Expectations

Lauren C. Ponisio

August 13, 2017

1 Authorship

There are five key parts of a study:

1. Idea generation

2. Funding

3. Data collection (including cleaning and curation)

4. Analysis

5. Writing

A good rule of thumb is a coauthor must be involved in at least three of the five pieces. 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.

We are part of community and should freely help each other whenever appropriate, i.e., bounce ideas off each other, ask for help with code, ran a statistical analysis by someone. That being said, it is important to discuss coauthorship early and often. Be upfront about:

1. Your interest in a project

2. What you can contribute "freely"v (i.e., without coauthorship)

3. What level of involvement you believe warrants coauthorship (i.e., I am happy to brain storm an analysis with you, but if I am developing code with you I would like to be more fully involved in the project and be coauthored)

2 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), must be make available publicly and associated with all publications. This includes:

1. 1. Curating all data in a relational database where automated data cleaning is conducted that is version controlled, including a record of any changes to the data and the raw data is never altered.

2. Version controlling and posting all protocols (collection, molecular, etc.) publicly, and associate each protocol DOI with its corresponding publication.

3. Version control all analytic code. Final products will be executable from a single script with an explanatory Rmarkdown file, publicly available and associated with corresponding publication.

4. Post all final data publicly

3 Data management

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

1. Data will first be entered from field sheets into a spreadsheet on UC Drive, the UCs version of Google Drive with FERPA and US-EU Safe Harbor protections, then downloaded as ċsv files into password-protected Dropbox (Box is also an option, but access disappears after your UC id expires).

2. We will use R (or programming language of your choice) to check each datasets fields for consistency and typographic errors and output the cleaned data files for analyses. After any updates, the Google spreadsheet is downloaded as a ċsv file, and cleaned and assembled using R protocols, while never transforming the original Google spreadsheet. All code for cleaning data will be version controlled on Github so there is a record of any changes made to the data.

3. Structure Query Language (SQL) Lite package in R will create a relational database that unites each of the relevant datasets for analysis. Traditional spreadsheets will be created for the data made publicly available.

4 Science communication

All lab members are required to communicate their research to a non academic audience at least once a year (i.e., blog for National Geographic, presentation at the Entomology fair, presentation to a school group etc.).