Version Control for Scientists

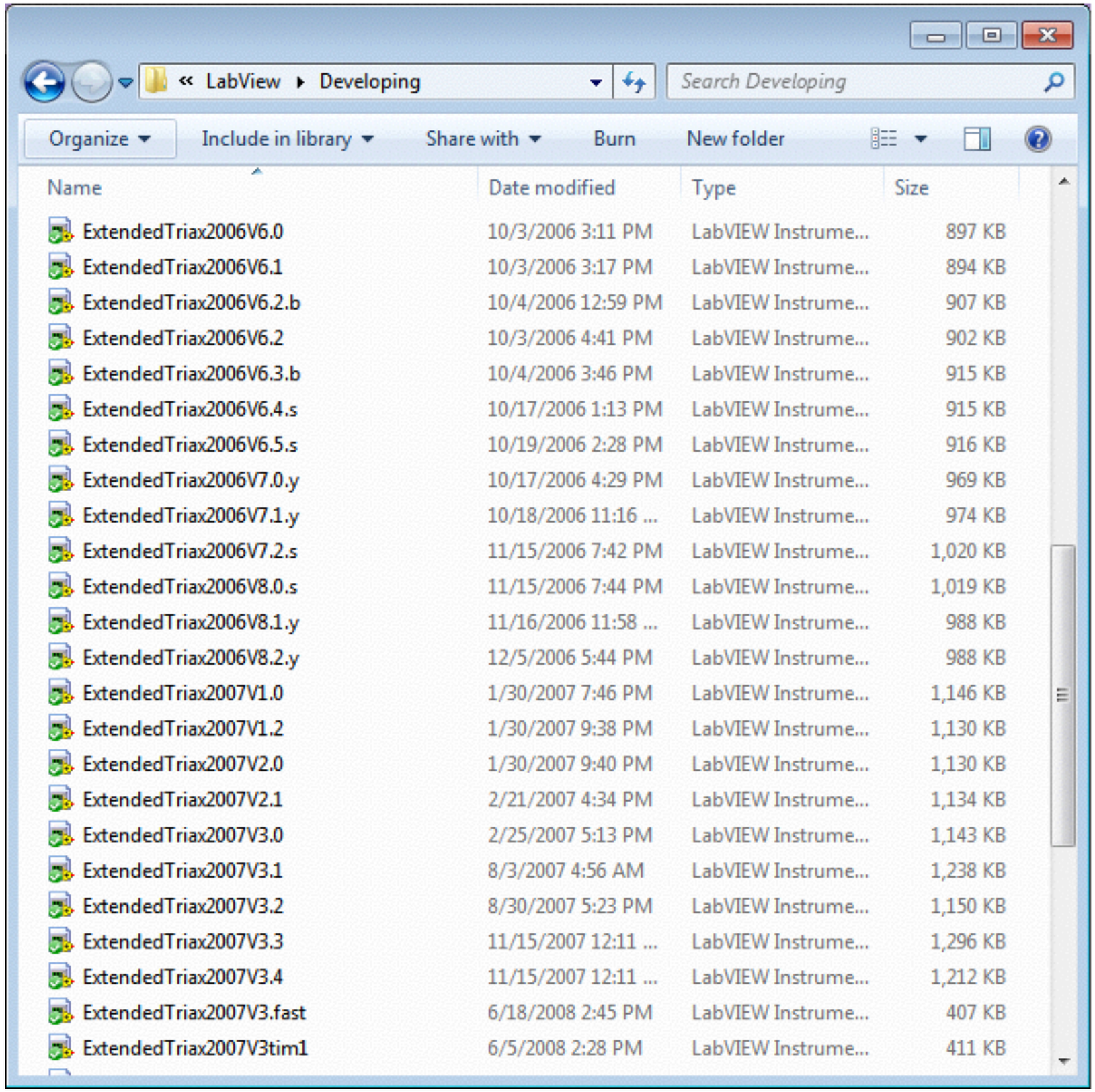
Thomas K. Baldwin

Version Control is an organizational skill

- keep files in order
- keep files backed up
- collaborate

for computer programs, papers, presentations…

- nvpaper.doc
- nvpaper_1.doc
- nvpaper_1_hailin.doc
- nvpaper_2.doc
- nvpaper_2.1.doc
- nvpaper_2.1_naturephysics.doc
- nvpaper_final.doc
- nvpaper_final_submitted.doc


« LabView ▸ Developing
Search Developing
Organize
Include in library
Share with
Burn
New folder
Name
Date modified
Type
Size
ExtendedTriax2006V6.0 10/3/2006 3:11 PM LabVIEW Instrume... 897 KB
ExtendedTriax2006V6.1 10/3/2006 3:17 PM LabVIEW Instrume... 894 KB
ExtendedTriax2006V6.2.b 10/4/2006 12:59 PM LabVIEW Instrume... 907 KB
ExtendedTriax2006V6.2 10/3/2006 4:41 PM LabVIEW Instrume... 902 KB
ExtendedTriax2006V6.3.b 10/4/2006 3:46 PM LabVIEW Instrume... 915 KB
ExtendedTriax2006V6.4.s 10/17/2006 1:13 PM LabVIEW Instrume... 915 KB
ExtendedTriax2006V6.5.s 10/19/2006 2:28 PM LabVIEW Instrume... 916 KB
ExtendedTriax2006V7.0.y 10/17/2006 4:29 PM LabVIEW Instrume... 969 KB
ExtendedTriax2006V7.1.y 10/18/2006 11:16 ... LabVIEW Instrume... 974 KB
ExtendedTriax2006V7.2.s 11/15/2006 7:42 PM LabVIEW Instrume... 1,020 KB
ExtendedTriax2006V8.0.s 11/15/2006 7:44 PM LabVIEW Instrume... 1,019 KB
ExtendedTriax2006V8.1.y 11/16/2006 11:58 ... LabVIEW Instrume... 988 KB
ExtendedTriax2006V8.2.y 12/5/2006 5:44 PM LabVIEW Instrume... 988 KB
ExtendedTriax2007V1.0 1/30/2007 7:46 PM LabVIEW Instrume... 1,146 KB
ExtendedTriax2007V1.2 1/30/2007 9:38 PM LabVIEW Instrume... 1,130 KB
ExtendedTriax2007V2.0 1/30/2007 9:40 PM LabVIEW Instrume... 1,130 KB
ExtendedTriax2007V2.1 2/21/2007 4:34 PM LabVIEW Instrume... 1,134 KB
ExtendedTriax2007V3.0 2/25/2007 5:13 PM LabVIEW Instrume... 1,143 KB
ExtendedTriax2007V3.1 8/3/2007 4:56 AM LabVIEW Instrume... 1,238 KB
ExtendedTriax2007V3.2 8/30/2007 5:23 PM LabVIEW Instrume... 1,150 KB
ExtendedTriax2007V3.3 11/15/2007 12:11 ... LabVIEW Instrume... 1,296 KB
ExtendedTriax2007V3.4 11/15/2007 12:11 ... LabVIEW Instrume... 1,212 KB
ExtendedTriax2007V3.fast 6/18/2008 2:45 PM LabVIEW Instrume... 407 KB
ExtendedTriax2007V3tim1 6/5/2008 2:28 PM LabVIEW Instrume... 411 KB

You're Reinventing the Wheel

The solution exists, and it is called **Version Control**

- a program that keeps track of files as they change

- invented for software development, but useful for any set of changing files

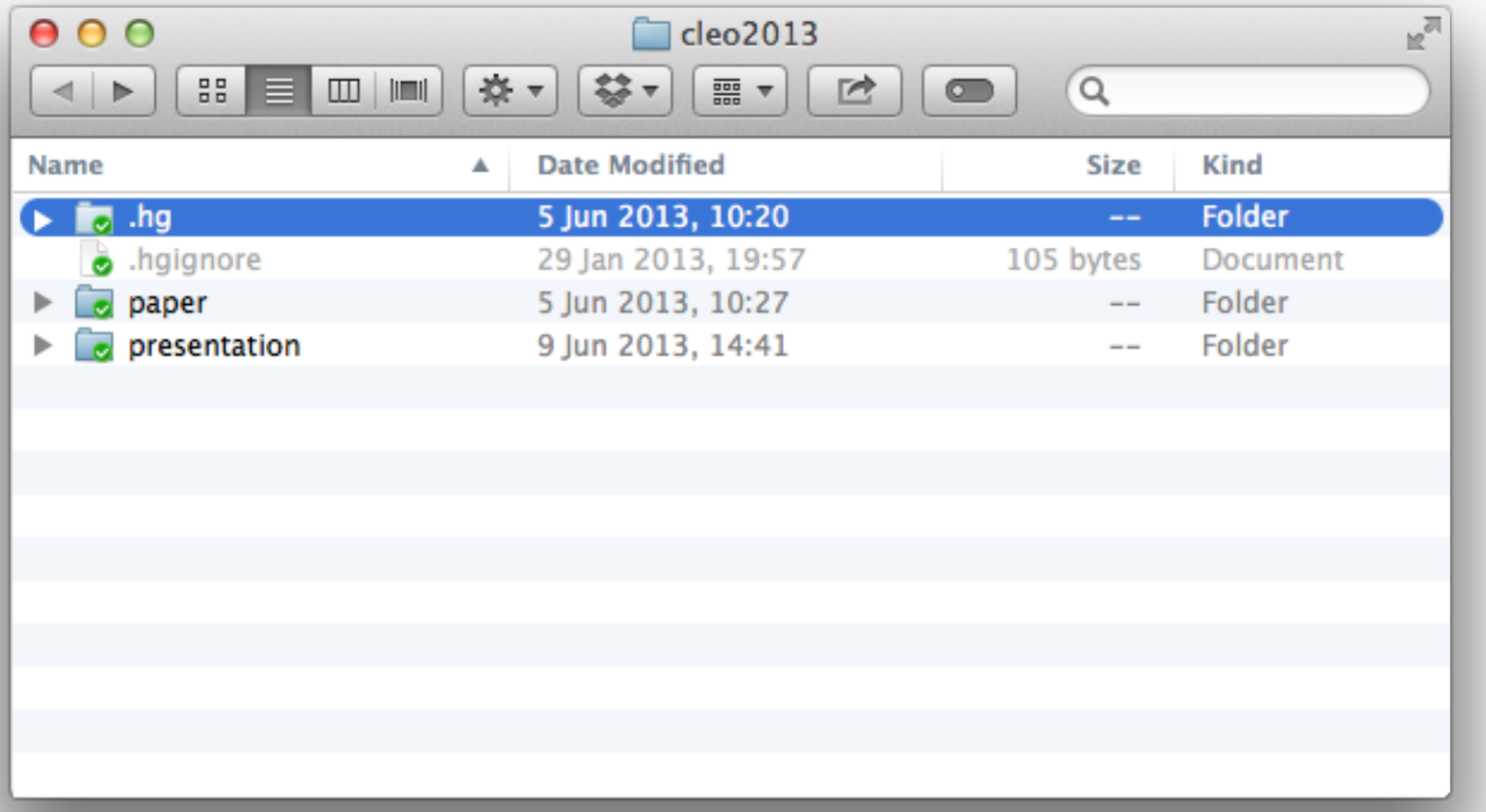
Version control programs

- **Distributed**: Git, Mercurial (hg), Bazaar (bzr)
- older: Subversion (svn), CVS, RCS
- Built-in to some file formats (e.g. Word)

The emerging standard is **git**.

The Repository

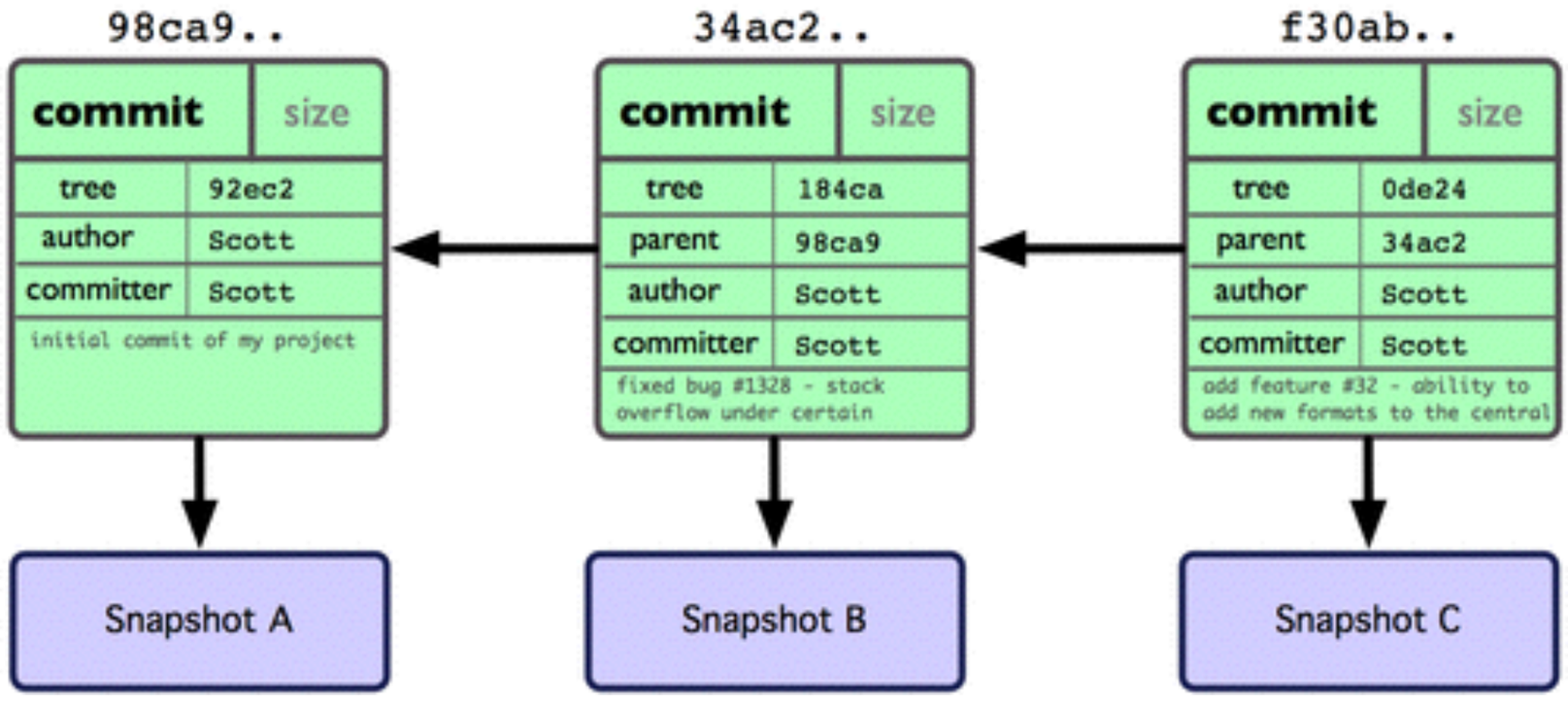
- A folder with stored history
- past versions stored in hidden subfolder (.hg or .git)
- only **added** files are tracked
- snapshots are saved when you make a **commit**



Commits

A snapshot of tracked files at a point in time.

- a unique identifier (SHA1 hash)
- a link to parent commit
- name of author/ committer
- a descriptive **message**



The commit message should briefly describe how the snapshot differs from the parent.

---

baldwint/wanglib

master ▾ | Changes | History | Branches | Sync

205 commits


Make sure density_plot inputs are arrays
3 months ago by Thomas Baldwin


Add property for pulse modulation
7 months ago by Thomas Baldwin


term_chars as keyword argument
8 months ago by Thomas Baldwin 2 ▸


add maxlen argument to plotgen
8 months ago by Thomas Baldwin


getters and setters for rf generator
11 months ago by Thomas Baldwin


density should return the image object
11 months ago by Thomas Baldwin


instead of forcing origin='lower', han...
11 months ago by Thomas Baldwin


Update plotgen to work in ipython not...
11 months ago by Thomas Baldwin

add maxlen argument to plotgen

Useful for monitors where you don't want to retain things forever


Thomas Baldwin — 475ef08 — 8 months ago

wanglib/pylab_extensions/live_plot.py

```
...  ...  @@ -43,13 +43,15 @@ This will print the data to STDOUT, but we could also:
43   43   """
44   44
45   45   from pylab import plot, gca, draw
     46 + from collections import deque
46   47
47      - def plotgen(gen, ax=None, **kwargs):
     48 + def plotgen(gen, ax=None, maxlen=None, **kwargs):
48   49       """
49   50       Take X,Y data from a generator, and plot it at the same time.
50   51
51   52       :param gen: a generator object yielding X,Y pairs.
52   53       :param ax: an axes object (optional).
     54 +     :param maxlen: maximum number of points to retain (optional).
53   55       :returns: an array of the measured Y values.
54   56
```

Demo

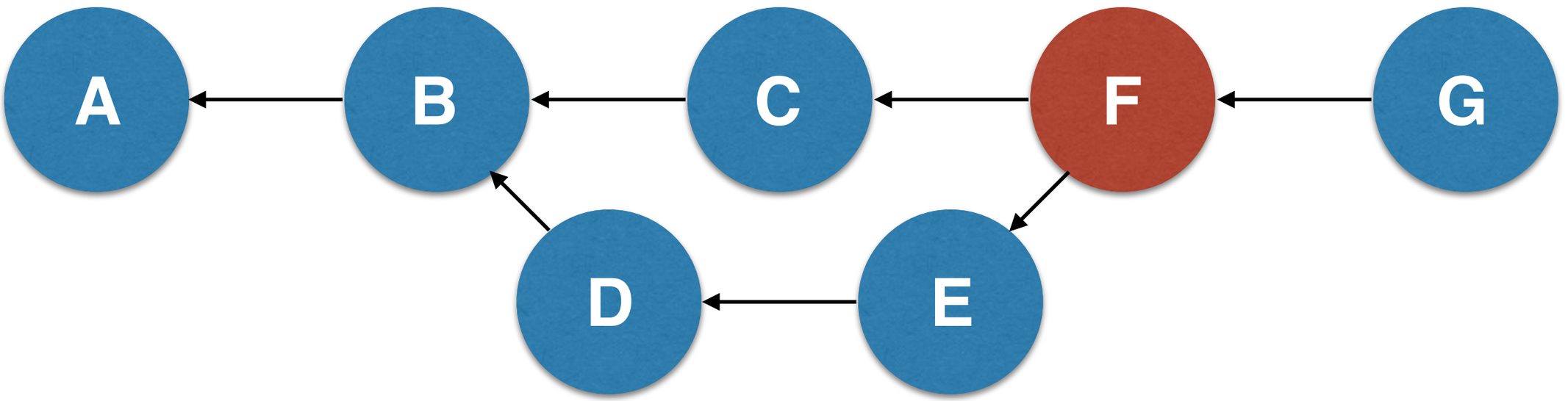
baldwint.com/vcs/important.tex

- Make a repository containing this file
- Add the file and make an **initial commit**
- Fix typos and **commit**
- Change author and **commit**
- Change date and **commit**

Advanced Topics

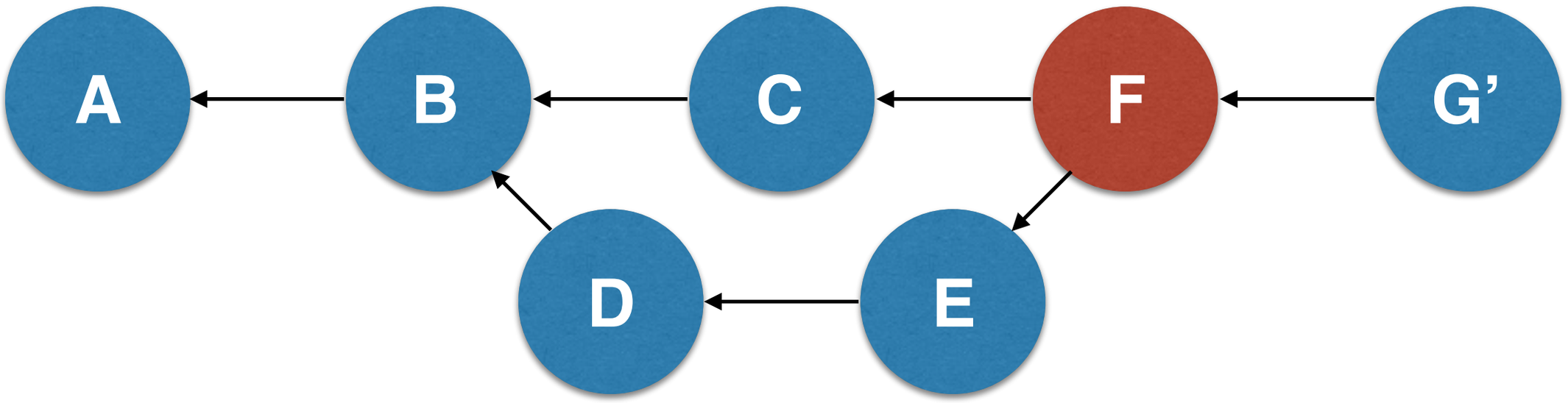
- keep files in order
- keep files backed up
- collaborate

Branches and Merges



branches: for dead ends

merges: for software dev



clones / **forks**: wholesale copies of a repository

Cloning, Pushing, Pulling

- **Cloning** copies a repository and its history.
  - This creates a **fork**. (it's a backup!)
- A **push** transfers changes from a local repository to a remote one. (A **pull** is the opposite of a push.)
- Pushes and pulls are done one branch at a time.

github.com

- **clone** your practice repository to github (“publish”)
- View it on the web
- Make a change to the local repository and **commit** it
- Is it on the web?
- **push** (“sync”) your new change to the remote copy

Collaborating with git

You can share commit privileges with others, and both collaborate on the same repository.

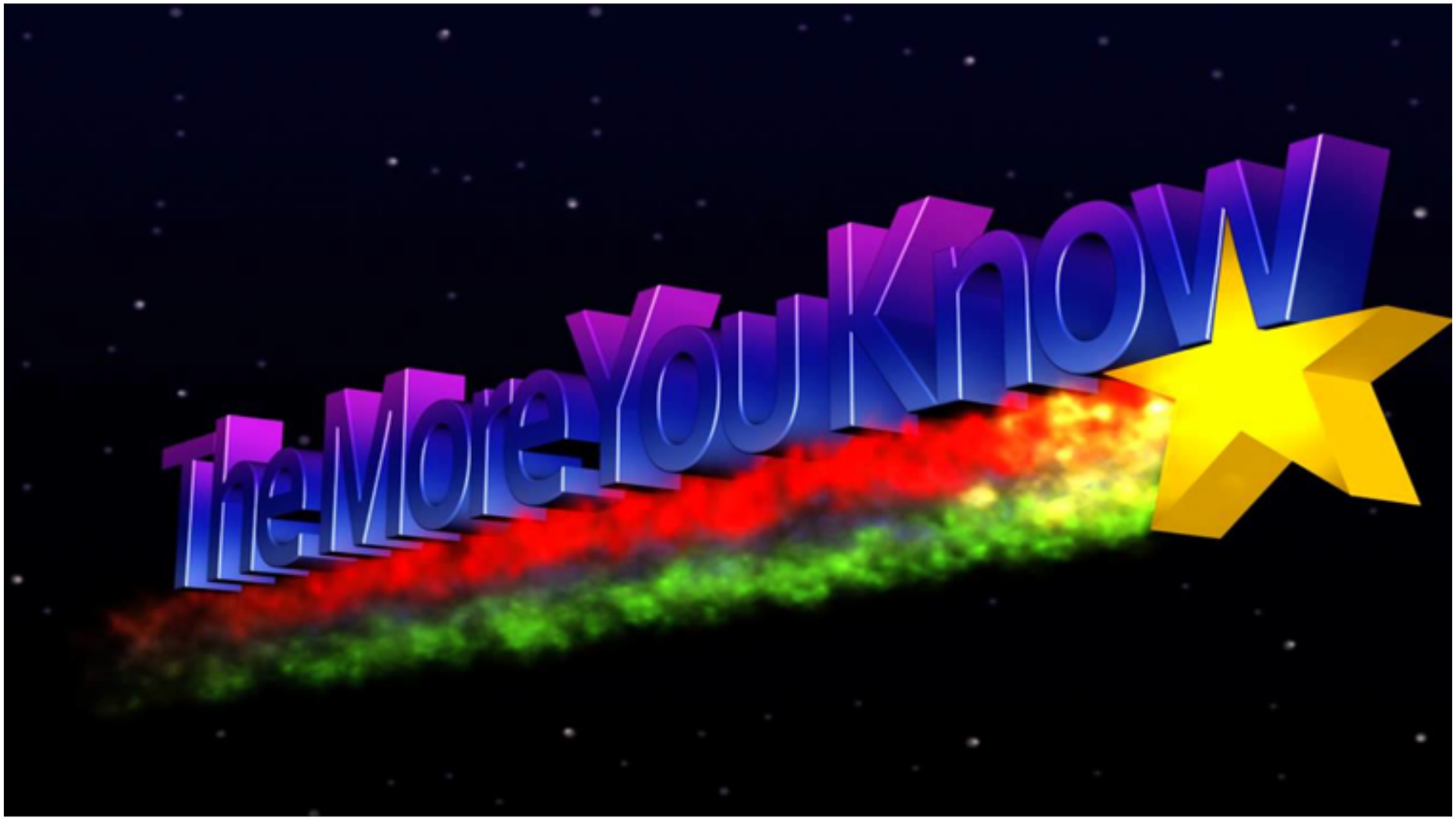
- Use github as an “official copy” and pull frequently

- If working simultaneously, use branches.

You can always commit to your own forks, but you’ll need commit privileges to push them back.

github.com/baldwint/uomlweb

- **clone** this repository to your computer
- Make a change and **commit** it
- one at a time, **push** your change to the remote copy
- I have no idea if this will work



Pro Git (free, official ebook): https://progit.org

Git for scientists: http://nyuccl.org/pages/gittutorial/

Try git without installing: https://try.github.io/