need to be preserved in their original digital format or through scans of the analog materials. Preserving these files will allow future NTHP employees or site managers to have access to documents that can provide context and documentation for practices and accessioning of materials within the sites' collections. This will prevent questions from going unanswered due to a lack of information and will provide necessary documentation for the sites and the national organization.

Scope of Digital Preservation

The materials held by the National Trust for Historic Preservation that are subject to digital preservation practices by the collections staff at the national headquarters and the individual Stewardship and Co-Stewardship sites are documents and images are the related to the collections held by these entities. These materials are legal documents relating to ownership of materials, documentation of repair or purchase, and materials created by the collections staff in relation to the sites. Photographs taken of the sites or of materials in the collection are also covered under this scope. These materials are not meant to show the institutional history of the NTHP or of legal practices not covering the collections.

Purposes of Digital Preservation

The over-arching purpose of digital preservation is for NTHP to maintain and preserve digital copies of collections materials for legal and historical purpose. These materials also include photographs of the sites and the collections within. Simply preserving the documentation is not enough, however, for the materials need to also be centrally located within the shared network and cloud-based resources. This will allow collections staff at the national headquarters to access documents that they may need that are held by the individual sites or the reverse, as both parties are involved with the collections. Following prescribed digital preservation practices will provide the foundation to support these purposes.

Digital Preservation Practices

The guiding principles of this digital preservation initiative focus on centralizing access and creating more documentation around document types and locations. Documents are meant to be accessible by multiple parties, with the goal of having the most valuable documents contained within a cloud-based provider. Collections documents that become static will be integrated into the digital preservation workflow and continual monitoring of active file management will enable headquarters and site staff to easily identify materials that they need and anything that is

no longer active. This is done through file inventories and documented file naming systems. Standardization of these practices will promote centralization and ease of use for the sites' staff and the staff at the headquarters. Centralization and preservation will also aid in the description of the documents that are digitally available through inventories and file directories, making the files more accessible and useful to NTHP staff and site staff. Specific file types (.pdf and .png) will be used to prevent file obsolescence and make monitoring file types for potential obsolescence easier. These practices privilege ease of access and a standard across the organization.

Challenges of Digital Preservation

The challenges of preserving these materials for the national headquarters are in the unique relationships that the national headquarters has with the individual sites. The individual sites maintain their own collections and order their day-to-day operations with degrees of oversight by the staff at headquarters. This can result in an imperfect implementation of methods that could result in data being lost by some sites. Full implementation of minimal digital preservation strategies will result in a lower risk of data loss, but only if it implemented across the sites. Another challenge is maintaining a file management system to help with identifying files that need to be preserved, both at headquarters and at the sites. The full documentation of practices will help the whole organization move beyond these challenges to preserve their digital materials into the future.

Digital Preservation Participants

Given the unique structure of the NTHP, the work of digital preservation is shared among staff at the national headquarters and among the individual sites. The materials collected and maintained by these different groups overlap and are used by both parties in various ways. This necessitates clear and easily adopted digital preservation practices so that there are no holes in the practice of digitally preserving material. These participants are also tasked with the continual process of digital preservation, rather than simply considering it to be a singular act. This mindset accompanied by defined workflows for preservation will work together to preserve the materials of the National Trust for Historic Preservation. The site staff and the headquarters staff work together on digital preservation, bringing all of the resources available to aid in the preserving of these files.

Resources

Altman, M., Bailey, J., Cariani, K., Gallinger, M., Mandelbaum, J., & Owens, T. (2013). NDSA Storage Report: Reflections on National Digital Stewardship Alliance Member Approaches to Preservation Storage Technologies. *D-Lib Magazine*, 19(5/6).

Library of Congress Recommended File Formats: <https://www.loc.gov/preservation/resources/rfs/>

Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). *The NDSA Levels of Digital Preservation: An Explanation and Uses*. IS&T Archiving, Washington, USA.

Schumacher, J., Thomas, L. M., VandeCreek, D., Erdman, S., Hancks, J., Haykal, A., Spalenka, D. (2014). *From Theory to Action: Good Enough Digital Preservation for Under-Resourced Cultural Heritage Institutions (Working Paper)*.