Brodie Lab
January 2014

Brodie Lab Standards and Practices

- The goal of all research is to collect publication-quality data. Take time and be careful whenever you're collecting data.

- Record your hypothesis and the brainstorming process leading up to it; your experimental design and protocol(s); data collection; data curation and clean-up; statistical analysis; and your interpretation of the results.

- As a rule of thumb, a year from now a collaborator should be able to look at data you collected, and interpret them correctly and with no missing information.

- Write legibly!

- Make each datasheet complete on its own. All collected data should clearly indicate, at a minimum **what** type of data it is, **who** collected the data, and **when** it was collected.

- Lab notebooks and datasheets should not leave the lab.

- Approach your lab notebook like a journal: you should use it to record everything that you do, as well as why it was done.

- Each entry in lab notebooks should contain the **date,** a brief **title** describing what was being done, the **name of the researcher** collecting the data

- If you are collecting data in a way that doesn't allow for datasheets, make sure the data are consistently and clearly organized and labeled.

- Enter data within one or two weeks of collecting it

- Proofread data as you enter it and check for errors after entering the data.

- Back up your data regularly and keep digital files well organized.

- Do not irreversibly save over changes in digital files.

- Take notes during analysis such that you can recapitulate the results later.

- There is a standardized format expected for collecting data and taking notes. See the Formats section at the end of the document.

Brodie Lab
January 2014

Record-keeping in Science

Record-keeping is important throughout the scientific process. You should record your hypothesis and the brainstorming process leading up to it; your experimental design and protocol(s); data collection; data curation and clean-up; statistical analysis; and your interpretation of the results. These records will likely be found in three places: in your lab notebook, on datasheets, and in digital form.

University policies
Read UVA's data rights and responsibilities guide:
http://dmconsult.library.virginia.edu/data-rights-and-responsibilities-guidance-1-0/

Key points from the guide

- UVA claims intellectual property rights to data generated with significant University resources, including grant funding. *This includes lab notebooks*.
- The PI (the Brodie Lab) must retain all original lab notebooks for five years after completion of the project

Data Collection

Typically you will be collecting data either on datasheets or in a lab notebook (whether personal or part of a collaborative project). In either case, the goal is to collect data that is accurate, unambiguous, and generally understandable. Make it as easy as possible to understand what is going on with the data, while reducing the chances that any important data are lost. **As a rule of thumb, a year from now a collaborator should be able to look at data you collected, and interpret them correctly and with no missing information.**

Data should be precise enough to capture all the currently relevant details, as well as any interesting observations that may prove important later. Conversely, collecting data in too much detail wastes time and effort. Finding the right balance for the amount of detail and the types of data to collect can be difficult, especially when first starting out. The best thing to do is to chat with your advisor and other researchers, then go out and collect some data. See what seems important, and repeat as needed.

General rules for data collection:

- Be consistent. Don't change the format of the data midway. If you do have to change things, make a note about what you changed and why.
- Remember that errors happen. Try to make the data you collect as robust as possible against likely errors.
    - Double check the IDs of individuals when you are collecting data. Errors from incorrect IDs are some of the most pernicious errors.
- Write legibly! Check with others to make sure they can read your writing.

- Close your 9s with a clear circle so they don't look like 4s or 7s.
- Write your 8s with two clear circles, so they don't look like 9s, 6s, or 0s.
- Cross your 7s (e.g., 7̶). This will make them stand out from 2s and 9s.
- When mixing letters and numbers, dot or cross your 0s so they are distinct from the letter O.

Metadata

Metadata is information about the data and the data collection itself. Every data point should have at least three types of metadata associated with it. The precision of the metadata will vary depending on what type of data is being collected. Most of the time, it will be fine to have a description at the top of the datasheet or lab notebook entry with the **what**, **who**, and **when** for all the data on the sheet/entry:

1. **What** type of data it is. This should be at least a short descriptive phrase (e.g., "behavioral observations"), or better yet a brief description (e.g., "Courtship and mating behavioral observations of FFB at BP01 population"). Descriptions should be sufficiently clear and detailed that a collaborator would know what the data are one year from now.
2. **Who** collected the data. Having a record of who collected the data is important for resolving any errors that arise (and they will…). If multiple people helped collect the data, it is best to list all of them, but make sure to indicate which person actually wrote down the data, as this is the person who will be contacted first in case of errors and ambiguity.
3. **When** it was collected. There should always be a date associated with any datum (and a time, if that level of precision is needed). Dates must include years. See the formatting section below for more details. Even if the data you collect doesn't change over time, knowing when data were collected allows them to be more easily sorted and connected to lab notebook entries.

Depending on the project, you may have other types of metadata as well, but you should never have fewer than these three.

Lab Notebooks

Approach your lab notebook like a journal: you should use it to record everything that you do, as well as why it was done. This is equally true whether you're working in the field, in the molecular lab, or on the computer. Because they are often the key record of the work that you've done, **lab notebooks should not leave the lab / greenhouse / molecular space**.

All entries should include appropriate metadata information: **what** was done, **when** it was done, and **who** did it. Additionally, lab notebooks are the principal place for you to record not just **what** was done, but **why** it was done. Ask yourself if your collaborator could pick up the entry and follow what you did and understand the logic underlying it. If they couldn't, you didn't write enough.

All lab notebooks should contain the following information at the front of the notebook

- The name of the person to whom the notebook belongs
- Their contact information (at minimum, an email)

- If project-specific, the name of the project
- A table of contents listing the page number, date, and title of important entries

Each entry in lab notebooks should contain the following information:

- The date
- A brief title describing what was being done
- The identity of the researcher collecting the data (if more than one person uses the lab notebook)

Datasheets

Datasheets are the primary place data are recorded. Just like lab notebooks, **datasheets should not leave the lab / greenhouse / molecular space**. Datasheets can differ from project to project, but most datasheets will be organized in tabular form — with columns for different data fields, and rows for each data point.

Rules for datasheets:

- Make each datasheet complete and interpretable on its own. Leave space at the top of each sheet for metadata and notes about the datasheet.
- If you are collecting data in a way that doesn't allow for datasheets, make sure the data are consistently and clearly organized and labeled. For instance, if you are taking photos of flowers, include a scrap of paper with the name of the plant in the photo, so it can be connected to other information about the plant.
- Don't rely on the ordering of the datasheets for metadata purposes (e.g., dates), unless datasheets are permanently bound together such that the order cannot change and sheets can't be omitted (e.g., if they are in a bound notebook, not just connected with removable binder or paper clips).
- Include a space for notes and other unexpected observations about each datapoint (e.g., a column in tabular form).
    - **Be careful not to overuse the notes section**. If you find yourself writing the same note repeatedly, it may warrant its own section on the datasheet.
    - If there are multiple notes for an observation, separate them with a semicolon, not a comma (commas occur too regularly in natural language).
- When collecting data on a particular day or timespan, you may want to number each datasheet, so that they can be easily sorted.
    - You can take this further and number each observation (e.g., row on tabular datasheets). Then you can refer to any given data point using the timespan (date and time in most cases), sheet number, and row number of the datum.

Data Entry and Management

Most, if not all, data will need to be entered into digital files (e.g., spreadsheets, databases) on the computer and stored digitally. This section deals with data entry and the management of digitally stored data.

It is important to take the time to logically lay out project directories to ensure that you and your collaborators will be able to find important data in the future. When possible, all files for a project should be stored in the same directory. Use subdirectories to store parts of a project, or to group files with the same types of data. Project directories should contain a README file (usually just a text file named "README.txt") that describes what the project is, as well as describing the contents of the important files and subdirectories for the project, and some forms of project metadata. For large projects, subdirectories should contain their own README files.

Metadata

All appropriate metadata from data collection (see above) should be transferred during data entry and remain associated with the data. Additionally, information about the data entry and data organization should be added. There are two appropriate methods for storing metadata:

1. A README file contained within the directory with the data.
2. If data are stored in a text or spreadsheet format, you may add a section or worksheet with the metadata. For instance, spreadsheets should have an additional worksheet with the metadata, and scripts should have comments at the top of the file.

Data Entry

Rules for data entry:

- Enter data within one or two weeks of collecting it, while it is still fresh in your mind. Waiting longer will only force you to struggle to reconstruct what you did, making data entry more difficult and less accurate.
- Keep in mind that the data may later be sorted or subset after you have entered them. As such, make sure each data point has the appropriate metadata. For instance, when entering tabular data from multiple datasheets this will often mean creating new columns for the metadata and repeating the metadata from a datasheet for each row entered from that sheet.
- Proofread data as you enter them.
- Check for common data entry errors after all the data are entered. For instance, when entering tabular data with grouping variables, list each variable's unique groups using PivotTable in Excel or the *table* function in R. This is a good way to catch any typos that may have occurred during data entry (e.g., entering "Fomes" as "Fome"), as well as making sure you are using consistent language (e.g., always "2014-01-12" with no "2014-1-12" mixed in). For numeric or continuous variables, use a histogram to look for extreme outliers that may have been created accidentally during the data entry process.
- Initial and date datasheets once you have entered the data. This results in a record of who entered the data, and when.

- Record the name of the file and the file path (folder where it can be found) of entered data in your lab notebook, along with a brief description of the data that the file contains.

Data Management

Digital files should be organized and managed in ways that prevent the loss of data, and should be regularly backed up.

The best way to accomplish both of these aims is to use revision control software (also called version control software). Revision control software allows you to save changes you make to files while still allowing access to past versions, and most such software will allow you to push the changes you make to another location or server, enabling easy data backup and collaboration. Two easy-to-learn revision control software programs are bazaar (http://bazaar.canonical.com/) and mercurial (http://mercurial.selenic.com/); git (http://git-scm.com/) is a more complicated and more powerful program. Dropbox (https://www.dropbox.com/) has some version control capabilities, but only stores previous versions for 30 days, and so should not be depended on as a form of revision control.

Most version control software also allows you to add comments when you commit changes. You should use this functionality to describe the major changes you have made since the last version.

Other data management rules:

- **Do not irreversibly save over changes** in digital files. When changing files:
    - If using version control software, commit the current version before embarking, so that files may be reverted if things go badly.
    - If not using version control software, “save as” to create a new spreadsheet with the changes, and append the date to the file name.
- You should always be able to identify the most current version of your digital data files. Do not rely on “last modified” timestamps, as these are often modified upon opening documents in many editing programs.
- **Back up your data regularly!** This can be on an external hard drive, a lab server or lab computer, or, ideally, on both.
- Dropbox is not a backup tool, it is a collaboration and file syncing program. You should back up the files in your Dropbox folder regularly.

Data Analysis

Both you and your collaborators should be able to recapitulate the results of your statistical analysis if/when they need to be rerun. You should have an entry in your lab notebook that at least describes the analysis you conducted, and the name and file path of any analysis script files, if used (e.g., in R). All publications now require that your data are deposited on open access sites in a format sufficient to recreate the analyses contained in your paper.

Together, your lab notebook entries and the annotations in your script files provide a record of the analyses you have run, and at the minimum, the following should be recorded:

- The name of the file in which the data used for the analysis were stored.
- Any data manipulation necessary to conduct the analysis (e.g., observations that were excluded (and why), data that were log-transformed, etc.)
- The model that was run, along with any information necessary to specify it (distribution of the residual error, the link function, etc.)
- Why you chose to run the model that you did.
- The results of the analysis, including what would be needed to publish it. **DO NOT** just record a p-value!

You should also save the output from the analyses that you run.

Collaboration

Collaborative projects often present their own unique set of challenges for record-keeping, data collection, and data management. One of the most important things when keeping records for collaborative projects is making sure that it is possible to find information/data from any stage of the project, regardless of who was responsible for collecting it.

- All members of the project should keep their own lab notebook according to the lab notebook protocol laid out in this guide.
- In some cases, it may make sense for there to be a shared notebook or binder containing all the datasheets produced by the project (e.g., PCR master mix recipes, plate maps, behavioral datasheets). Individual lab notebooks should contain entries indicating when to look in the communal binder (e.g., “13 January 2012: Ran PCR and put the plate map for PLATENAME in the project binder”)
- Record the full names, initials, and contact information for everyone involved in the project. Once this information is written down somewhere, everyone involved can be referred to using their initials.

Formats

Abbreviations

Where possible, use only standard abbreviations (e.g., forked fungus beetle behavioral projects may use “CRT” for “courting”). If there isn’t space to write the full word or phrase on a datasheet,

or if repeatedly writing the same phrase would be too onerous, it is acceptable to abbreviate it, but you must have a definition for the abbreviation on the datasheet or in the lab notebook entry.

Names
Names should be unambiguous. For brevity, it is acceptable to use initials (e.g., EDB3, MEA), but keep in mind that many people may have the same initials, so be careful. If you do use initials, it is better to use the full set of initials, including middle names (e.g., MEA, not MA). Include an entry in your lab notebook that records the full names of everyone involved so you (or a collaborator) can figure out to whom the initials refer.

Dates
Dates should ideally be written as YYYY-MM-DD. E.g., January 13th, 2014 is written as 2014-01-13. Include the year. Avoid the MM/DD and DD/MM formats as they are ambiguous for the first twelve days of each month.

Time
Times should be recorded in the 24 hour clock (e.g., 13:00, 00:15, 18:59). It is far too easy for people to forget the AM and PM, and data collection can happen at all hours.

Brodie Lab
January 2014

Brodie Lab Policies

Undergraduate Research Students

General

- If you're the last person in lab, be sure to lock the door when you leave. **This includes all three doors** in the Brodie-Cox lab space.
- Food should be kept in the conference space/kitchen area. You're welcome to leave food in the fridge on the shelf labeled "undergrads", but **label it clearly**.

Time Commitments

- You are expected to work ~5-10 hours/week, including meetings and reading groups.
- Meet regularly with your mentor to discuss the work you have done and any questions that have arisen.
- You must volunteer for at least a semester before doing independent research.

Lab work

- Clean up your station after doing lab work. If you were doing messy work (e.g., working with beetles, fungus, or plants), be sure to wipe down the bench when you're done.
- Some lab equipment is communal and some is not. Ask your mentor before using any new equipment or computers.

Lab notebooks and datasheets

- Your lab notebook should be a record of your work, including what you have done, and when you did it.
- If you are doing independent research, it is a part of your grade.
- It is your responsibility to show your lab notebook to your mentor on a regular basis and solicit feedback. Prior to these meetings, review past entries and enter important entries in the notebook table of contents.
- **Never remove your lab notebook or your datasheets from the lab / greenhouse / molecular space**. Keep it on the shelf space provided.
- Your lab notebook and all datasheets must stay in lab after you leave the project.

Reading papers

Whenever you read a paper with your mentor, you should write your own abstract for the paper (one paragraph) by answering the following:

- What was main point of the paper, or the hypothesis it tried to test?
- Briefly (two or three sentences) summarize the methods and results (e.g., "They did X and found Y").
- What were the authors' conclusions? Do you agree? Why or why not?
- Write down any confusing terms or ideas to bring up during discussion.

Bring your summary to the paper discussion, and turn it in to your mentor after discussion.

Contact

- When in doubt, ask questions. It is better to interrupt your advisor for 5 minutes than to make mistakes or get stuck all day.
- Under most circumstances, it is fine to find your mentor and ask them questions without having a scheduled meeting, but discussions longer than you will need to schedule a meeting.

Grading

When doing independent research, your grade will be based on a combination of the following:

- attendance, reliability, and diligence.
- lab notebook
- paper summaries
- project proposal and report (including drafts)