

Meeting 2 - how do we store our data and code at present?

Present: Freddie, Alex, Rich

Overall impressions

Following reading Wilson et al 2017 ‘Good enough practices in scientific computing’ <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000424#s6> it is clear we are doing some of these things already (as in meeting one) which is positive.

Ways we store our data:

We are backing up raw data and storing it in an unadulterated form. Where we store the raw data depends on our need. For example, Freddie will store all raw data within the R project file, where Rich will store it externally so there isn’t duplication of the datasets across projects (a space constraint specific to the data being used).

Ways we share our methods:

- Through our published papers
- Through labelling throughout R markdown files or commenting in R

Ways we share code now:

- Zip the whole project and send to whoever wants to run it
- Sharing through github to be cloned or forked and send large raw_data files separately

Things we liked from the paper:

- Highlighting the importance of the ReadMe file for documenting what people need to do on accessing your code and storage
- The importance of shifting the naming convention for figures away from numbering (i.e. Fig_1, Fig_2) to a descriptive label that doesn’t need to be changed when the figure has been
- Providing a file structure that could be applied by someone who wanted to try shifting the organisational structure of their code and data

What is the need for within the size ecology lab?

- We thought that having a repo structuring example that we can all play with (pushing, pulling, creating new branches for from GitHub) could be a really good way for people to gain confidence in GitHub.

Other useful thoughts that came out of the discussion

- Could be that raw_data stored outside of the project could also be stored openly on something like Zenodo (if it is possible to do so)
- file.exists statements can help with preventing on running tidy data scripts if the tidied raw data already exists within the project.
- Streamlining storage for project to one place so that data are not across multiple sources. Lab members need to be abreast of the various options and advantages of different file storage solutions provided at UTAS re storage vs privacy vs disaster recovery.
- Can we find out how to link to our UTAS OneDrive through an API?
- We could add our accepted (non-formatted) versions of manuscripts to the project/Github repo too which should increase open accessibility.