Aalto University
School of Science

Git and version control tools at BECS

how to use them to keep your code in order and to keep your mind sane

Enrico Glerean (MSc), Brain & Mind Lab, BECS, Aalto University
www.glerean.com | @eglerean | becs.aalto.fi/bml | enrico.glerean@aalto.fi

What?

WHAT IS GIT?

What is git?

- **Git** /gɪt/ is a **distributed version control** and **source code management** (SCM) system
- Git is a **Finnish invention**! (By Linus Torvalds, the inventor of Linux)
- Git is British English slang roughly equivalent to "**unpleasant person**"
- **Linus explains**: *"I'm an egotistical bastard, and I name all my projects after myself. First 'Linux', now 'git'."*

What?

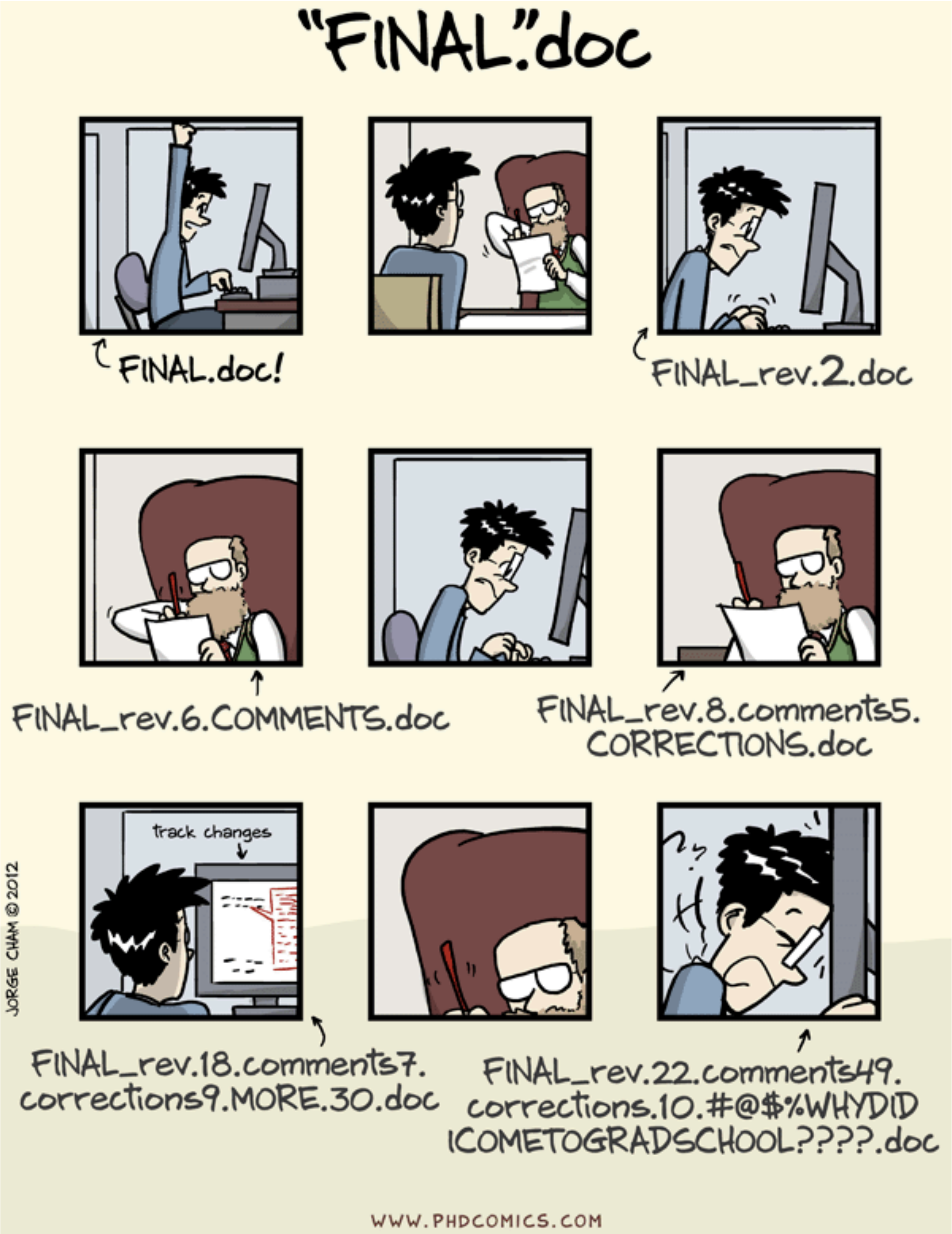
What is a version control system?

Revision (or version) control

Version control is the management of changes to documents, computer programs, large web sites, and other collections of information

Examples of version control

- **Version control on file-system**
  - *Folders with meaningful names*
  - *Files with meaningful names*
  - *Use modification date to see which one was the recent version*


"FINAL".doc
FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.CORRECTIONS.doc
track changes
FINAL_rev.18.comments7.corrections9.MORE.30.doc
FINAL_rev.22.comments49.corrections.10.#@$%WHYDIDICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

Examples of version control

- **Version control on file-system**
  - *Folders with meaningful names*
  - *Files with meaningful names*
  - *Use modification date to see which one was the recent version*
  - ***Requires loads of brain energy***
  - ***Often impossible for other to take over if you suddenly disappear***
  - ***Requires more work when you need to pass to other people***

Examples of version control

- **Version control of big software projects**
  - *http://www.mathworks.se/help/matlab/release-notes.html*
  - *http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/WhatsNew#anchor1*

- **Version control of wikipedia pages:**
  - *http://en.wikipedia.org/w/index.php?title=Connectome&action=history*

- **Version control is often embedded in applications**
  - *Microsoft word*

- **You can even version control back-ups**
  - *Dropbox (stores file changes up to 30 days – for free)* https://www.dropbox.com/help/11/en
  - *Time Machine for Mac users*

But the main question is…

Why?

WHY AM I LISTENING TO THIS ?

WHY DO I NEED VERSION CONTROL?

Scenarios

Case 1: Are you a programmer?

- You write **recyclable** software e.g. **modules** and **functions** that will be used also by other people
  - *E.g. you have made a matlab toolbox*
  - *You work in a company and do some software bits for a part of the whole system in the company*
- Most likely you are also **collaborating** with other programmers
- You need to **tell end users whenever you have changed something in your software**

…THEN YOU MUST USE VERSION CONTROL

Case 2: Are you aspiring to be a programmer?

- You are **learning** languages and tools for programming
- If you go to an interview for a job position on programming they will ask “What code have you written? Where’s your **software repository**?”
- Even a simple project with your scripts on a neat software repository can be used in your **CV** to show your **programming** as well as **organizational skills**

Case 3: You're not a programmer, you're not aspiring to become one, but you're working on a project

- You are working **alone** on a **small single project**, not collaborating with anyone…
  But:
  1. ***You need to document what you are doing to show your employer***
  2. ***You most likely need to hand-over what you did to somebody else***
- A **project** in science is often a **paper: you might be the responsible person for a project** from data collection to the final paper although other people will do the bits of the work (scanning, preprocessing, analysing, etc…)

How?

How is your lab notebook implemented?

Importance of Lab Notebooks

- **You can't try to chase a student who did preprocessing for you** and to find which value for the smoothing filter was used.
- **Every bit has to be stored in the notebook** even before running the actual preprocessing.
- **Version control will make sure that you have a reliable place to store all the parameters used** (as well as tracking history if there were other values attempted)
- You can then **easily publish and share** the software and scripts used
- **TLDR: Git is the digital part of your lab notebook!**

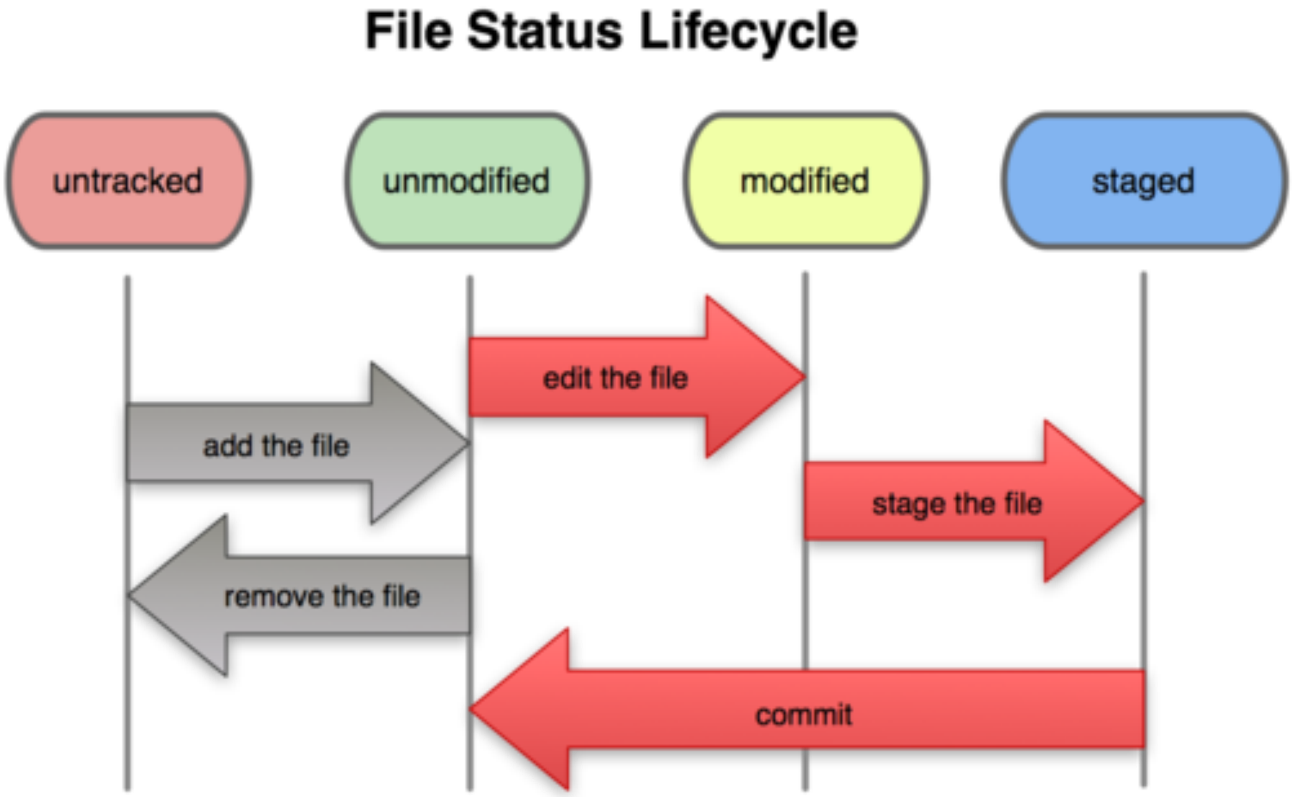
What?

WHAT IS GIT?

What is git?

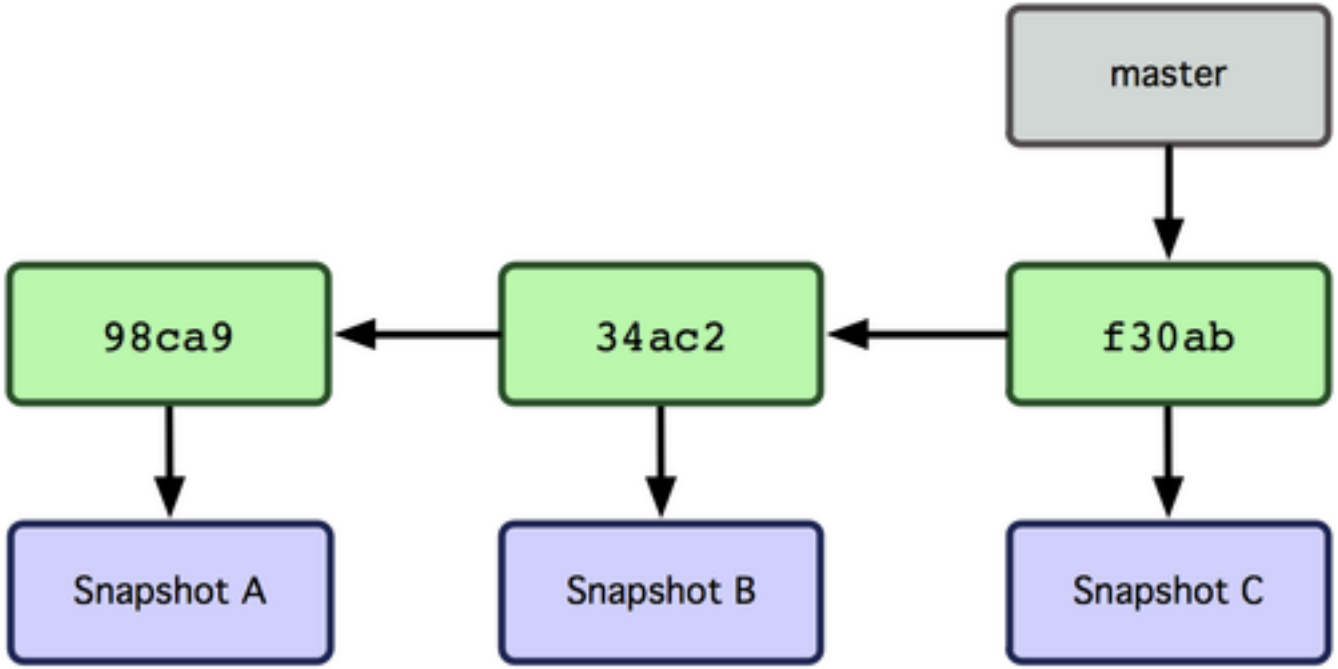
- **Git** /gɪt/ is a **distributed version control** and **source code management** (SCM) system with benefits such as:
  - ○ **Reproducibility** : logging of every step
  - ○ **Peace of mind** : a robust backup system
  - ○ **Flexibility** : zero-cost branching
  - ○ **Collaboration** : synchronization across multiple computers/people

How does it work?



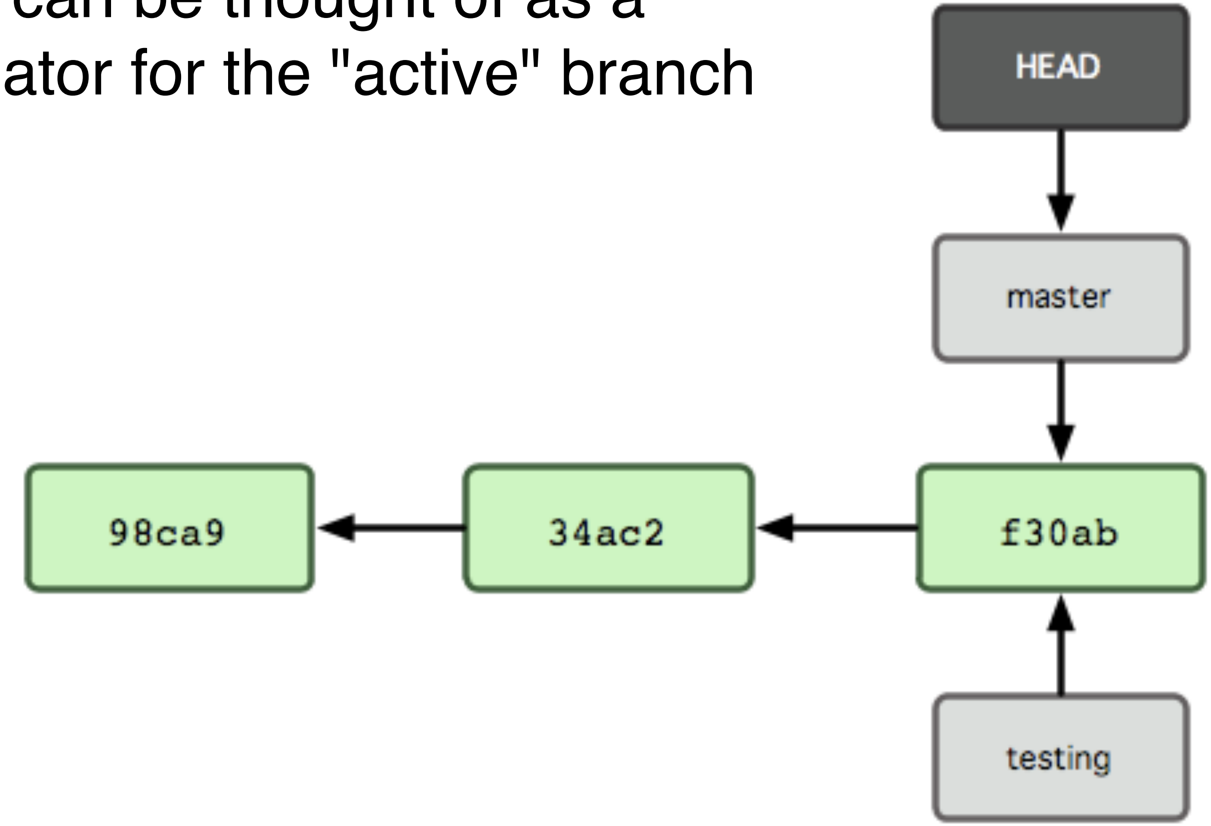
Git and branches

- Current main **branch** is the **MASTER**



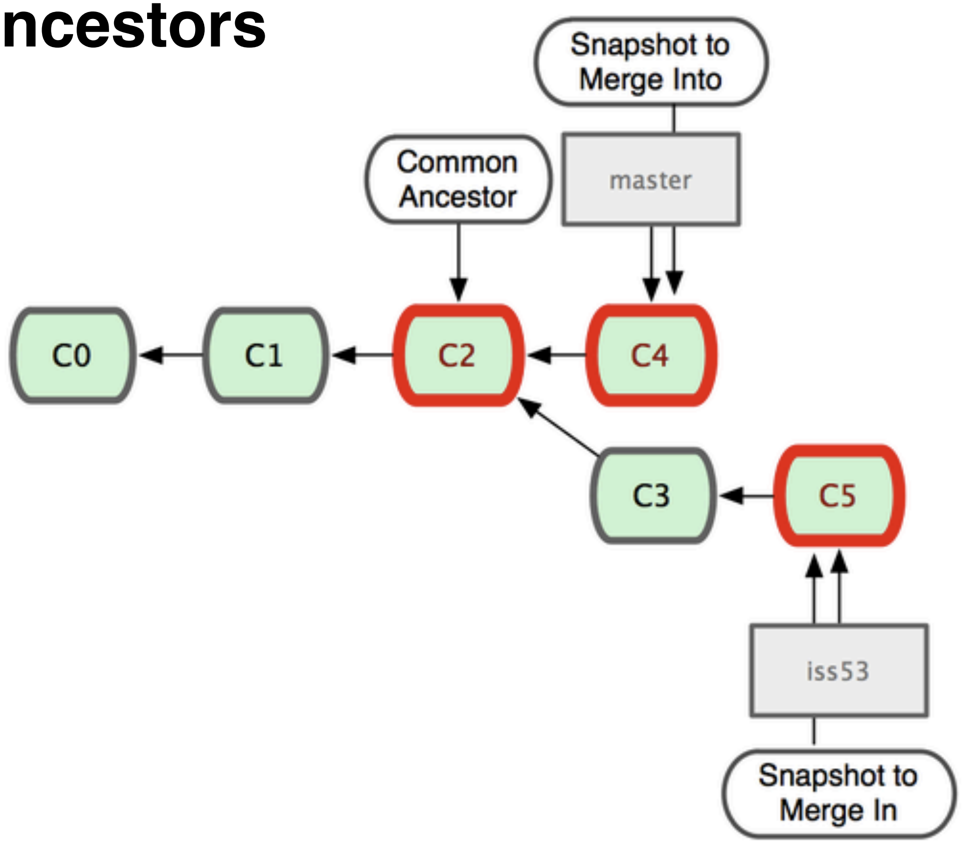
More branches and a head

- **HEAD** can be thought of as a designator for the "active" branch



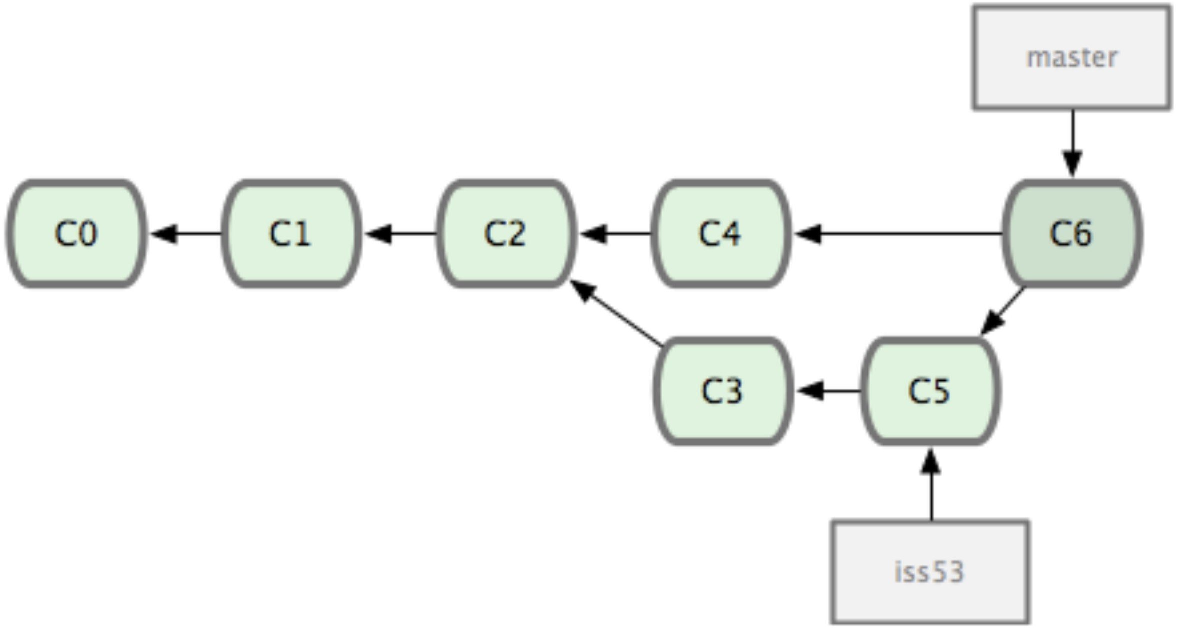
Merging

- More people (or branches) can have **snapshots with different ancestors**



Merging

- You can then merge the branches by **looking at the code and resolve conflicts**



Resources

- To get started BECS intranet pages: https://wiki.aalto.fi/display/becsintra/Gitlab
- Wiki of our **BraMiLa tools**: https://git.becs.aalto.fi/bml/bramila/wikis/home
- There you find our **Git micromanual** https://git.becs.aalto.fi/bml/bramila/wikis/git-micromanual
- **Richard** made some more advanced examples: http://rkd.zgib.net/wiki/DebianNotes/git/GitlabAndCollaboration
- You don't need to use the terminal! Gitx (mac), GitExtensions (all), Gitk (Linux)

Here's some other links

- http://arokem.github.io/2013-09-16-ISI/lessons/git-notebook/git-for-scientists.slides.html#/
- http://nyuccl.org/pages/GitTutorial/
- http://gureckislab.org/blog/?p=2521
- http://blogs.biomedcentral.com/bmcblog/2013/02/28/version-control-for-scientific-research/
- http://oss-watch.ac.uk/resources/versioncontrol
- http://www.dnacoil.com/tools/git-for-scientists-in-2-minutes/
- https://github.com/blog/1840-improving-github-for-science