



National Park Service
U.S. Department of the Interior

Natural Resource Stewardship and Science

GRSM Standard Operating Procedure: Electronic File Archival

Suggested Citation: National Park Service. 2016. GRSM Standard Operating Procedure: Electronic File Archival. Version 2016-07-20. National Park Service, Great Smoky Mountains National Park. Gatlinburg, Tennessee.

Contents

Purpose.....	3
Scope	3
Introduction	3
Relevant NPS and GRSM Policies	3
Project Archival.....	3
Archival.....	4
CD and DVD Handling Guidelines.....	7
Roles and Responsibilities.....	7
Contact Information	7
Attachments	8
References.....	8

Version History		
Version	Update Date	Changes
20160719	20160719	Revised draft to appropriate format

Purpose

This document provides basic guidance for creating archives intended for long-term storage and distribution of park data.

Scope

This document pertains to electronic/digital data sets. The preservation of hard copy archives are presented in a separate appendix: GRSM Record Management Standard Operating Procedure. Digital archives refer to non-changing data sets and information products that will be write-protected to ensure against inadvertent editing. These guidelines apply to all GRSM staff, cooperators, contractors, and other partners who are likely to generate data. The NPS Project Lead in collaboration with the Data Manager and GIS Specialist has primary responsibility for ensuring adequate archival of project data.

Introduction

Security and longevity are two of the critical components of information management. Data longevity can be enhanced through proper documentation (e.g. metadata) and by maintaining the data in currently accessible and interpretable formats. Digital data must also be maintained in environments that protect against loss from a variety of factors including: improper storage conditions, hardware failure, software obsolescence, storage media deterioration and natural disasters (e.g. flooding and hurricanes). This document describes how GRSM maintains and protects its digital assets to ensure their long term security and longevity.

Relevant NPS and GRSM Policies

- GRSM will keep current with all Department of Interior (DOI) software platform migrations. Currently the standard office software platform is Microsoft products (e.g. Microsoft Office) for all DOI Agencies (DOI, Assistance Secretary for Policy, Management, and Budget, Findings and Determination, September 12, 2002).
- GRSM will update and maintain datasets no more than two versions behind the current software version – or will store the data set in American Standard Code for Information Interchange (ASCII) format, complete with data and file documentation.

Project Archival

Electronic files should be consolidated and packaged for electronic archival when a protocol reaches periodic milestones (e.g. annually at the conclusion of each field season) or at the end of a project (for any short term projects). Generally, the pertinent information is contained within an organized project folder therefore will not need much manipulation prior to archival.

The following items should be bundled together, if applicable:

- Certified data stored in a relational database housed in a Microsoft Access database that follows the current NRDT standards (NPS, 2007);
- Clear distinction should be made between draft and final files—this can easily be accomplished by using the Directory Structure SOP.
- Raw data in its native format— Raw data can come in many forms such as continuous monitoring data downloaded from a logger or data from an outside researcher in its original format before being processed for migration into one of the park's database. Archiving these kinds of data is a crucial step in data security and integrity as the original format can never be recreated once processed in any way.
- An xml metadata record explaining data and the database structure in which it is stored;
- A complete copy of the Protocol and Standard Operating Procedures (SOP) which governed data collection and management;
- A document specifically detailing any restrictions on the use of the data based on the sensitivity or commercial value of the data when appropriate;
- "README.TXT" file describing each of the elements.

Other items to archive, either in project bundle or otherwise:

- Photographs – These should reside as a sub-folder or as a zipped file and if included will require their own xml metadata records;
- Digitized copies of field sheet - These should reside as a sub-folder or as a zipped file;
- Project metadata—applicable especially for projects conducted by non-NPS researchers; GRSM Study ID as well as project contacts should be included in the bundle.

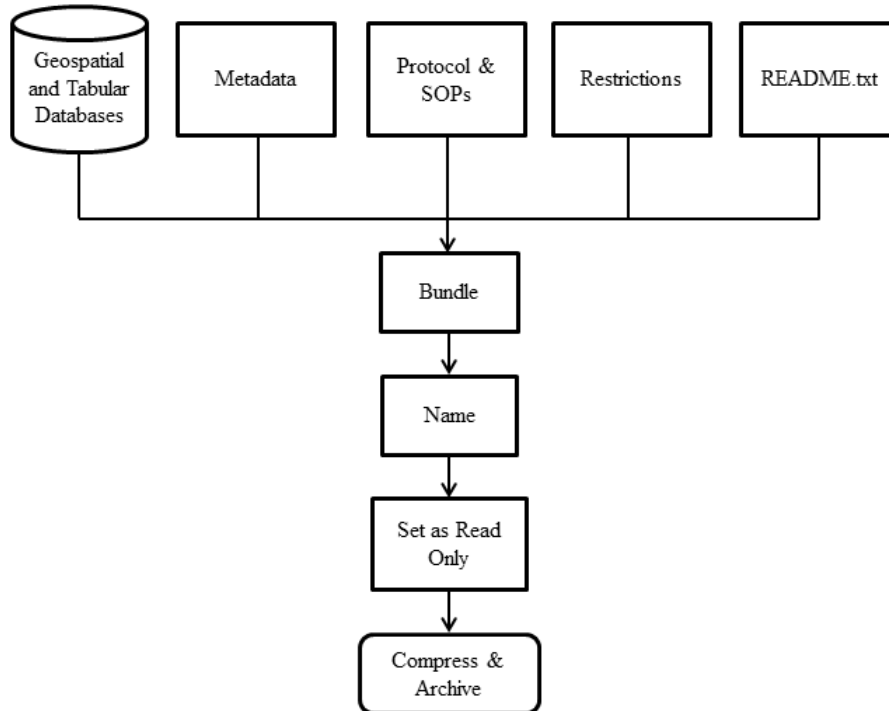


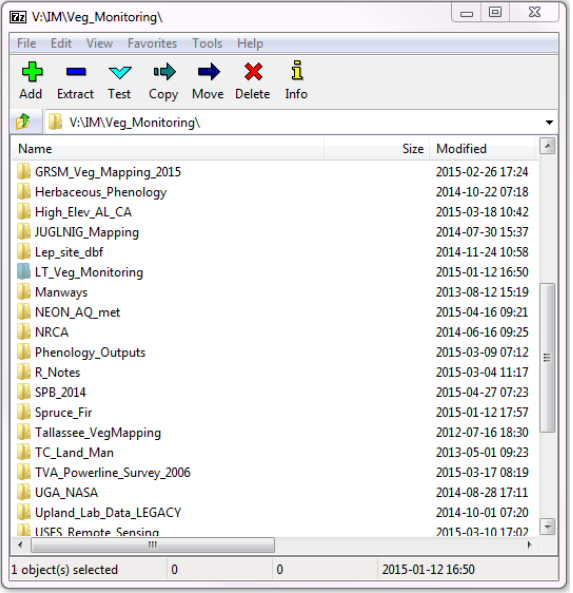
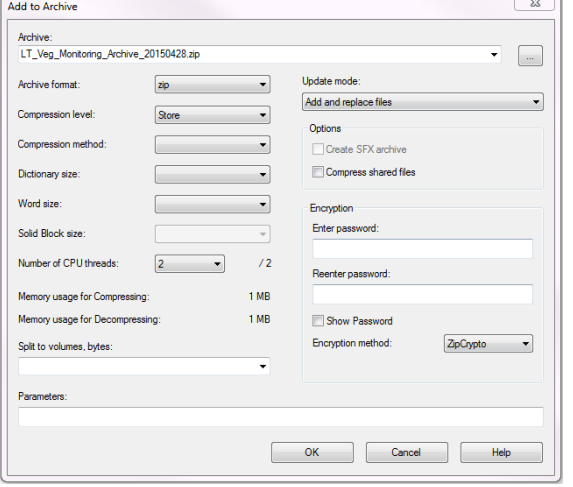
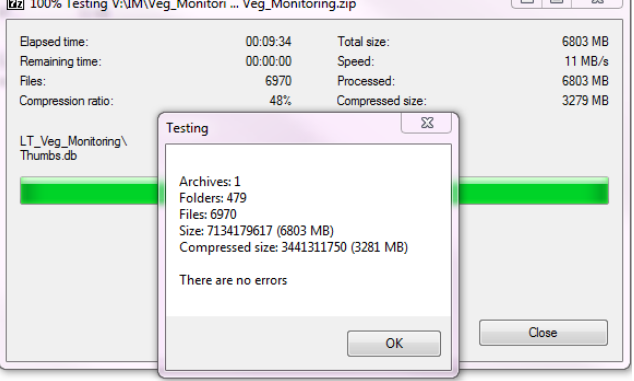
Figure 1. General steps for bundling data into a zipped archive

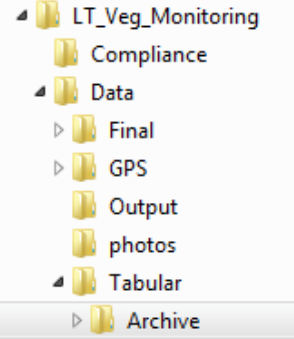
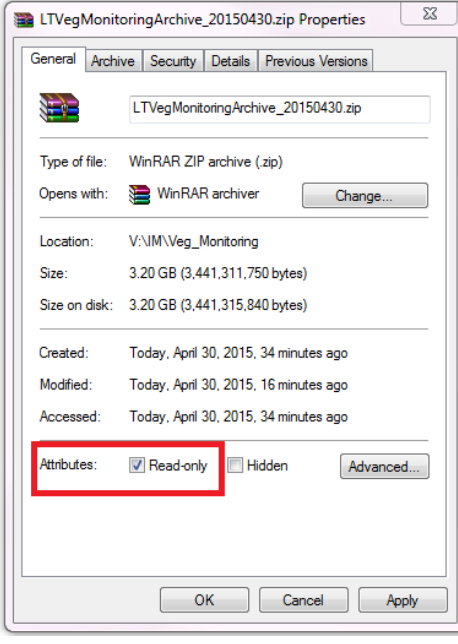
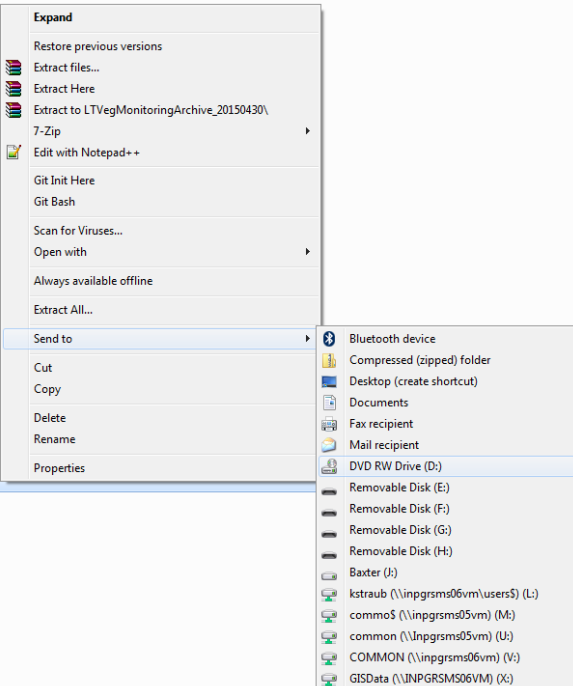
Archival

The 7-Zip archiving program or some similar program will be used to create a compressed archive file containing the documents mentioned above.

Archival Steps:

Document archival by using the GRSM GIS and Data Management work order system to track the task (http://inpgrsms04tc/gisworkorder/) Use the 'DM: Archive' work order product Record the location of the folder to be archived in the appropriate field Use the comment field to capture your reason for archival	
Create an electronic record of the files in the folder being archived by utilizing the cmd.exe utility. Include the resulting text file in the project folder so that the files archived are searchable.	Example: Open up command prompt Type X: then enter Change directory to the project folder: cd X:\GIS_Projects\129_Property

	<pre>X:\GIS_Projects\129_Property>dir /S > X:\GIS_Projects\129_Property.txt</pre>
Navigate to the 7-Zip program and browse to the folder you plan to archive. When you have selected the folder, click the green plus sign, “Add” to create the archive.	 The screenshot shows the 7-Zip File Manager window with the directory V:\IM\Veg_Monitoring\ selected. The file list includes folders like GRSM_Veg_Mapping_2015, Herbaceous_Phenology, High_Elev_AL_CA, JUGLNIQ_Mapping, Lep_site_dbf, LT_Veg_Monitoring, Manways, NEON_AQ_met, NRCA, Phenology_Outputs, R_Notes, SPB_2014, Spruce_Fir, Tallassee_VegMapping, TC_Land_Man, TVA_Powerline_Survey_2006, UGA_NASA, Upland_Lab_Data_LEGACY, and USFS_Remote_Sensing. The status bar at the bottom indicates '1 object(s) selected'.
Name and format the archive. Archive files will be named using the GRSM's file naming convention (See File Naming Conventions SOP) Archive Format: zip—“7z” is a proprietary archive file format therefore, using “ZIP” file format files can be extracted regardless of end user's software. Compression level: Default (Store)	 The screenshot shows the 'Add to Archive' dialog box. The 'Archive' field is set to 'LT_Veg_Monitoring_Archive_20150428.zip'. The 'Archive format' is set to 'zip', 'Compression level' is 'Store', and 'Update mode' is 'Add and replace files'. The 'Encryption' section is expanded, showing 'Enter password' and 'Reenter password' fields, and the 'Encryption method' is set to 'ZpCrypto'.
Test the archive using the built in testing functionality in 7-Zip	 The screenshot shows the 7-Zip testing window titled '100% Testing V:\IM\Veg_Monitori ... Veg_Monitoring.zip'. It displays progress bars and statistics: Elapsed time: 00:09:34, Remaining time: 00:00:00, Total size: 6803 MB, Speed: 11 MB/s, Files: 6970, Processed: 6803 MB, Compression ratio: 48%, and Compressed size: 3279 MB. A 'Testing' dialog box is overlaid, showing: Archives: 1, Folders: 479, Files: 6970, Size: 7134179617 (6803 MB), Compressed size: 3441311750 (3281 MB), and the message 'There are no errors'.

Save archive folder in the appropriate subfolder (see GRSM Directory Structure SOP) of the project to be stored indefinitely.	
Once saved to the appropriate location on the server, the archive will be set to read-only using Windows Explorer file properties	
Burn compressed archive folder to an archival quality CD-R or high quality DVD-R for long-term storage	

Label the CD/DVD case with the project name, GRSM Study ID if applicable, archival date, number of discs, point of contact/ project lead, name of individual who archive the folder and work order reference number (see step 1). Store the disc in a fire proof safe until project is complete and regular access to the data is no longer required. Refer to Archival of Physical Materials SOP (not sure what this will be named) for guidance on archival CD and DVDs in the park library.	GRSM Long Term Veg Monitoring Project Archived Date: 2015-04-28 Archive By: Kendra Straub POC: Troy Evans, Joshua Albritton, Tom Remaley Work Order #3652
If the project is complete and the project lead no longer needs regular access to the information the disc(s) should be turned over to the librarian-archivist.	

CD and DVD Handling Guidelines

- Use archival media such as MAM-A Gold Archive CD-R if appropriate (for long term storage)
- Always handle discs by the outer edge or the center hole, never touch the surface of the disc
- Never write on or label the disc with markers, pens or adhesive labels, only the disc case will be labeled
- Always transport and store discs with the protective jewel case
- Always store discs vertically (book style)
- Always avoid exposing discs to sunlight, extreme temperatures or humidity

Roles and Responsibilities

Role	Responsibility
Project Manager	Gather pertinent files and materials, ensuring that folder structure and file naming is correct
	Notify DM/GIS when ready to archive once per year
	Work with DM/GIS to facilitate the archival of data originating from outside sources including collaborators
Data Manager	Maintains physical supervision and maintenance of protected storage area/safes at Twin Creeks
	Responsible for archival of permanent data;
	Work with Project Manager to facilitate the archival of data originating from outside sources including collaborators
GIS	Maintains physical supervision and maintenance of protected storage area/safes at Twin Creeks
	Responsible for archival of permanent, park-wide GIS data
	Performs portion of archival of project data that is housed on remote, enterprise databases upon notification
	Work with Project Manager to facilitate the archival of data originating from outside sources including collaborators
Park Librarian	Catalog and store CD/ DVD

Contact Information

Great Smoky Mountains National Park
National Park Service

Gatlinburg, TN

Kendra Straub
Data manager
(865)436-1703
Kendra_Straub@nps.gov

Thomas Colson
GIS Specialist
(865)436-1701
Thomas_Colson@nps.gov

Attachments

None

References

Byers, B. R. (2003) Care and Handling CDs and DVDs: A Guide for Librarians and Archivists. NIST Special Publication 500-252

NPS-ROMN (2007). Archive Format Standards for Network Data, Version 1.0, Rocky Mountain Network-National Park Service, [<http://www1.nrintra.nps.gov/im/units/romn/>], Fort Collins, Colorado.

Wright, C. J., J. C. DeVivo, and K. Funk. 2008. Information management and archiving plan, Southeast Coast Inventory and Monitoring Network. Natural Resource Report NPS/SECN/NRR—2008/062. National Park Service, Fort Collins, Colorado.