***Before class***

Review the following resources to prepare for this data management activity:

* Watch the NEON overview video: <https://youtu.be/39YrzpxVRF8>
* Read and be ready to discuss: Abbreviated NEON Small Mammal Trapping Protocols
* Read and be ready to discuss: Sutter, R.D., Wainscott, S.B., Boetsch, J.R., Palmer, C.J. and Rugg, D.J. (2015). Practical guidance for integrating data management into long-term ecological monitoring projects. Wildlife Society Bulletin, 39: 451–463.
* Read and be ready to discuss: Borer, E.T., Seabloom, E.W., Jones, M.B., and Schildhauer, M. (2009). Some simple guidelines for data management. Bulletin of the Ecological Society of America, 90(2): 205–214.

Make sure to bring your laptop to class/meet in the computer lab on [**Instructor Note:** insert date of class session] and make sure Microsoft Excel is installed. You’ll be working in partners to enter and analyze data that day.

***Introduction***

Data are the backbone of scientific research and exploration, so learning how to collect and process data efficiently is a critical skill for scientific professionals. Most people are not born understanding how to collect, record, enter, and analyze data, but with guidance and practice you can learn how to handle information and create world-class datasets.

Scientific organizations, especially large ones, spend a lot of time and effort determining the best ways to process data. The National Ecological Observatory Network (NEON) is one such organization that has made efficient data collection and processing a priority. NEON was designed to collect long-term ecological data on a continental-scale to help researchers address questions related to climate change, land-use, and invasive species. Data are collected at field sites called domains using standardized protocols, which allow for comparisons across large geographic ranges. Data on dozens of different variables and species will be collected every year for 30 years, yielding a comprehensive look at ecological processes across the entire United States. Regardless of the variables being measured, the general flow for data in these projects progresses from data collection to data files and metadata files as shown in Figure 1.



Figure 1. A workflow from data collection (can be in the field, lab, or other venue) to data collection sheets (paper, app, or entry form) to data files and metadata files.

Given the scale and scope of the project, they will create literally terabytes of data every year, so the information needs to be very well organized to be useful. In this activity, you’ll get practice translating field data into a usable format for long-term archiving and then explore how real NEON data can be used to detect ecological patterns, in this case the change in small mammal abundance over a year.

***Field Collection & Data Sheets:***

Small mammals were chosen by NEON to be bioindicators because they are present across the country in a wide variety of habitats. Their small size and short lifespan makes them sensitive to environmental changes, and they are responsible for spreading or maintaining a wide diversity of zoonotic diseases in an environment. They are also easy to safely collect as live specimens using arrays of traps like those described in the *Abbreviated NEON Small Mammal Trapping Protocol.* Live trapping has the advantage of being able to return the animal to their habitat without having to destructively sample. As you learned in the readings and the YouTube videos, in just a few minutes you can collect a lot of information from an individual animal. Because researchers want to reduce the stress on the animal while it is captured, it’s important to have an efficient framework for recording that data.

Take a few minutes now and review the NEON Small Mammal sampling data sheet. See if you can identify what variable is being recorded in each of the column categories. Make sure you know what codes refer to what type of animal being collected

Now look at the example data sheet. On the data sheet, circle the column headings for the following variables:

* Plot ID
* Date of capture
* Species
* Individual ID
* Sex
* Weight
* Whether individual is a recapture?

***Data Sheets & Data Files:***

Processing raw data sheets into a data table is only easy if the data table is well designed. Thinking about the presentation and the principles described in Borer et al. 2010 and Sutter et al. 2015, work in pairs to create an Excel data file that displays the information from your example data sheet for the variables you identified above. Make sure that your data table adheres to the best practices for data file construction that we talked about.

**Discussion questions:**

1. Why might field (or lab) data collection sheets look different from the digital data files?
2. Do you think the NEON data sheets are well designed to transfer the information to a data file? What makes the process easier and what makes it challenging?
3. Imagine you were responsible for entering data from hundreds of data sheets. How would you make sure you were not making mistakes? What types of checks could you do to make sure you were correctly transferring the data?

***Public Data & NEON:***

Another hallmark of NEON is that the data are all publically available. NEON has created an online data portal (<http://data.neonscience.org>) that allows access to all of the NEON data from any Domain across the country. This portal will serve as the long-term repository and clearinghouse for all of the NEON data in perpetuity.

We will use a series of data files downloaded from this portal to estimate the abundance of several small mammal species in different seasons (spring, summer, and fall) at NEON’s Smithsonian Conservation Biological Station field site during 2014 and 2015.

***Metadata:***

As we talked about during the presentation, metadata are another important component of collecting good data. A good metadata file can help someone unfamiliar with a data file interpret the codes and variables presented – and will help you remember what you did when you come back to the data later. It also provides an opportunity to discuss any irregularities in the data set.

**Discussion Questions:**

1. Examine the metadata file for the NEON data file. Briefly discuss with your partner how this file could have helped you interpret the data sheet and create your own data file or perform data analysis. Be prepared to share your observations with the class.
2. Now that you better understand the kinds of data collected in the NEON data file, think about what types of questions could be answered using this data. Discuss with your partner and draft 2-3 questions that could be asked with this data.

***Data Analysis:***

Once you have a well-designed data file, you can use that information to determine interesting patterns. One of the simplest ways that the NEON small mammal datasets can be used is to calculate abundance estimates for individual species within the plots. There are many ways to estimate abundance. One of the simplest is the Lincoln-Peterson method. This calculation uses data about the recapture of marked individuals of a species to estimate how many individuals of that species are present in a particular habitat. Because NEON small mammal protocols include marking individual animals with unique numerical identifiers, we can easily use NEON datasets to calculate small mammal abundance using Lincoln-Peterson according the following equation.

$$\frac{n\_{1}}{N}=\frac{n\_{2}}{m\_{2}}$$

N = total population size estimate

n1 = Number of individuals captured and marked in first sampling bout

n2 = Number of individuals captured in second sampling bout

m2 = Number of marked individuals in second sampling bout

It’s important to note that there are several assumptions that should be met for this calculation to generate an accurate estimate of population size:

* Individuals are randomly distributed between captures
* There is no change in the population (i.e. births, deaths, immigration, emigration) between sampling bouts
* Marking individuals does not impact their likelihood of being captured again in the future

**Discussion questions:**

1. Lincoln-Peterson estimation depends on several assumptions about the population. Knowing what you do based on the sampling methods outlined in the *Abbreviated NEON Small Mammal Trapping Protocols* document, do you think any of those assumptions have been violated in this data set? Why and what could be done to address those issues?

Complete the species population analysis using the Lincoln-Peterson model, following the instructions on next page, before completing the next discussion questions.

1. Why is it important that the m2 value is the number of individuals captured in the first month that were captured in the second month and not the total number of marked individuals captured?
2. Based on everyone’s data, how does the population abundance change for white-footed mice between plots? What are some hypotheses for why this pattern may exist?
3. Based on everyone’s data, how does the population abundance change for white-footed mice over the year at this site? What are some hypotheses for why this pattern may exist?
4. Looking back at the spreadsheet data table you created in section on Data Sheets and Data Files, how does this compare to format of the data from NEON? In what ways, if any, would you revise your spreadsheet?

**Step-by-step directions for Lincoln-Peterson model**

Working in pairs, use the workbook – NEONSmallMammalDataAbundanceWorkbook.xlsx – as a guide to calculate the Lincoln-Peterson estimation of population abundance using the following protocol. Please note that the workbook is laid with each worksheet showing each step, you do not need to create new worksheets for each step. You can perform most of the steps in a single worksheet:

1. Open the data file (NEON.D02.SCBI.DP1.10072.001.mam\_pertrapnight.072014to052015.csv) using Excel. Resave the file as an Excel workbook with you and your partners names starting the file name.
2. You and your partner will perform the analysis for sampling bouts either in the spring (April to May 2015), the summer (July to August 2014), or the fall (September to October 2014) for samples collected in your assigned plot. Record your time frame and plotID below:

Timeframe: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Plot ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Copy all the data to a second worksheet and name that worksheet. Remember, keep raw data raw!
2. Sort the data by plotID and then by collectDate. Now you can see when trapping occurred at each plot.
3. To start, you will perform the Lincoln-Peterson calculation for white-footed mice (*Peromyscus leucopus*, PELE) a common species in this area. Therefore, filter your data for the specified taxonID and plotID. Now you see only the data of interest for you Lincoln-Peterson calculation.
4. Identify unique individuals collected throughout your time frame. Record the following:

Number of individuals captured during the first month of the time frame (n1): \_\_\_\_\_\_\_\_\_

Number of individuals captured during the second month of the time frame (n2):\_\_\_\_\_\_\_

Number of individuals from the first month recaptured during the second month (m2):\_\_\_

*Remember this is the number from the first month that were recaptured in the second month. Not the total recaptured in the second month.*

1. Use these numbers to calculate the population abundance of PELE for your site and time frame:\_\_\_\_\_\_\_\_\_\_\_
2. Share your results with the class.