

The EEB & Flow

Blogging all things ecology and evolutionary biology

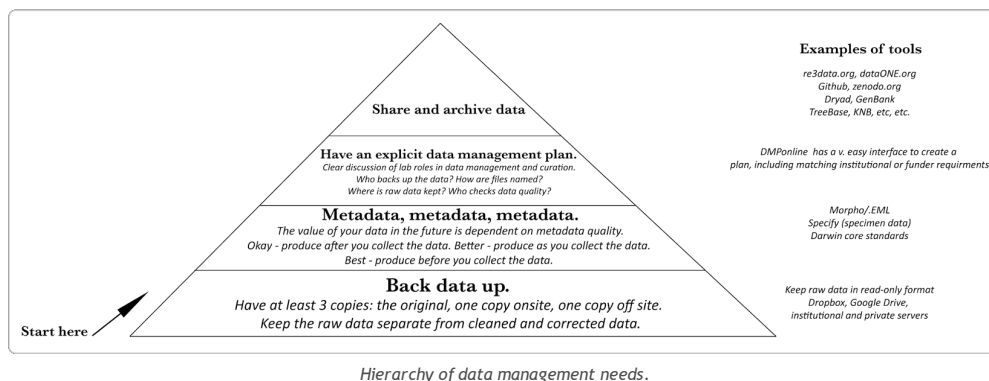
Thursday, March 9, 2017

Data management for complete beginners

Bill Michener is a longtime advocate of data management and archiving practices for ecologists, and I was lucky to catch him giving talk on the topic this week. It clarified for me the value of formalizing data management plans for institutions and lab groups, but also the gap between recommendations for best practices in data management and the reality in many labs.

Michener started his talk with two contrasting points. First, we are currently deluged by data. There is more data available to scientists now than ever, perhaps 45000 exabytes by 2020. On the other hand, scientific data is constantly lost. The longer since a paper is published, the less likely its data can be recovered (one study he cited showed that data had a half life of 20 years). There are many causes of data loss, some technological, some due to changes in sharing and publishing norms. The rate at which data is lost may be declining though. We're in the middle of a paradigm shift in terms of how scientists see our data. Our vocabulary now includes concepts like 'open access', 'metadata', and 'data sharing'. Many related initiatives (e.g. GenBank, Dryad, Github, GBIF) are fairly familiar to most ecologists. Journal policies increasingly ask for data to be deposited into publicly available repositories, computer code is increasingly submitted during the review process, and many funding agencies now require statements about data management practices.

This has produced huge changes in typical research workflows over the past 25 years. But data management practices have advanced so quickly there's a danger that some researchers will begin to feel that it is unobtainable, due to the level of time, expertise, or effort involved. I feel like sometimes data management is presented as a series of unfamiliar tools and platforms (often changing) and this can make it seem hard to opt in. It's important to emphasize good data management is possible without particular expertise, and in the absence of cutting edge practices and tools. What I liked about Michener's talk is that it presented practices as modular ('if you do nothing else, do this') and as incremental. Further, I think the message was that this paradigm shift is really about moving from a mindset in which data management is done *posthoc* ('I have a bunch of data, what should I do with it?') to considering how to treat data from the beginning of the research process.



Once you make it to 'Share and archive data', you can follow some of these great references.

Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. (2016) Ten Simple Rules for Digital Data Storage. *PLoS Comput Biol* 12(10): e1005097. doi:10.1371/journal.pcbi.1005097

James A. Mills, et al. Archiving Primary Data: Solutions for Long-Term Studies, *Trends in Ecology & Evolution*, Volume 30, Issue 10, October 2015, Pages 581-589, ISSN 0169-5347.

<https://software-carpentry.org/blog/2016/11/reproducibility-reading-list.html> (lots of references on reproducibility)

Home

Seasons greeting caRds

Blog Archive

▼ 2017 (6)

▼ March (1)

Data
managemen
t for
complete
beginners

► February (2)

► January (3)

► 2016 (38)

► 2015 (51)

► 2014 (48)

► 2013 (83)

► 2012 (30)

► 2011 (43)

► 2010 (33)

► 2009 (70)

► 2008 (6)

Search This Blog

 Search

Subscribe to RSS feed

Subscribe in a reader

Follow by Email

 Submit

Twitter

Follow @EEB_Flow

Logo

K.A.S. Mislan, Jeffrey M. Heer, Ethan P. White, *Elevating The Status of Code in Ecology*, *Trends in Ecology & Evolution*, Volume 31, Issue 1, January 2016, Pages 4-7, ISSN 0169-5347.

Thanks to Matthias Grenié for discussion on this topic.

Posted by Caroline Tucker at 3/09/2017 03:44:00 AM



 Recommend this on Google

Labels: Academic life, data, data management, data sharing, open access

1 comment:

Anonymous said...

THANKS! BUT still in how many universities is the concept and use of meta data from the beginning! taught? Working at a research institute with a full unite dedicated to it, researchers still hesitate to use meta data and take the effort to annotated their work properly. cheers Marten

March 9, 2017 at 5:32 AM

[Post a Comment](#)

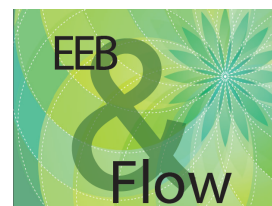
Links to this post

[Create a Link](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)



Blog roll

Dynamic Ecology

Poll: What does the "Are you willing to review a revised manuscript" question mean to you?
1 day ago

Scientist Sees Squirrel

How to handle a useless review
2 days ago

EcoTone

How is a good ecologist to react to the new administration? Letter from ESA President David Lodge
2 days ago

Eco-Evo Evo-Eco

Maladaptation to chemical exposure – what may be happening and where do we go from here?
5 days ago

ECOLOGY BITS

First steps
2 weeks ago

Denim & Tweed

The Yoder Lab opens in fall 2017
1 month ago

BioDiverse Perspectives

How would we survive without Meiosis??
8 months ago

Thoughts For Breakfast | Some topics that two scientists chat about

Hiring postdocs: the PI perspective
11 months ago

Oikos Blog

How does multiple climate variables and consumer diversity loss together "filter" natural communities?
2 years ago

Blogging from Canadian
perspectives



Statcounter

214274

Awesome Inc. theme. Powered by [Blogger](#).