What is open and reproducible science?

Lowndes et al 2019 define it as:

"Our idea of open data science blends R developer Hadley Wickham's definition of data science —
"turn[ing] raw data into understanding" — with open science tools and practices, such as using
collaborative version-control platforms for code and project management"

How do size-ecologists see open and reproducible science?

- Aligns with the ethical onus as public researchers for making our research public
- It is a key shift in transparency in methods and moves away from the dependence on English descriptions in methods too – speeds up scientific progress.
- Can be a source of livelihood through social media and other avenues
- It provides multiple avenues for statistical programming and improves understanding – coding can get down to the nuts and bolts of an analysis without being too much of a black box (e.g. minitab, spss)
- Increases community efficiency in tackling challenges
- Unclear guidelines – talked about a lot –
- Maybe not a destination but a journey - incremental improvements to your research are inevitable – not searching for perfection.
- Documentation and description of code
- Open and reproducible science is two pronged - manuscripts and code are different parts of clarity and reproducibility that we need to work on.
- Also requires understanding copyright/IP/proprietary information needs – should be a clear description for what is allowed- can stipulate this within licenses.

Interesting resources from BES on reproducible code

https://www.britishecologicalsociety.org/wp-content/uploads/2019/06/BES-Guide-Reproducible-Code-2019.pdf?utm_source=web&utm_medium=web&utm_campaign=better_science

Who feels comfortable that their code is clear and reproducible – could someone pick it up and run it with no guidance?

- Not about making it perfect, not to over-plan to the point of being daunting - those incremental improvements key
- Open domain helps push it towards being clear and reproducible

Where is the need for open and reproducible code in general and within Size Ecology?

- Could we use size ecology as halfway house for code and data products – a resource hub to avoid email-based back and forth or the loss of data from different people's presence in the group
- Central place to fork to for important resources
- Create an open lab book.
- A resource to help avoid repetition in terms of learning skills/meetings

- Forking all published works into Size Ecology Github as a rolling record of work done within Size Ecology
- Specific codes for certain skills – flagging key tasks and useful code to approach them. A key place for collaboration. Some key requests for new topics or existing code for:
    - Chlorophyll data to size spectrum slopes/coefficients
    - Calculate slopes from observational data
    - Mapping – shapefiles/rasters
    - Using Git – ssh, terminal and bash
    - Basic stats – linear models, GAMs
    - Visualisation – ggplot, base
    - DataTas – resources
    - Clarify the use of licenses Challenges of copyright/proprietary information/ Adding license queries
    - Create Cheat Sheets for regular tasks?

Time for reading Lowndes et al 2019: https://www.nature.com/articles/d41586-019-03335-4

Poll: Is this course something we want to get involved in?

Resounding yes from the group!

Main things that the 10 week (every other meeting – so really once a month and so 10 months) program should aim to do – with some questions we could consider throughout the process:

Normalize data discussions

- Create digital and physical spaces for sharing questions and expertise – how do we do this through the medium we already use? **
- Regular data centric meetings – are we happy with once a month (every other meeting)?
- Who wants to know more – become more engaged in coding culture – and what are the things we know or can recommend?
- Horizontal leadership – we can alternate who leads data discussions?

Identify shared needs

- How do people currently organise their workflows?
- What software do people use?
- Which packages help?
- How do people currently work collaboratively?
- So, what are the main shared priorities going forward?

Think ahead

- Is there scope for collaboration in the group?
- How could we make the workflow clear and defined going forward for students and new members – can we use it to collaborate.

- What are the existing training opportunities that we know of for students esp. But the wider group too, available at UTas/online.

Week 1 - What does an open science data flow look like and how does transition to one occur?

Read Lowndes et al 2017 and discuss: How do we approach reproducibility, collaboration and communication within size-ecology and our wider networks?