

# HOLTGRIEVE ECOSYSTEM ECOLOGY LAB

## DATA MANAGEMENT & ARCHIVING PROTOCOL

*Gordon W. Holtgrieve*

*Revision date: 29 May 2017*

## INTRODUCTION

Proper data management is critically important and taken very seriously within the HEEL. Most funding agencies require that data be curated in easily accessible formats with appropriate metadata, archived securely and in perpetuity, and widely available after the research has been completed. It is therefore absolutely critical that all data generated as part of HEEL activities be *properly curated and archived continually throughout the research process*. The protocol below describes general requirements for data management as well as specific actions for some types of data. All HEEL personnel are required to follow this protocol when working on HEEL projects.

## DATA MANAGEMENT REQUIREMENTS

1. HEEL researches must save and curate all data generated during their work in an easily readable/understandable format (preferably .csv text files). This should be done as soon after collection as is possible. It is not acceptable to leave data on instruments or field/personal computers for extended periods of time.
2. All HEEL project data must have, from the outset, accompanying metadata that includes at a minimum the four components given above (workflow, data dictionary, data citation, and data access). Metadata should be in a text file or similarly readable format.
3. Data in hard copy form (i.e., field notebooks, lab data sheets) must be scanned into digital form and curated along with any associated digital data. Hard copies should remain in the lab (FSH 232).
4. Workflow and metadata should be continually updated as the research process progresses.
5. All HEEL project data must be backed up on a daily basis and be stored in a least two findable locations.
6. At the completion of a project, the researcher must upload to the HEEL NAS system a final project folder that includes, at a minimum, the essential components listed above.

## DATA ARCHIVING & CURATION STANDARDS

All data must be 1) adequately described via metadata, 2) managed for data quality, 3) backed up daily in a secure manner, and 4) archived in an easily reproducible format.

### Metadata

All research data must be accompanied with a thorough description of that data from the beginning of the work. Metadata describes information about a dataset, such that a dataset can be understood, reused, and

integrated with other datasets. Information described in a metadata record includes where the data were collected, who is responsible for the dataset, why the dataset was created, and how the data are organized. Proper metadata includes the following four components.

**Workflow Capture:** A workflow is the formal description of how the data have been processed to get to the current state, which includes a description of the researcher's method for experimental data. It conceptualizes the data inputs, transformations and analytical steps to achieve the final data output.

**Data Dictionary:** A Data Dictionary is a repository of information which defines and describes the data resource with the goal of making it useable by someone unfamiliar with its collection. At a minimum, a data dictionary defines terms and variable in a data set including column headings, codes, etc. Data dictionaries can also include additional information such as units, measurement precision and accuracy, detectable limits, NA values, etc.

**Data Citation:** A suggested way this data set should be cited going forward. Often similar to citation of a journal article. Also includes reference to other data sets incorporated into the current dataset.

**Access Controls:** Defines who "owns" the data and allowable uses for the data including any limits to sharing or copyright concerns.

## **Data Quality Standards**

All HEEL researchers must take care that protocols and methods are employed to ensure that data are properly collected, handled, processed, used, and maintained, and that this process is documented in the metadata. HEEL methodological protocols are designed to prevent errors or defects (quality assurance, QA), however researchers should strive to improve these methods (they are in no way perfect) and continually check that the data they are generating matches expectations (e.g., by using quality control, QC, standards in analytical methods). The importance of QA/QC carries through to data entry and management; data should continually be check for errors and discrepancies. In short, check your data and seek ways to make it better.

## **Data Backup**

Bottom line: It is unacceptable to lose project data due to computer hardware failure/loss. Thus, all HEEL researchers are required to have a daily (or better) backup system. This system must: a) be in two distinct locations, and b) include a findable "key" that describes how someone other than yourself can access the data. The Computing Resources section below describes available options, including the HEEL NAS system, but ultimately it is up to the researcher to decide what works best for them and to document where data are stored.

## **Data Archiving**

In contrast to data backup, which is to prevent catastrophic loss of work, the goal for data archiving is to make your research easily understandable and reproducible in the future. It is therefore incumbent upon the researcher that, by the end of a project, care and effort is given to providing a highly organized and traceable accounting of the research that is archived in perpetuity. At a minimum, this archive should include: raw and full processed data, complete metadata, all computer code, and any research products (manuscripts, published articles, figures, etc.).

# COMPUTING RESOURCES

The HEEL has bought into a network attached storage (NAS) system operated by the Fisheries Acoustics Lab. We have a total of 8 TB of storage space and the ability to do automated backups of personal computers. This resource is available to anyone in the lab, but you are also free to use a different service for daily backups. Time Machine, Dropbox, CrashPlan, CarbonCopyCloner, Carbonite are all viable alternatives. However, if you use a different system, please leave a text file on the HEEL NAS system that describes how someone might retrieve your HEEL data if necessary.

## Connecting to the HEEL NAS system

### Apple macOS

1. You should either be connected to the UW network directly or via the UW virtual private network (VPN). More information on installing and using the VPN can be found [here](#).
2. Open the Finder window and type command & K.
3. Enter `smb://acoustics.washington.edu/HEEL` into the Server Address box. Hit enter.  
You be prompted for the username and password. They are:
  - HEEL
  - TreyReil1995

A new Finder window should open showing the folders within the HEEL NAS. This is where data archives should go and where you can use general data storage (e.g., if you need to transfer lots of data). Daily backups of machines do not go here.

### MS Windoze OS

1. You should either be connected to the UW network directly or via the UW virtual private network (VPN). More information on installing and using the VPN can be found [here](#).
2. Open Windows Run by pressing Windows Key + R.
3. enter `\\acoustics.washington.edu\HEEL` into the Run box. Hit enter.  
You be prompted for the username and password. They are:
  - HEEL
  - TreyReil1995

A new Explorer window should open showing the folders within the HEEL NAS. This is where data archives should go and where you can use general data storage (e.g., if you need to transfer lots of data). Daily backups of machines do not go here.

## Installing Synology Cloud Station Backup system (*You can use any backup system you want. This is one available option.*)

### Apple macOS

1. Download software as a dmg file [here](#). Double-click to install software as normal.
2. You should either be connected to the UW network directly or via the UW virtual private network (VPN). More information on installing and using the VPN can be found [here](#).
3. Launch software.  
You be prompted for the server address, username, and password. They are:

- acoustics.washington.edu
- HEEL
- TreyReil1995

4. Select where to store the backup on the NAS. The main directory path is: `home\CloudStation\Backup`. Within this folder create a new folder that identifies the computer. It should include at least the computer name, your name, and location. This is where the backup should go.
5. Select the folders on your computer to be backed up. You are welcome to back up personal files as well as project files.

MS Windoze OS

1. Download software as an msi file [here](#).
2. Follow the instruction for the macOS above.

## OTHER RESOURES

USGS Data Management standards and guidelines.

```
##
##
## -----
## Printed on:          25 Mar 2019
## R version:           3.5.1
## HEEL package version: 1.1.2
## -----
```