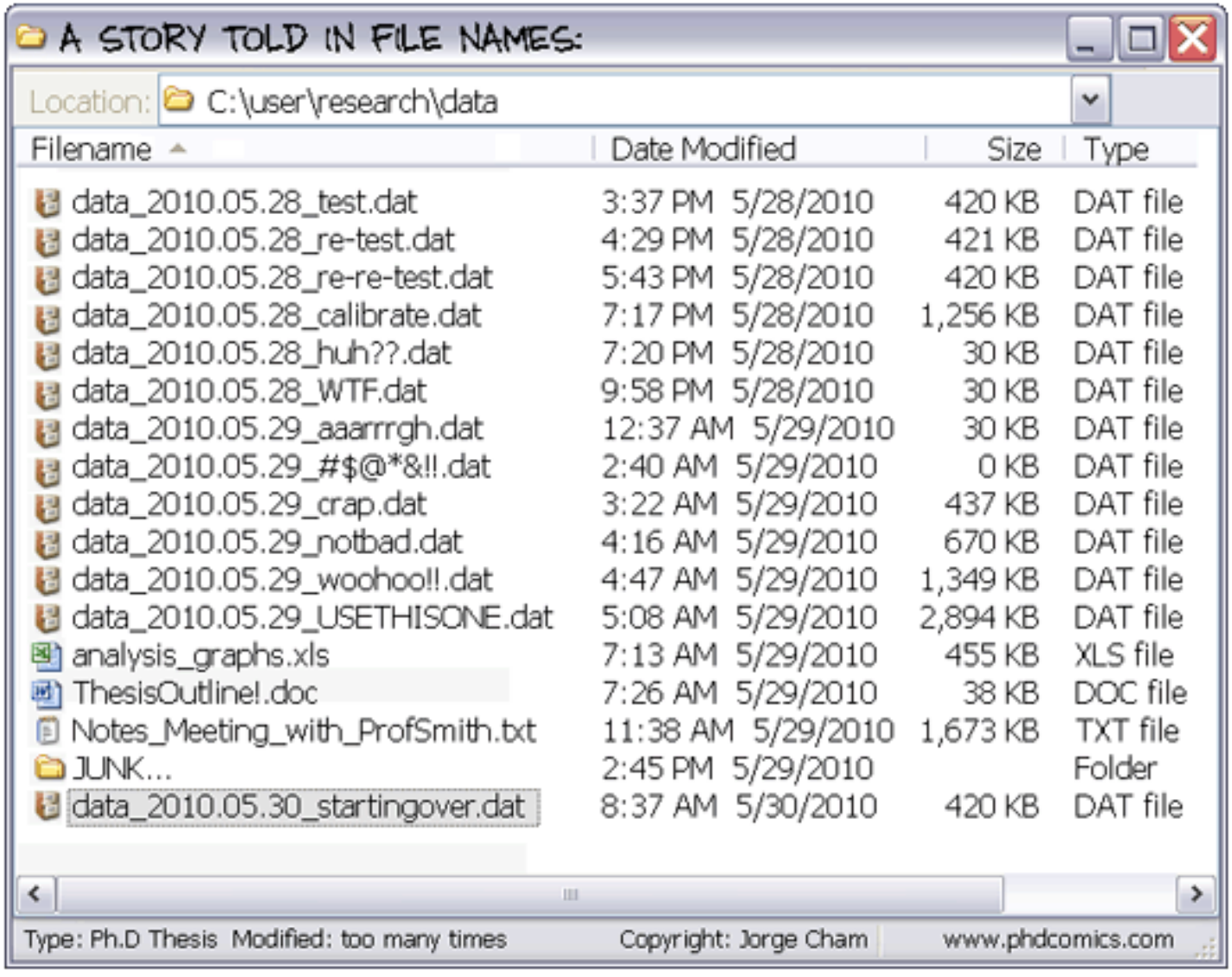
Version control with Git

2015-04-21

A common problem



Reproducibility of research

Repoducibility

- Research should be reproducible by others.
- This refers to the experiments generating the data, but also to the analysis of the data.
- The first researcher who will need to reproduce your results is likely to be you.

A lab notebook for analyses?

- Lab books make lab work traceable. Analyses should also be traceable.
- Analysis steps must be recorded, and reverting to any previous step must be possible.
- This ensures that we always exactly know how a result was generated.

A possible solution...

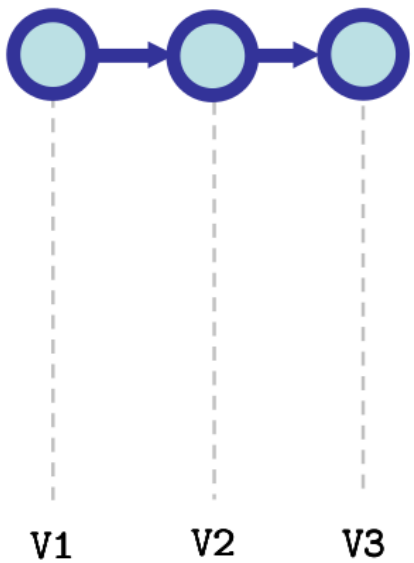
Version control

- Version control is a tool to keep track of file changes.
- However, version control softwares offer more than simply recording successive versions of a file:
  - projects can be shared with collaborators
  - projects can be forked and merged (branching)
- Interesting both for collaborative work and for single developper

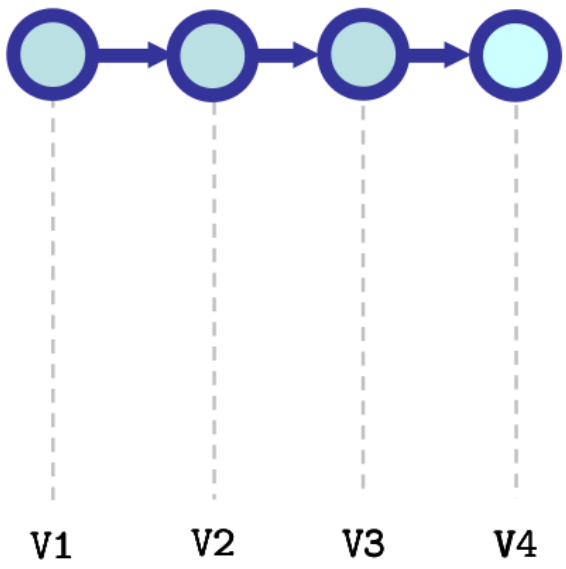
Example of a version control workflow

Example of a version control workflow

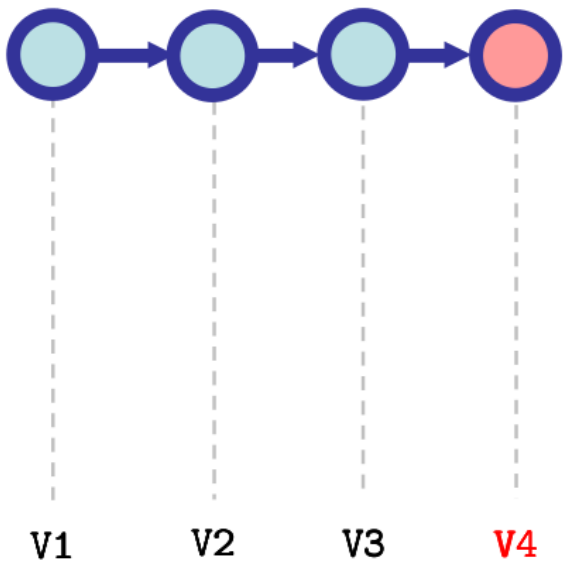
Example of a version control workflow



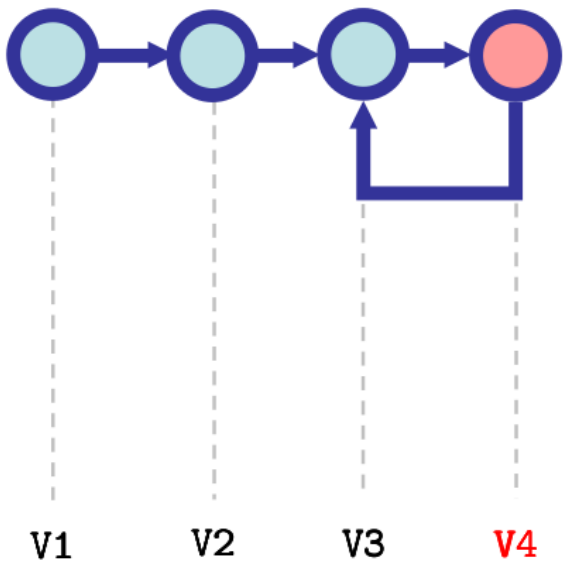
Example of a version control workflow



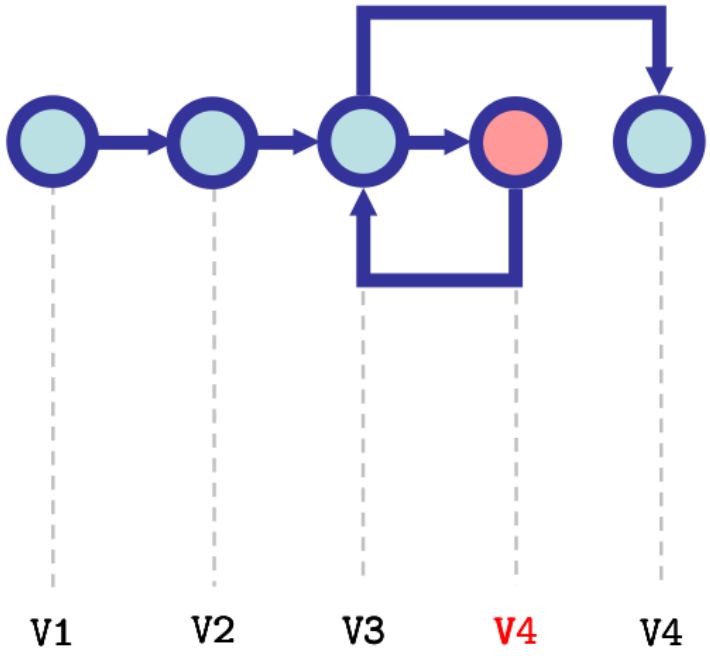
Example of a version control workflow



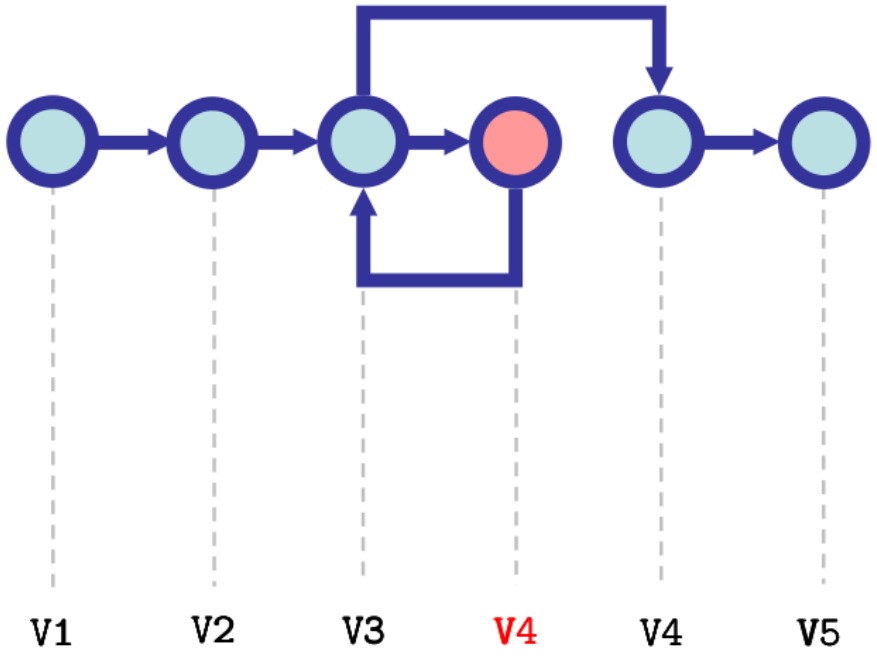
Example of a version control workflow



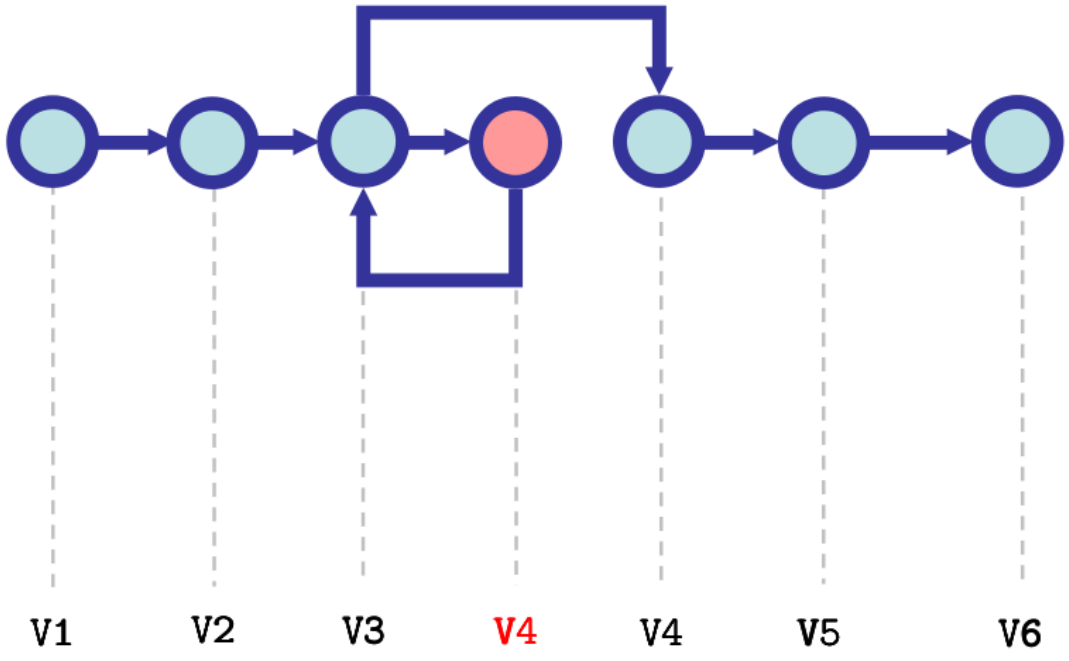
Example of a version control workflow



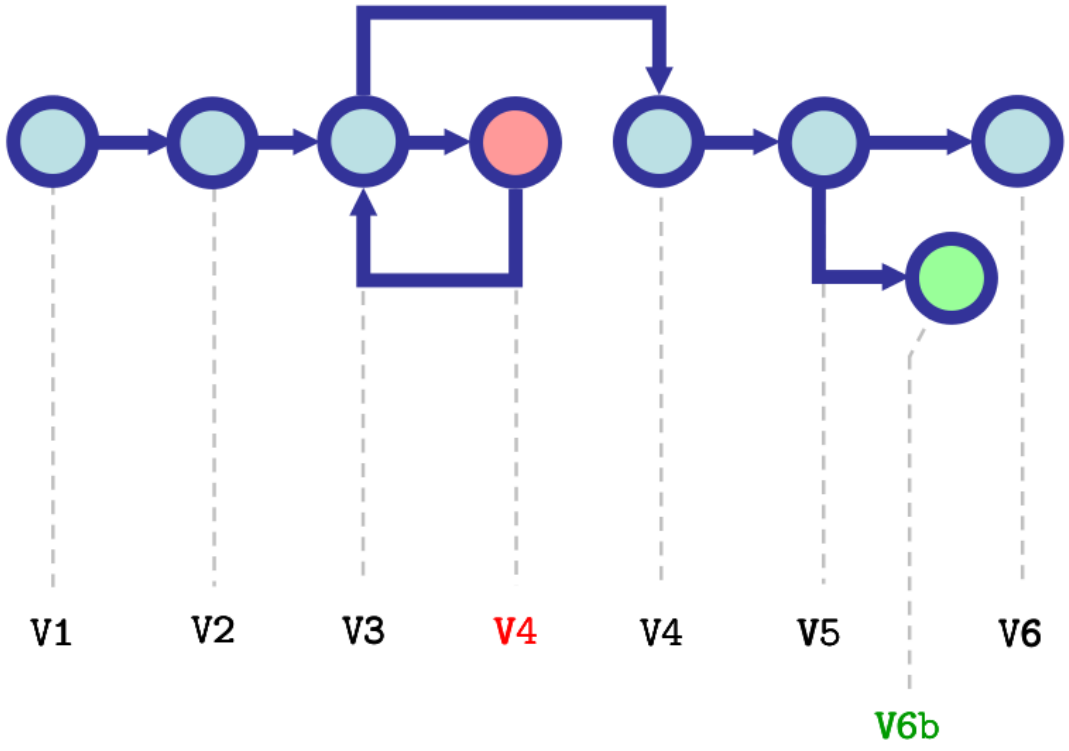
Example of a version control workflow



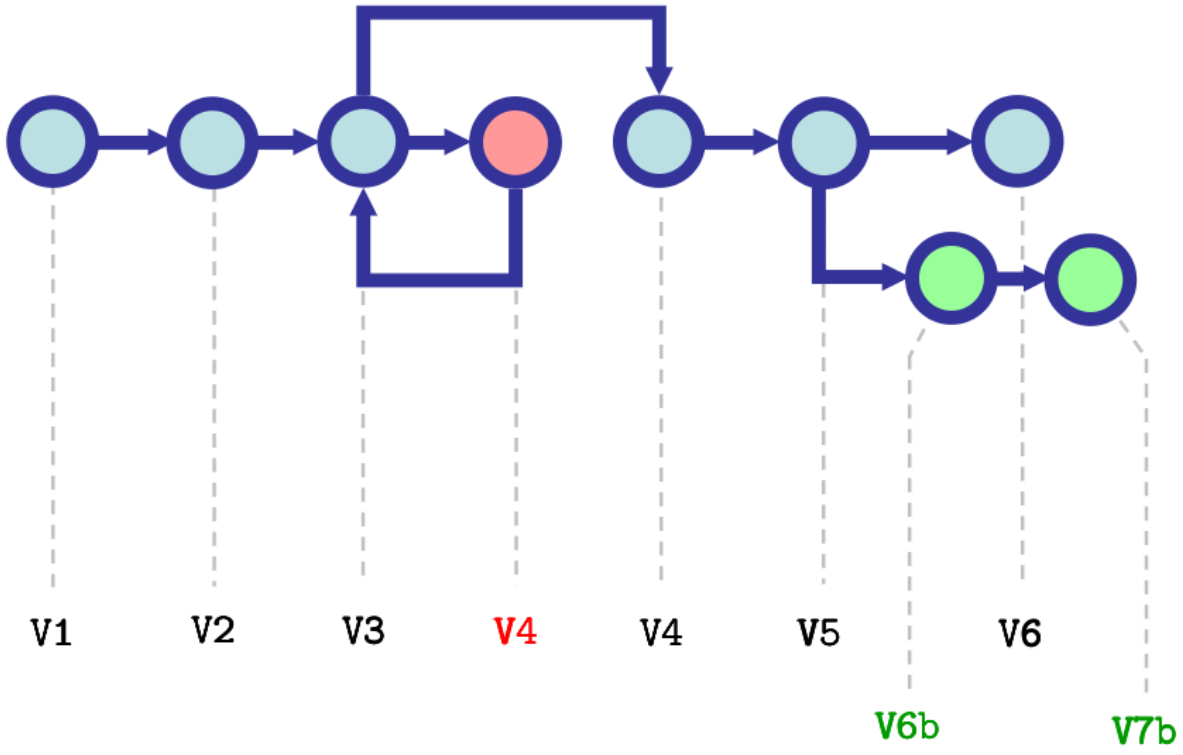
Example of a version control workflow



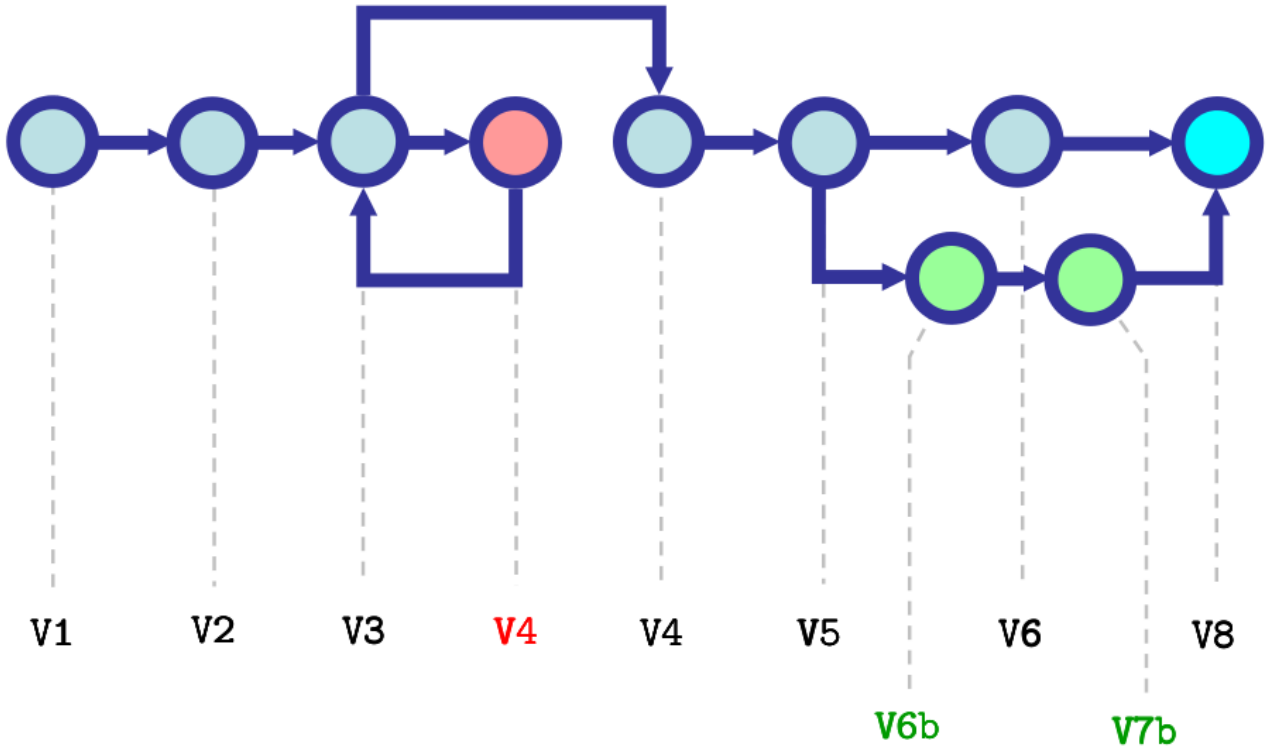
Example of a version control workflow



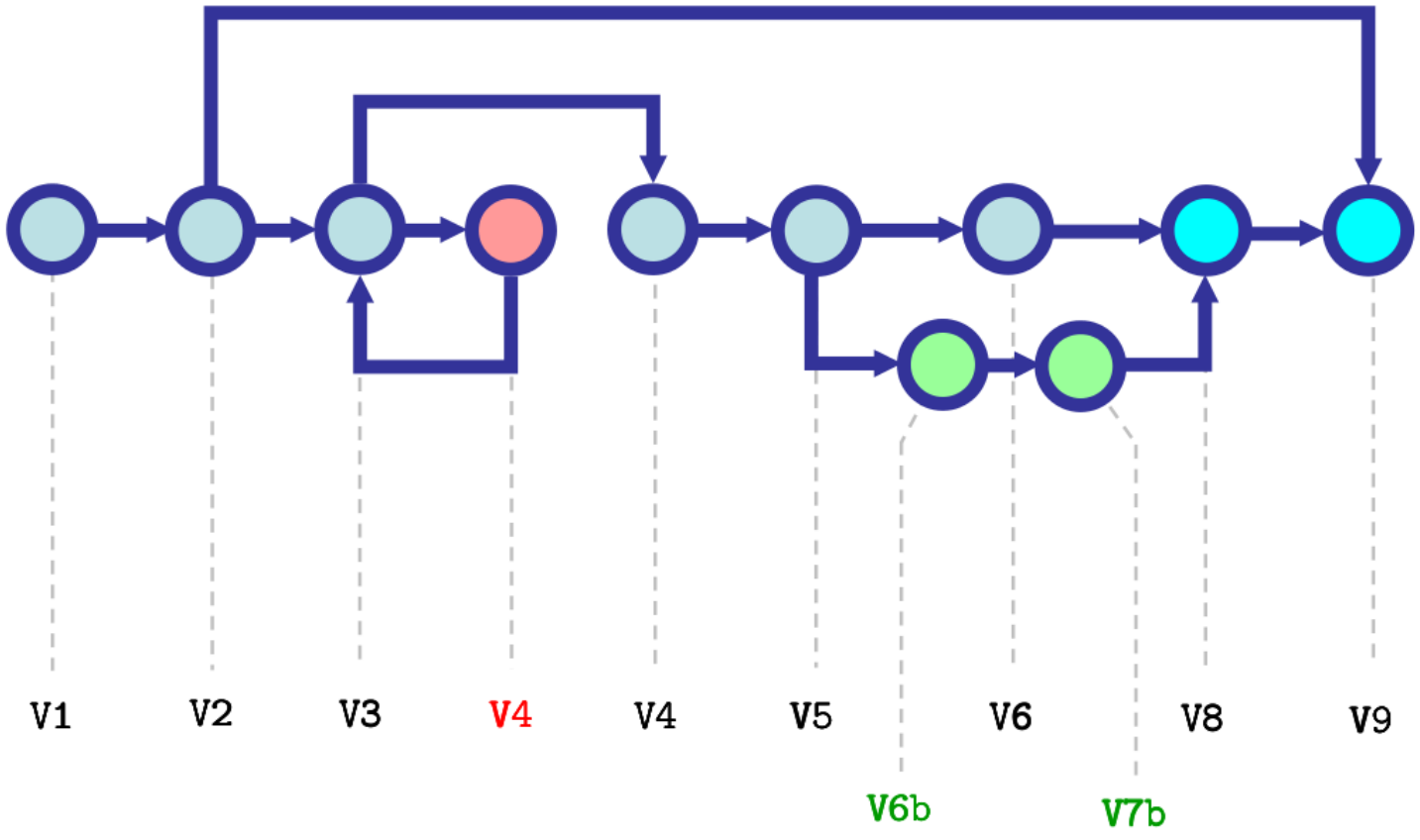
Example of a version control workflow



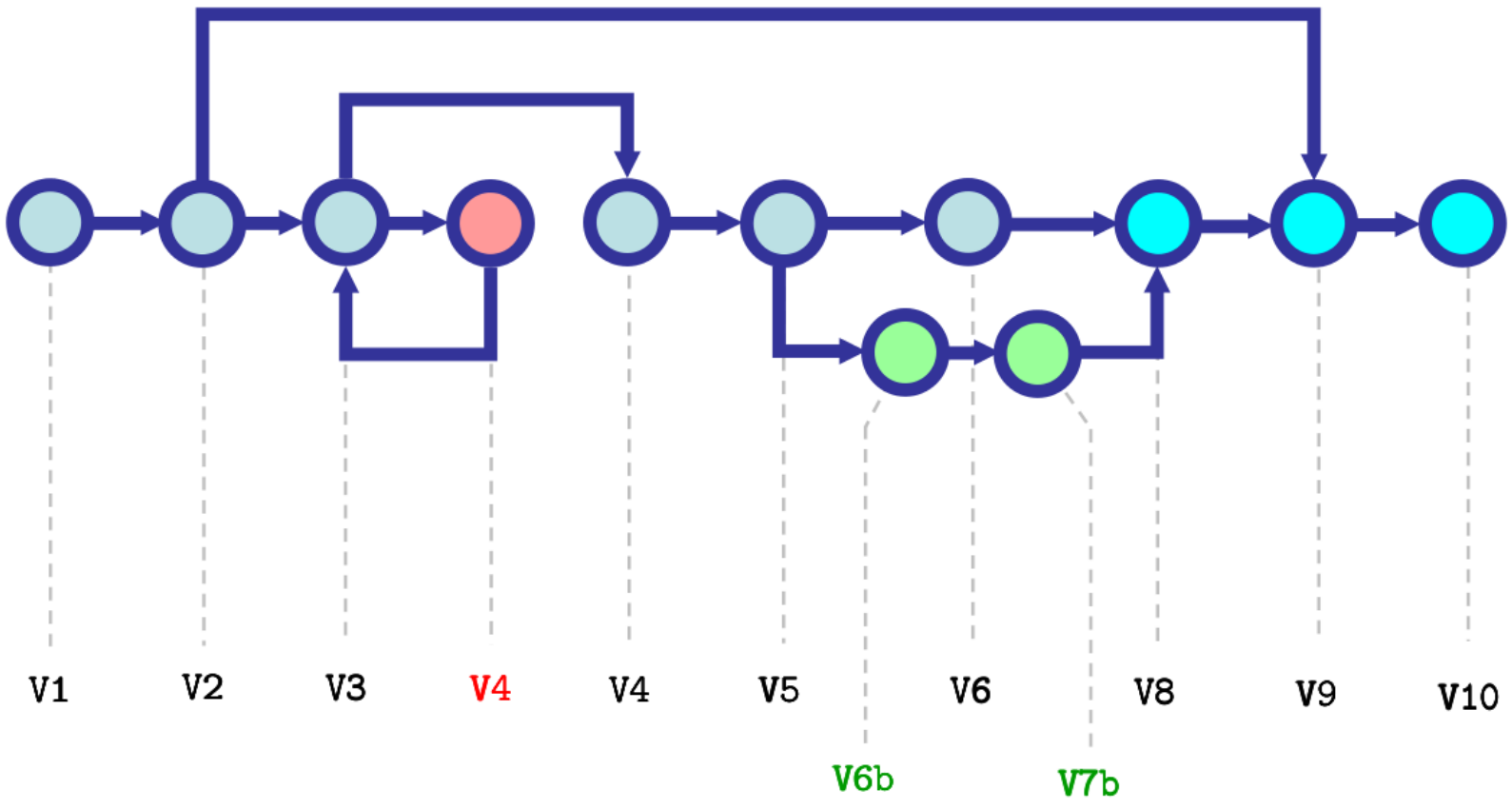
Example of a version control workflow



Example of a version control workflow



Example of a version control workflow



What are the available tools?

Existing version control tools

- Subversion
- Bazaar
- Mercurial
- Git

Online servers for repositories

- BitBucket (free private repositories)
- GitHub (free for public repositories but not for private repositories)

Git

Git

- Relatively recent but very mature
- Decentralised model
- Reputed fast

Practicals

Hands-on exercise

- You need to develop a script to analyse some sequence data
- Let's develop a small Python script together using version control
- At the end of this exercise, you will have produced:
  - a fully traceable version of your work
  - a remote repository where you can share your work

Structure of a project folder

What is version-controlled, what is not