Saving & Organizing Data Using Box

- What type of data do I save to Box?
- When do I save data to Box?
- How do I get permission / access to Box?
- What format(s) should I save the data in?
- What is metadata and what do I need to do about it?
- Why am I being asked to save data to Box? Can't I hand over my lab notebook?
- How do I name data files?
- How should I enter data into spreadsheets?
- When/how often do I save? How many backups should I have?

What type of data do I save to Box?

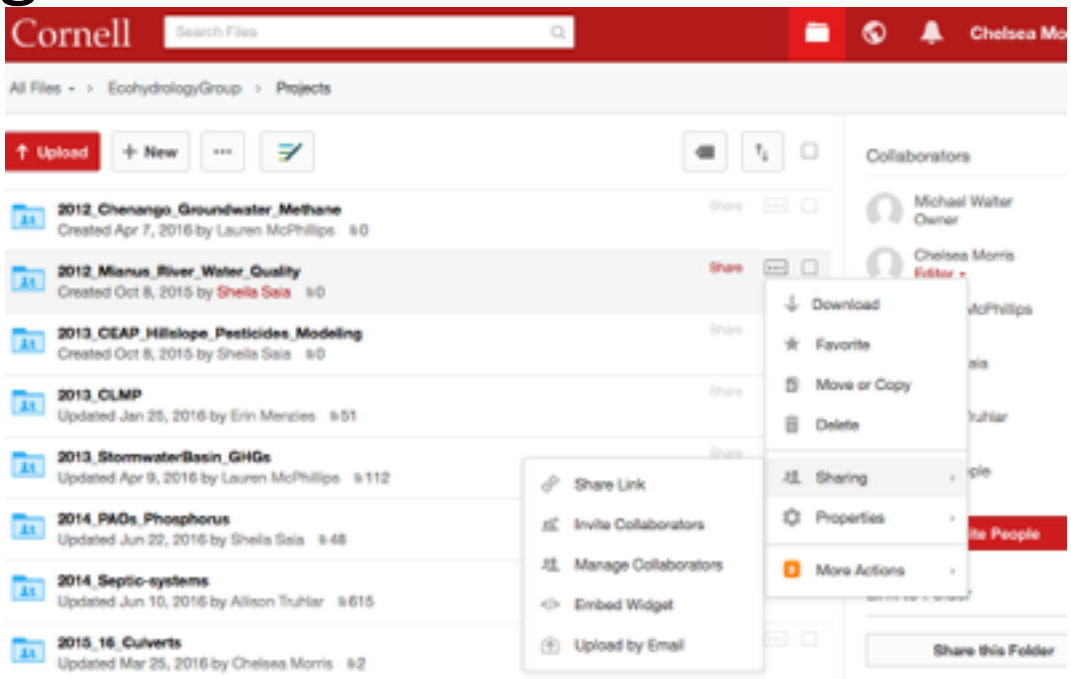
- *The Simple Answer:* EVERYTHING.
  - We have unlimited storage space, so it's possible to save every datasheet, note, or output that you generate.
- *The Nuanced Answer:* All that will help you (or your grad student) reconstruct the experiment
  - Saving every piece of information may be overkill and clutter your project folder. Consider keeping the most up-to-date version of a file. See the version control features in Box.
  - Save rough output from analyzers, clean versions of data, narrative notes regarding data collection

When do I save data to Box?

- It’s best to work off of Box directly. Don’t wait until the semester/summer’s end to upload data.
  - Use Box Sync to map a drive to your computer directory for easy access.

How do I get permission/access to Box?

- The graduate student you are working with will give you access to the project folder(s)
- Click the three dots next to the folder and navigate to “Sharing” then “Invite Collaborators”



What format(s) should I save the data in?

- Non-proprietary formats
  - Examples: comma-separated values (.csv); text (.txt); images (.tiff, .svg)
  - .xls and .doc are okay, but the above formats work best for large datasets (think DNA sequences) & minimize errors when read into R.
- Lossless formats
  - Formats that compress the information in a file are often smaller, but the compression often permanently removes data from the file.
  - Examples: audio (.wav); images (.tiff)
  - Avoid: audio (.mp3); images (.jpeg)
- Unencrypted and uncompiled files

From Cornell's Research Data Management Service Group

What is metadata and what do I need to do about it?

- Metadata is documentation that describes the data
  - How the data was collected
  - How the data was cleaned, analyzed, changed from the original output
  - Units, analytical procedures used (e.g. KCl-extraction versus water-extraction)
- Metadata may point to well-commented R scripts for analyses.
- The file is often named “ReadME.txt” or “Metadata.txt”
- Ask your graduate student for their preferred metadata format
- Metadata standards and examples from Cornell’s Research Data Management Service Group

Why am I being asked to save data to Box? Can't I hand over my lab notebook?

- We want to guard against data loss. Paper notebooks can be misplaced, damaged, or unreadable.
- We want to collectively use a cloud-service to store our data for the long term. Box is Cornell's preferred data management product.

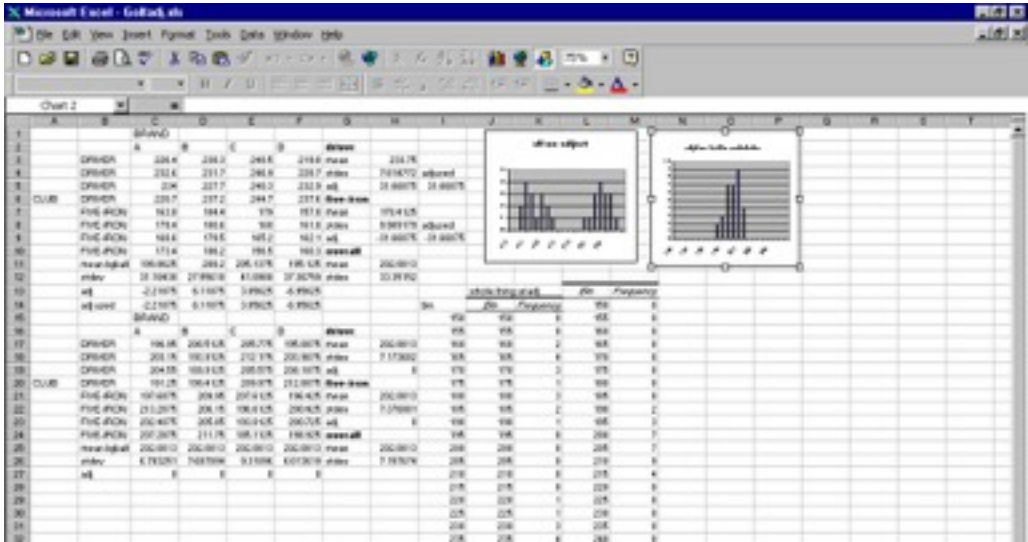
How do I name data files?

- Ask your graduate student for a template with file name, column headings, & missing data method.
- File name: meaningful, but not overly long. Possibly incorporate data collection stage & time stamp.
  - Example: “150331_CH4_raw.csv” for raw methane values output by the GC from samples collected on March 31 2015
- Column headings: meaningful, but not overly long. Do not duplicate column headings within a file. Use only alphanumeric characters, underscores, or hyphens. Do not use spaces.
- Missing data: use a standard method such as -9999 or NA. Do not leave blank or use zeros.

How should I enter data into spreadsheets?

- Include only data in a spreadsheet. Not formulas, figures, or analyses
- Use scripts for formulas & analyses
- Data in spreadsheets should look rectangular with every row and column in the rectangle filled.

DO NOT create your spreadsheets like this



When/how often do I save? How many backups should I have?

- Save a raw data version (consider locking the file to prevent changes); a cleaned file; and an analyzed/ results file.
- Save different versions of the cleaned and analyzed files as needed for the project.
- You can view the version history of a file by clicking on the 'v' below the file name or when previewing a file. You can open the version history window through the More Options menu (Properties > Version History).
  - For more help see "How To Track Your Files and File Versions (Version History)"