Do Open Science

Julien Colomb, Jon Tennant, Peter Grabitz, Anna Krystalli

About this document

The document is meant to be read as a **pdf**, while it is written with **rmarkdown**.

I am looking for help: co-authors, reviewers and commentators are welcome, see https://github.com/jcolomb/doopenscience.git to see how you could help.

food for thoughts

check the opensciencesskills document (another github repo of J colomb)

Audience

PhD students with no coding skills ?

Timing

Goal is to set the workshop up for end of mai and do it middle of June

Topics

1. Version control (git)
2. Documenting (markdown)
3. Metadata and standards

how

- Getting contest (best solution to a problem) via github fork and pull
- Work with participant private data with one survey before and one servey after the workshop: github to give the data (same standard), R to concatenate it and ggplot2 it!
- Maybe push it to Figshare ? Data is at the same time the content we work with and the presentation of the results of the workshop.

Open data has no cost

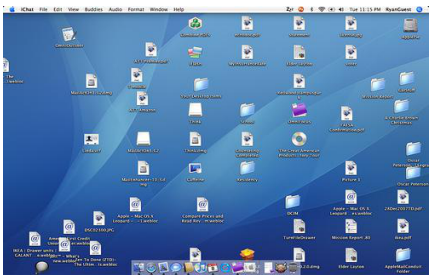
Organizing data for it to be shared represents a lot of work. This is often presented as an extra cost for scientists, there would be a cost to share own's data. I do not agree: there is no cost in making own's data sharable, but there is a time **investment**. An analogy will best show my point: The very basic rule in a molecular biology lab is: tidy your bench, make sure that your solutions, samples and equipment are kept in

the right place and temperature and that the equipment is well maintained. It will save you time for your next experiment (finding the right reagents and your tools). The same apply for data: **tidy your data files**, make sure that your data is kept in the right place and format and that your analysis tools are well maintained.

The amount of time saved by organizing your data overwhelm by far the time invested to organize it. Everybody aggrees that the extra cost to share organized data is tiny. In conclusion, Open data is not costly, it is only a question of training and getting the right habits.

You won't let your bench messy.

Why do you let your files be?



Tidy up you data and analysis code!

That's just efficient scientific practice

Vestion control for one's work

You would never think of labeling a chemical solution or a sample with a name like `ringer_solution_final_final`, but you have probably done that with reports and other documents on your computer. And you ended up with problems, didn't you?

Version control is there to avoid that. The same way as you would record the change of your chemical solution on a piece of paper, you shall record the change of your digital documents in a separate document, writing a comment for each new version about what was changed and why.

The good news is that you do not need to do that by hand, there is a program to help you with the task, it is called `Git`.

git explanation here: http://backlogtool.com/git-guide/en/intro/intro1_1.html

In addition, there is no need to code, you can use github desktop (for example)!

collaboration using Git and Github

What is great with version control is when you work collaboratively. Let's bring it to test. source: http://mozillascience.github.io/working-open-workshop/github_for_collaboration/

- fork the https://github.com/jcolomb/doopenscience_workshop
- clone it on your PC
- make a new branch with your name
- copy the survey/template.csv file and name it with "YOURNAME_before"
- open the file and enter your answers (use openoffice or libreoffice)
- save the .csv file: delimiter = "," decimal = ".", encoding = "UTF-8"
- commit, push
- make a pull request
- answer putative questions, make necessary changes,...
- yeah, you have collaborated!