

# Expectations

Hallett Lab

September 25, 2020

## 1 Authorship

There are five key parts of a study:

1. Idea generation
2. Funding (through own grants or contributions to Lauren's)
3. Data collection (including cleaning and curation)
4. Analysis
5. Writing

A good rule of thumb is a coauthor is involved in at least three of the five pieces, or has contributed substantially to two. When an individual's support or input is sought for an early or critical stage of a project, it is lab practice to give them the opportunity to be involved to the level of authorship. For example, if the project involves follow-up at someone's study site or relies on their expertise for system knowledge/experimental design/methodological approach, the project lead should make sure that they have the opportunity to continue to engage with the project should they be interested.

It is widely acknowledged that underrepresented groups in science are coauthored less because their contributions are not as easily valued or apparent. Lab members are encouraged to keep this in mind when considering the contributions of others, and also when advocating for their own contributions. Especially when undergraduates are involved in a project, they should be offered the chance to remain involved past graduation (if they are motivated and interested to do so).

Lead authors are expected to keep a running tally of folks who contribute to the project to ensure that everyone is appropriately credited (either as authors or in the acknowledgement). It is recommended that this tally be maintained and updated in the project's Github wiki.

## 2 Data management

All lab members are responsible for the curation of the data they collect. The lab's data management protocol is:

1. Data for projects in progress will be stored either in a project OneDrive or a project Dropbox folder that is shared with Lauren. Code for projects in progress will be backed up on Github and in a repository that is shared with Lauren (for grant-funded projects, this repository will typically live on the Hallett Lab Organization page).

2. Hand-written field notes, lab notes, and datasheets will be scanned promptly and organized in a Raw data folder for each project. Field notes will be digitally archived in a Github wiki associated with a designated project repository to facilitate sharing among project members (located at <https://github.com/HallettLab>).
3. Entered data will be saved in .csv files and stored within an “Entered data” folder for each project. Data will be error-checked in R (e.g. for missing values and outliers signifying typos in entry). Corrections to the data will be documented within an associated R cleaning script. This script will also format the data to comply with “tidy” data format. The workflow for data cleaning will be stored in a Github repository associated with the project. The final, cleaned data product derived from this script will be exported to “Cleaned data” folder and will be the version that is publicly archived and shared. Lab derived data will follow a similar process, except that data may be entered or imported digitally (skipping the raw data scans).
4. Protocols for all methods will be developed and stored within a “Protocols” folder; updates to the protocols will be noted with the date of revision. Protocols will serve to guide all project members in consistent methodologies and will also be the basis for final metadata. Detailed keys for column headings will be stored in a paired .csv file for each entered dataset (or alternatively be stored at the head of relevant .csv file).
5. Upon project completion or when a lab member is preparing to graduate/move on, the data will be archived on the UO Server for permanent storage.

### 3 Reproducibility

We are committed to using the best practices in scientific computing and reproducible science. All the materials needed to reproduce the study entirely (from data collection to analysis) should be made public upon publication. This includes:

1. Cleaned data products and associated metadata, uploaded to an appropriate digital archive. Often this will be the Environmental Data Initiative or the Knowledge Network for Biodiversity.
2. Well-documented analytical code to run the paper’s analyses, archived either on Github or Zenodo (Zenodo is more stable for finished code).

### 4 Science communication

The lab is encouraged to engage in science communication and outreach, and to keep a record of this engagement. For example, this might involve presenting at the Willamette Valley restoration meeting or CPOP meeting, attending the ARS Advisory meeting, leading a workshop on Github/data management for the broader community, etc.

It is helpful to keep a record of our community partners and how we have built relationships with them. To this end, outreach events should be logged in a lab google document as short journal entries that include why/when/who/what/where.