

# DIGITAL PRESERVATION REPORT: SURVEY, NEXT STEPS, DRAFT POLICY

*PREPARED BY TINA HABASH*

*FOR THE ARCHEOLOGY PROGRAM OFFICE*

*PRINCE GEORGE'S COUNTY PARKS AND RECREATION*

*MARYLAND-NATIONAL CAPITAL PARK AND PLANNING COMMISSION*

*DECEMBER 2018*

# TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>2</b>
<b>SURVEY .....</b>	<b>3</b>
OVERVIEW .....	3
SCOPE OF HOLDINGS.....	3
CURRENT MANAGEMENT OF DIGITAL HOLDINGS.....	4
PERCEPTION OF THE STATE OF DIGITAL CONTENT.....	5
RESOURCES .....	5
<b>NEXT STEPS PRESERVATION PLAN .....</b>	<b>6</b>
OVERVIEW .....	6
STORAGE AND GEOGRAPHIC LOCATION .....	6
METADATA.....	8
FILE FIXITY AND DATA INTEGRITY .....	8
INFORMATION SECURITY .....	9
FILE FORMAT .....	9
FUTURE STEPS.....	10
<b>DRAFT POLICY .....</b>	<b>11</b>
INTRODUCTION .....	11
PURPOSE.....	11
SCOPE .....	11
STANDARDS .....	11
STORAGE.....	11
DATA INTEGRITY .....	12
ACCESS AND CONFIDENTIALITY.....	12
METADATA.....	12
FILE FORMAT .....	12
REVIEW.....	13
<b>REFERENCES .....</b>	<b>13</b>

## EXECUTIVE SUMMARY

The Archeology Program Office of the Prince George's County Department of Parks and Recreation was established in 1988 to excavate, preserve and protect archeological sites in county parks. It is part of the Maryland-National Capital Park and Planning Commission. As part of its mission, the program curates millions of artifacts, and over the years, related documentation has been created in various formats and on disparate media. The program's goal is to have all digital content centralized on one shared network which is currently in development. This report is comprised of a collection survey, recommended next steps and a draft preservation policy.

A survey of the office's records revealed that the majority of digital content is stored locally on five desktop computers, two laptops, a back-up drive of files from former staff, approximately 300 compact disks and approximately 700 3.5" floppy disks. This content is currently hard to find, on inaccessible media, or in unusable formats. All artifacts and documentation are stored in the same facility, a house in Upper Marlboro, Maryland. The perception from the archeology staff is that there is digital content that will add either known or potential value to their mission.

Recommendations are based on the National Digital Stewardship Alliance's Levels of Digital Preservation. Recommended next steps are:

- Maintain a second copy of all digital content in a second geographic location in order to prevent loss from equipment failure or a catastrophic event
- Set up an unnetworked workstation to copy over files from floppy disks
- Develop a naming system to organize all content within the shared directory
- Install fixity software to monitor the data integrity of files and to generate a manifest of files
- Separate historical from working files in order to prevent loss or changes to historical files
- Work with the department's IT staff to monitor obsolescence of storage media and file formats

This report also includes a draft policy which will serve as a basis for establishing a sustainable preservation plan for the program's valuable cultural heritage collection. This policy should be reviewed annually to evaluate compliance and to amend as needed.

# SURVEY

## OVERVIEW

The Archeology Program Office of the Prince George's County Department of Parks and Recreation was established in 1988 to excavate, preserve and protect archeological sites in county parks. It is part of the Maryland-National Capital Park and Planning Commission. As part of its mission, the program curates millions of artifacts, and over the years, related documentation has been created in various formats and on disparate media. The office has been located in a house in Upper Marlboro, Maryland, since around 1999/2000 where all artifacts and documentation are kept. It is currently staffed by three archeologists who are the primary users of the documentation.

The goal is to have all digital content centralized on one shared network which is currently in development. Ideally physical copies would be digitized. This content would likely have variable levels of access; some files could be made available to the public through a research portal while some would remain for internal use only.

## SCOPE OF HOLDINGS

Documentation is primarily in the form of digital and/or physical copies of reports, catalogs, slides, print photographs, negatives, drawings, maps, and videos. The following includes a description of digital content and some of the physical holdings since there is interest in digitizing them.

### *DIGITAL STORAGE MEDIA*

Digital content currently resides on five desktop computers, two laptops, and a back-up drive of files from former staff. The three current staff people also keep back-up drives at their desks. Content is primarily kept and backed up locally, but the staff are just beginning to test a shared network.

Additionally, files are also stored on approximately 300 compact disks and 700 3.5" floppy disks. This may be an underestimate as some disks are kept with hard copy reports. The disks are labeled but not necessarily enough to discern the contents. Approximately 600 of the 3.5" floppy disks are kept in boxes with a consistent labeling format. Staff have an external floppy drive provided by an offsite IT department to read the 3.5" disks but the drivers have yet not been installed.

### *REPORTS AND CATALOGS*

While the exact contents of the storage media are not known, staff anticipate that many reports and catalogs were originally written with WordPerfect. More recent reports are written using Word. Versions will vary. Catalog data was originally kept using Paradox relational database software and staff currently

have no means of reading these files. Current catalog information is kept on both Excel and Access. There are approximately ten shelves of archeological site reports (25 linear feet) and four filing cabinets of site-related files.

#### *PHOTOGRAPHS, CONTACT SHEETS, NEGATIVES, AND SLIDES*

Digital images are saved as JPEGs. These are both born-digital and digitized from a portion of the photograph and slide collection. There are two shelves (approximately 5 linear feet) of binders containing photographs, contact sheets, negatives and slides. Descriptive information is written on the dividers and on the backs of some photographs. There is also a flat file drawer of matted photographs.

#### *MAPS AND DRAWINGS*

Mapping files have been created using CAD (computerized aided drafting), and more recently, GIS (geographic information systems) technologies. In addition, there are fourteen flat file drawers of maps and drawings.

#### *VIDEO*

There are approximately twenty VHS tapes the contents of which are unknown.

#### **CURRENT MANAGEMENT OF DIGITAL HOLDINGS**

Digital content on floppy disks is inaccessible at the moment. Current staff are relying on institutional memory and lengthy searches to find files from former staff stored on a back-up drive. This is based on an understanding on what the person's role might have been and when they worked there. At times, searching based on file extension can provide a hint at the content.

Most digital content is still stored and backed up locally, but staff are testing a new shared drive with directories allowing different levels of access for onsite staff and the rest of the department. Within the directory for onsite staff, subdirectories currently relate to specific archeological sites, but the exact organization has not yet been determined since content can be classified in more than one way. Implementing this shared network offers an opportunity to integrate disparate digital holdings and would facilitate creating a back-up copy in a different location.

## PERCEPTION OF THE STATE OF DIGITAL CONTENT

The perception from the archeology staff is that there is digital content that will add either known or potential value to their mission. This content is currently hard to find, on inaccessible media, or in unusable formats. Identifying and moving these files to one drive would: save staff time by making the files easier to find, save space currently taken up by heterogeneous storage media, eliminate duplicates, offer an opportunity to keep all information about specific sites or artifacts under one subdirectory, and prevent irreversible loss.

## RESOURCES

The Archeology Program Office will be able to benefit from the technical expertise of the Prince George's County Parks and Recreation IT department, which provides and maintains the staff's equipment. In addition to the equipment mentioned above and the shared network, the staff have two scanners that have been used to digitize slides, files and print photographs in the past.

The IT department also regulates information security, one of the key preservation activities that will be addressed in the coming preservation plan. This does, however, limit the types of software staff can download themselves – the drivers for the external floppy drive mentioned above, for example. Additionally, they may serve as a resource in implementing another of the key activities – data integrity checks.

The program also benefits from a small but dedicated staff who are vested in maintaining the integrity of this digital content. The effort is being spearheaded by the Assistant Archaeology Program Manager. The level of effort will likely be limited to a couple of hours per week until a student intern or volunteer can be hired and trained to do the work.

# NEXT STEPS PRESERVATION PLAN

## OVERVIEW

This report outlines recommended next steps to preserve their digital content following a framework designed by the National Digital Stewardship Alliance (NDSA). NDSA's Levels of Digital Preservation presents a series of recommendations based on five areas of concern for digital preservation – storage and geographic location, file fixity and data integrity, information security, metadata and file format ([Phillips, 2013](#)). Level 1 recommendations relate to the most urgent activities need for preservation and serves as a prerequisite to the higher levels.

## STORAGE AND GEOGRAPHIC LOCATION

**Establish a central storage system for all digital content and maintain at least one complete copy stored in another location.**

Digital content is currently stored locally on five desktop computers, two laptops, a back-up drive of files from former staff, approximately 300 compact discs and about 700 3.5" floppy disks. Staff have been working on a shared drive to organize content in one place. While staff are backing up their own drives locally, there is no complete copy of all digital content.

Following NDSA's Level 1 recommendation, staff should move all files off disparate media and into a storage system and create a complete copy of files that should be stored in a different location. This serves a number of purposes:

- Transferring to a new storage system will guard against potential loss of data on old media.
- Minimizing number of storage locations will facilitate data integrity monitoring.
- Creating a complete copy will provide a back-up in case files are corrupted or lost.
- Storing a copy in a different location will guard against loss specific to one location such as damage to equipment as a result of severe weather or a catastrophic event.
- This process is also a necessary first step in achieving a goal for the program, which is to have all files organized and accessible without extensive searching.

1. Create a complete copy of what is accessible right now.

The development of the shared network in coordination with the program's IT department will likely serve as their established storage system. A portion of files are currently inaccessible on floppy disks. However, the bulk of digital content is stored locally on various hard drives. Any content that is currently

accessible on hard drives should be copied and stored in another location to safeguard against loss. See “File Fixity and Data Integrity” below for checking data integrity and creating a file manifest. In the short term, storage can be on an external hard drive that is kept in another location in a secure place. Alternatively, staff could use a cloud storage service such as Dropbox which would have the added benefit of storing files offsite. Once files are integrated into the shared drive, a copy can be created from this set of content. Staff can work with the IT department to see if they can use a back-up storage system that is already in place with other departments.

## 2. Establish a file structure in the shared directory

Staff can begin integrating current files that they work with into the networked drive in order to develop a workable structure that will be practical for finding files. Historical documents can then be organized into that structure. Once a structure and naming system has been established, staff should document and follow this structure. This can be done concurrently with the following step.

## 3. Copy files from compact disks and 3.5” floppy disks using an unnetworked workstation

It will likely take a while to sift through decades worth of files. Given the urgency of potential loss from old media and a potential risk of viruses from disks used by former staff, copy over all the files from compact and floppy disks first onto an unnetworked drive and run a virus scan before combining with current files.

a. Drive space needed: Assuming approximately 700 MB of data per CD and 1.44 MB per floppy disks, there is approximately 211 GB of data if all the disks are full.

b. Install anti-virus software on unnetworked workstation: Speak with the IT department about installing antivirus software so that it can run in the background while copying over files. Establish a schedule to run full virus scans depending on how regularly files are being copied over.

c. Install drivers for 3.5” external floppy drive on unnetworked drive: At the time of the survey, staff had an external drive to read 3.5” floppy disks but lacked the software to run the drive. This step is necessary in order to access any of this information.

d. Install fixity software on unnetworked and networked drives: See the section “File Fixity and Data Integrity” below.

e. Transferring files from CDs and floppy disks to unnetworked hard drive: Currently staff use context such as file extensions, file location, file name, and the author name to find content. Therefore, it’s recommended that they preserve these elements and related descriptive information that may be written on the disks as much as possible until they can sort out all the content. Disks were labeled in a number of ways. The majority of floppy disks were organized in boxes of about 10 disks with labels on



the boxes suggesting that there were groups of related disks. Some disks had vague or no labeling. Related files kept together on disks or boxes of disks should be copied into a directory folder with the descriptive information from the label. This can either be transcribed onto a text file or a picture can be taken of the label and included in the directory.

f. Timing: While each floppy disk may hold a relatively small amount of data, the process of copying over hundreds of disks can be labor intensive. However, during this window, the work is not being backed up. Therefore, consider either setting a short window of time to copy over all of these files, or setting a schedule where small batches of files are copied over from disks and scanned for viruses before being copied to another location. This can either be added to shared drive or another external drive that is stored in another location. Once files have been safely transferred, they can be integrated into the file structure on the shared drive.

4. Set a schedule of back-ups to maintain a complete copy.

## METADATA

The NDSA Level 1 recommendation is to create an inventory of digital content and storage location. Like the digital content itself, the inventory should be backed up and stored in another location. As mentioned in the Section “File Fixity and Data Integrity” below, avp’s Fixity software includes a function to generate a manifest of file paths that will assist with inventory. Staff can maintain this inventory as they continue to develop the file structure on the shared drive.

For file and directory naming, consider creating a controlled vocabulary and syntax to make it easier for staff to find files. This can include specific terms for archeological site names, document type (e.g., site form, report), and a version, year or other modifier (e.g., draft, final) when needed.

## FILE FIXITY AND DATA INTEGRITY

File fixity is a way of ensuring that files have not changed. It is recommended to run fixity checks whenever files are transferred ([Owens](#), p. 110). This will generate an alphanumeric string called a checksum that can be compared before and after the transfer. Changing the content of the file including the format will change the checksum value. If the IT department does not already use fixity software, Fixity is a free tool from avp that can be used to generate and compare checksum values and ensure that all files have been transferred ([avp](#), 2018). The software also generates a manifest of file paths along with the checksums that could prove useful in establishing an inventory of digital content.

The Level 2 recommends virus checking high risk content while Level 3 recommends virus checking all content. Virus checking high risk content is addressed 3b of “Storage and Geographic Location.” Staff should have antivirus software installed at their workstations and run scheduled scans.

Level 3 also recommends checking fixity as fixed intervals to ensure data integrity over time. Consider establishing a yearly schedule of validating fixity. Any corrupt or missing files can be replaced with a copy that passes fixity validation.

## INFORMATION SECURITY

This step will outline who has access to the content and what they can do with it. This will prevent files from getting deleted or changed by unauthorized staff. NSDA Level 1 recommends to identify who is authorized to read, write, move or delete individual files. Related to this, Level 4 in the section on file fixity also recommends that no one is authorized to have write access to all copies. This reduces the likelihood of changing or deleting all copies of one or more files.

Staff have taken initial steps in the process by creating three different directories with different levels of access for their users: one directory for onsite archeology staff, one directory for the rest of the Prince George's County Parks Department, and one directory for Dinosaur Park, another program of the Prince George's County arm of the Maryland National Capital Park and Planning Commission that shares the same workspace.

However, more levels of access may be necessary if only one person in the office is allowed to have read or write permissions. In addition, staff should clearly delineate working files from historical files that should not change. This will help to prevent the document from being changed or deleted. It will also help with fixity validation since working files will likely involved changes in content which will change the checksum value. This can be accomplished by setting permissions to specific subdirectories or to specific sets of files. Document access restrictions and store in a location that all users can access.

## FILE FORMAT

File formats can become obsolete. In some cases, once the format is obsolete the file might not be able to be opened in another format or will not be rendered in exactly the same way. The purpose of this section is to minimize these problems by using formats that are less likely to become obsolete, or that can be effectively rendered in another format. Widely used formats are generally considered to remain accessible because there will be a demand to either keep them accessible or develop a means of migrating them ([Owens](#), 121).

Since some media have not yet been accessed, a current inventory of file formats that have been used is not available. Formats currently in use are jpegs and files generated from different versions of Microsoft Word, Excel and Access. Mapping files are created using GIS (geographic information systems) technologies. Older files have been created using WordPerfect, CAD (computerized aided drafting) software, and Paradox relational database. Staff are currently having trouble with opening Paradox files since the database is no longer supported and cannot be opened using current versions of Access or Excel.

Formats such as jpeg and Microsoft Word and Excel are commonly used, although the latter two undergo regular updates which may render slight changes if a file is opened in a new version. As the files are incorporated into the new directory structure on the shared drive, staff should develop an inventory of formats that they are using, work with the IT department to monitor them for obsolescence, and be prepared to migrate as needed.

## FUTURE STEPS

NDSA Level 2 for storage recommends creating a third copy of the content and Level 4 recommends at least three copies stored in locations with different disaster threats. The Archeology Program Office could combine this recommendation with a means of sharing some of their content. This could be through a subject-specific repository for archeology or something more general like the Internet Archive.

Levels 2-4 address steps to maintain storage media so that files continue to be accessible in the long term. Staff should work with their IT department to document storage used for their shared drive and back-up copies, monitor for obsolescence, and have a plan in place for updating systems.

Staff also expressed an interest in resuming digitization of their physical documentation. Some reports and slides have already been digitized. As a starting point, staff can discuss their experiences with past efforts and lessons learned to establish goals for the program and how this will fit into the file structure that they are creating for current digital content. The Still Image and Audiovisual Working Groups of the Federal Agencies Digital Guidelines Initiative can be a good resource to establish best practices for digitization ([FADGI](#), 2018).

# DRAFT POLICY

## INTRODUCTION

The Archeology Program Office of the Prince George's County Department of Parks and Recreation was established in 1988 to excavate, preserve and protect archeological sites in county parks. It is a program of the Maryland-National Capital Park and Planning Commission (MNCPPC). As part of its mission, the program curates millions of artifacts. Over the years, related documentation has been created in various formats and on disparate media, some of which have become obsolete. This practice has led to content being hard to find and inaccessible.

## PURPOSE

This policy will establish practices for organizing and preserving digital content. Such practices will help save staff time by making the files easier to find and access, consolidate digital content into one more easily manageable system, and minimize loss of irreplaceable cultural heritage materials.

## SCOPE

This policy only covers preservation of digital content although documentation exists in both digital and physical form. Reports, catalogs, photographs and maps were historically created in print but are now born digital. A selection of print photographs and slides have been digitized, and there are plans to continue digitizing physical copies.

Content includes both historical and working files. Historical files are those that are considered in their final form, the content of which should remain unchanged. Working files are those that continue to be updated.

## STANDARDS

This policy draws from recommendations of the National Digital Stewardship Alliance's Levels of Digital Preservation and the Federal Agencies Digital Guidelines Initiative. Documentation follows MNCPPC's Guidelines for Archeological Review ([Phillips, 2013](#); [FADGI, 2018](#); [Prince Georges County Planning Board, 2005](#)).

## STORAGE

The department's IT staff will maintain the storage systems, and monitor storage media to preempt obsolescence and system failure.

All content will be organized within the office's shared drive and backed up nightly.

A complete copy of content will be stored in a different geographic location, either through another departmental storage system or by purchasing cloud services. This will protect content from loss due to equipment damage or a catastrophic event in one location.

## DATA INTEGRITY

Directories within the shared drive will allow for three different levels of access: the Archeology Program Office, all of the Department of Parks and Recreation for Prince George's County, and Dinosaur Park (another Prince George's County program that shares workspace with the Archeology Program Office).

A designated staff person will maintain a list of staff with read, write, move and delete permissions to directories, subdirectories and special files. The list will be kept in the shared directory.

No one staff person will have write, move or delete access to all copies of files.

Historical files will be kept in separate subdirectories from working files, and will have limited write and delete permissions.

Fixity software will be used to create checksums. A designated staff person will compare checksums annually against all historical files, and for all files whenever transferring to a new storage system. This will ensure that files have not been altered or deleted.

## ACCESS AND CONFIDENTIALITY

The Archeology Program Office will explore potential repositories for sharing content with other researchers and with the public while still observing regulations protecting the locations of archeological sites.

## METADATA

Subdirectories and files within the Archeology Program Office directory will follow a naming system developed by staff. A file documenting the naming system will be kept within the directory so that it can remain accessible. All staff creating records will be responsible for following naming guidelines and saving content in the appropriate directory.

## FILE FORMAT

In order to keep files accessible, one or more designated staff people will maintain an inventory of file formats in use and monitor them for obsolescence. Whenever possible, staff will select commonly used file formats and use them consistently. This will help files remain accessible and facilitate migration if

necessary. Currently, images are saved as jpegs. Microsoft Word, Excel and Access are used for text, spreadsheets and relational databases, respectively.

Prior to digitizing hard copies, staff will establish a standard for format and image quality appropriate for the material being digitized.

## REVIEW

This policy will be made available on the department shared drive, and undergo annual review by representatives of the Archeology Program Office and the Department of Parks and Recreation. Review will take place to evaluate compliance and amend the policy as necessary.

## REFERENCES

avp (2018). *Fixity*. Retrieved from: <https://www.weareavp.com/products/fixity/>

Federal Agencies Digitization Guidelines Initiative (2018). Retrieved from: <http://www.digitizationguidelines.gov>

Owens, T. (2018). *The theory and craft of digital preservation*. Baltimore: Johns Hopkins University Press. (2018).

Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA Levels of Digital Preservation: An Explanation and Uses. *IS&T Archiving*, Washington, USA. Retrieved from: [https://ndsa.org/documents/NDSA\\_Levels\\_Archiving\\_2013.pdf](https://ndsa.org/documents/NDSA_Levels_Archiving_2013.pdf)

Prince George's County Planning Board, Maryland National Capital Park and Planning Commission (2015). Guidelines for Archeological Review. Retrieved from: <http://www.pgparcs.com/DocumentCenter/View/423/Guidelines-for-Archeological-Review-PDF>