Best Coding Practices

Lecture 01.1: FAIR Guiding Principles

Lauren Sullivan

Module: Data Management, Visualization & Reproducibility

Helpful Readings

- Wilkinson et al. (2016) The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data.* 3:160018.
- Sandve et al. (2013) Ten Simple Rules for Reproducible Computational Research. *PLOS Computational Biology.*
- Powers, S.M., S.E. Hampton (2019) Open Science, Reproducibility, and Transparency in Ecology. *Ecological Applications.* 29:1.
- GO FAIR Initiative

Four Foundational Principles of Data Management

1. **F**indability
2. **A**ccessibility
3. **I**nteroperability
4. **R**eusability (Reproducibility)

If we care about long term data useability, sharing and stewardship, then we need to think about how and when to make our data FAIR. This is especially important if we want our work to be accessed by multiple stakeholders (i.e. industry, public, education, policy, *computers?*)

These ideas will come in handy when writing future "data management plans" required for many grants.

Quick Note

The FAIR guidelines were originally set up to allow for use across all platforms and stakeholders, and to be "machine-actionable" to allow for large-scale, internet-driven data discovery.

In many cases our work will fall outside this scope, but the principles are still important. Where appropriate, I have modified the FAIR Guiding Principles to match work that may never be fully machine-actionable. For those seriously interested in machine-actionable data sharing, there are many resources on the internet for you (e.g. FORCE11).

Findability

To be *Findable*...

1. data and code must be non-obscure
   - Is there a statement in your paper about where to find your data and code?
2. data must be described by detailed metadata.
3. data and metadata are assigned globally unique and persistent identifiers so as to link across datasets.

Accessibility

To be *Accessible*...

1. data and code are open, free, and able to be implemented easily (you want others to be able to reproduce your work).
   - This includes clear annotation of code.
2. metadata must also be available in order for others to understand what you did.

Interoperability

A quick definition. Interoperability is the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort.

To be *Interoperable*...

1. data and metadata use a language that's common and broadly applicable (e.g. R).
2. data, and metadata especially, use vocabulary that follow FAIR principles.
   - You want others to be able to understand what you did.
3. data and metadata include references to other sets of data and metadata or studies when appropriate.

Reusability

To be *Reusable*...

1. metadata are described in depth.
2. data and metadata use field-specific community standards.
3. data and code must be produced in formats that are common to the field and can be repeated.
   - Make sure to include the version of your programs and packages used.

Problems with FAIR data?

There can be a lot of issues with open source, FAIR data. Some are listed below. We will discuss these issues in class.

- Privacy of those not protected by IRB (e.g. land owners)
- Potential poaching
- Interest in writing future follow up papers with data

Where to Store Data, Metadata and Code

Here are some resources for storing data, metadata and code for your papers. These options range in FAIRness.

Cost	Source
Cost of pub.	Supplemental information for paper
Free	Zenodo
	Open Science Framework (OSF)
	Environmental Data Initiative (EDI) {*Uses FAIR practices. Free for most ecologists. Large datasets may require fees.}
	Dataverse {*Pretty sure its free.}
	Figshare
Small fee	Dryad {*free @ some journals where required.}
Not sure	GenBank
	Worldwide Protein Data Bank (wwPDB) {*Uses FAIR practices.}

If you have other ideas to add to or improve this list please share!