

Code versioning & Git

GIGA Doctorate School



Program

- Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



Program

- Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



Example 1, "last version?"

Real life example

- Hey can you send me the source of that article XYZ?
- Sure, ...hum, well, ...

```
article.tar.bz2

article_final.tar.bz2

Article_final2.tar.bz2

article_final2.tar.bz2

Article_last.tar.bz2

Article_20180705_bis.tar.bz2

There it is!

No this one is more recent

Wait, this is one even more so

Hold on, that should be it

Or maybe...
```

Poor man's versioning → date & comment in archive file name



Example 2, "collaborate?"

One person in charge

Send an email with:

"Changes made:

- updated help part of file1.m
- corrected a bug in file2.m
- Added a new feature to handle .png images in file3.m

See the attached files."

- One shared file, e.g. through Dropbox or on server
 - → Incompatible parallel versions, overwritten files, lost changes,... depending on "who saved last"



Example 3, "mess with yourself!"

A simple way to "shoot oneself in the foot":

- Take a snap shot archive of current stable version commonly, = copy your code in a new folder
- 2. Begin implementing your new crazy experimental idea.
- 3. Fix some bugs in old code, revealed during testing.
- 4. Your idea was crap, discard experimental version.
- 5. Start back from stable version archive.
- 6. You lost your bug fixes, which also applied to the stable version...

 Or was it?



Why Version Control

Key questions:

- Do you work in a team?
- Has it ever happened that you were working on a file, and someone else was working on the same file at the same time? Did you lose your changes to that file because of that? Or ended up with incompatible code?
- Have you ever saved a file, and then wanted to revert the changes you made? Have you ever wished you could see what a file looked like some time ago?
- Have you ever found a bug in your project and wanted to know when that bug got into your files?

If any "Yes", then use a VC system!

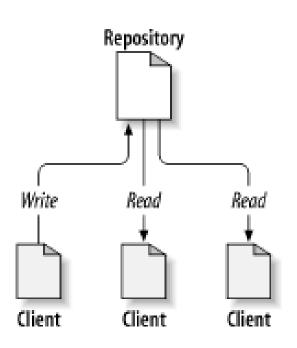


Program

- ▶ Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



Centralized file management

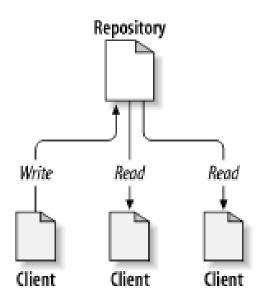


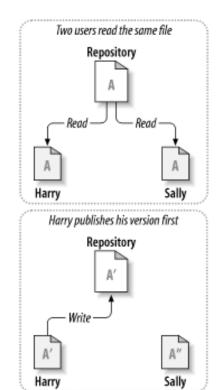
- One central repository, on a server.
- (Stores the files and their history.)
- Many clients, i.e. users, connecting to the repository.
- Each client has one or more working copies, i.e. a local copy of the files, where changes are made

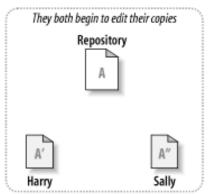


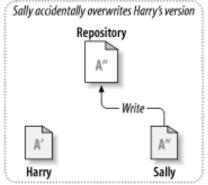
File sharing & Collaboration Problem

Centralized VC model



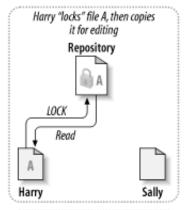


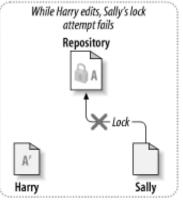


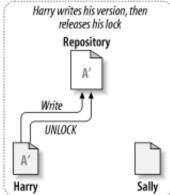


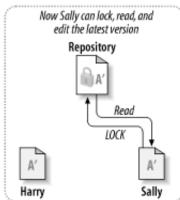


Locking solution



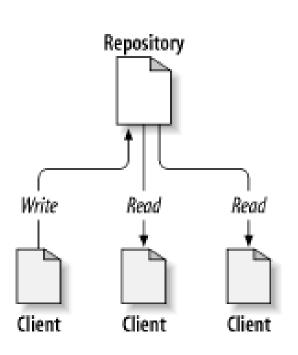








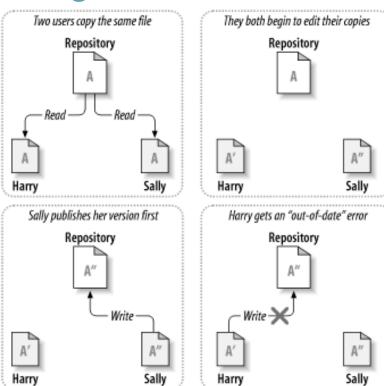
Centralized VC model



- One central repository, on a server.
- Stores the files and their history.
- Many clients, i.e. users connecting to the repo
- Each client has one or more working copies, i.e. a local copy of the files, where changes are made
- A revision identifies a point in time of the repo, it is denoted by a number.

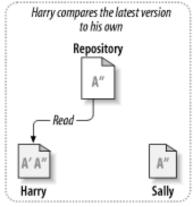


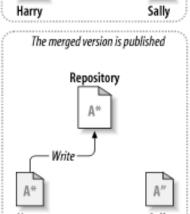
Copy-Modify-Merge Solution

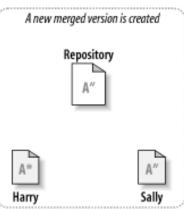


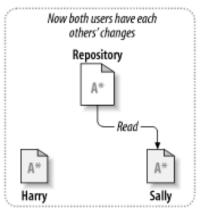


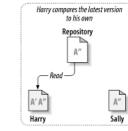
Copy-Modify-Merge Solution

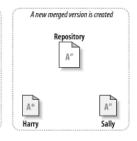












File merging & conflicts

When updating files are "updated" automatically.

Merged files:

all changes, yours & from server, are automatically merged into *your* files (if possible).

→ manual check recommended...

Conflicted files:

your changes and those on the server are NOT compatible, no automatic merging possible

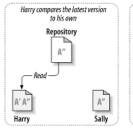
→ manual intervention necessary! **Your** responsibility.



Resolving conflicts

When updating your working copy:

- If some files have changed both in the repository and in your working copy, there can be a conflict
- It is your responsibility to fix conflicts, by inspecting the diverging changes and
 - choose your own version, or
 - choose repository version, or
 - choose previous version, or
 - mix both versions





Binary files...

- Merging works on text-based files (code/document)
- With binary files (images, .ppt, .pdf, .doc, .xls, ...)
 - → Updating overwrites the file... but previous versions still available in history!
- Use Markdown (.md), comma-/tab-separated values (.csv/.tsv) or JSON (.json) files instead of Word or Excel files!



How to...

- Create repository or get code from repository:
 - check out/clone code, or update code
- Work on your code/files:
 - bug fixes and/or new features
- Publish your changes to the repository
 - re-updating and fixing conflicts, if necessary

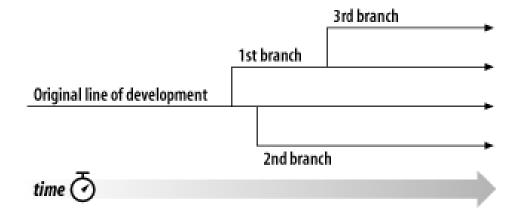
Note:

- Split your commits into logical steps
- Add description!!!



Code branch

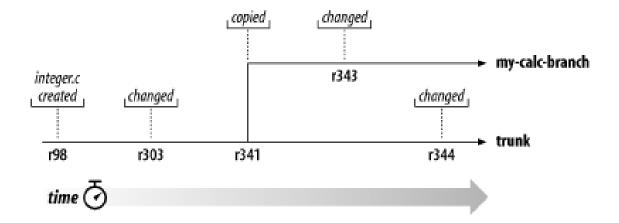
"...a line of development that exists independently of another line, yet still shares a common history if you look far enough back in time. A branch always begins life as a copy of something, and moves on from there, generating its own history."





Branching

- Work on a branch as you would on any other folder,
 e.g. code_v1, code_v2,...
- File histories in branches also stored!





Branch merging

- = synchronizing two branches
- When developing a branch, you'll want to synch with "main trunk" from time to time (e.g. for bug fixes)
- When merging, you can encounter conflicts, to be resolved as before
- If you want to integrate a branch back to "main trunk", you can merge it back (e.g. adding new features).



Program

- ▶ Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



What is "Git"?

- currently the most popular distributed versioning system
- free open-source software
- cross-platform (originally for Linux but now also on MacOS and Windows)
- very efficient, very powerful but can be very complex
- some GUIs and IDEs plugins
- no global revision numbers, "hashes" instead
- created by Linus Torvalds, 1st release in 2005



Git, pro's & con's

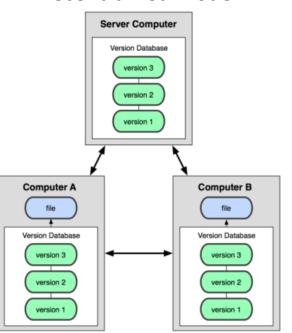
Pro's

- Every working copy is a full backup of the data
- You can work off-line
- You can do micro-commits
- Allows private work, eases experimental jump in

Cons

- More complex (decentralized → "parallel worlds")
- Less control on project evolution
- Less sharing?

Decentralized model



Sheet Cheat

Git Basics		Rewriting Git History	
git init <directory></directory>	Create empty Git repo in specified directory. Run with no arguments to initialize the current directory as a git repository.	git commitamend	Replace the last commit with the staged changes and last commit combined. Use with nothing staged to edit the last commit's message.
git clone <repo></repo>	Clone repo located at <repo> onto local machine. Original repo can be located on the local filesystem or on a remote machine via HTTP or SSH.</repo>	git rebase <base/>	Rebase the current branch onto <base/> . <base/> can be a commit ID, a branch name, a tag, or a relative reference to HEAD.
git config user.name <name></name>	Define author name to be used for all commits in current repo. Devs commonly use —global flag to set config options for current user.	git reflog	Show a log of changes to the local repository's HEAD. Addrelative-date flag to show date info orall to show all refs.
git add <directory></directory>	Stage all changes in <directory> for the next commit. Replace <directory> with a <file> to change a specific file.</file></directory></directory>	Git Branches	
git commit —m " <message>"</message>	Commit the staged snapshot, but instead of launching a text editor, use <message> as the commit message.</message>	git branch	List all of the branches in your repo. Add a <branch> argument to create a new branch with the name <branch>.</branch></branch>
git status	List which files are staged, unstaged, and untracked.	git checkout -b dranch>	Create and check out a new branch named <branch>. Drop the -b flag to checkout an existing branch.</branch>
git log	Display the entire commit history using the default format. For customization see additional options.	git merge <branch></branch>	Merge dranch> into the current branch.
git diff	Show unstaged changes between your index and working directory.	Remote Repositories	
Undoing Changes		git remote add <name> <url></url></name>	Create a new connection to a remote repo. After adding a remote, you can use <name> as a shortcut for <url> in other commands.</url></name>
git revert <commit></commit>	Create new commit that undoes all of the changes made in <commit>, then apply it to the current branch.</commit>	git fetch <remote> <branch></branch></remote>	Fetches a specific <branch>, from the repo. Leave off <branch> to fetch all remote refs.</branch></branch>
git reset <file></file>	Remove <file> from the staging area, but leave the working directory unchanged. This unstages a file without overwriting any changes.</file>	git pull <remote></remote>	Fetch the specified remote's copy of current branch and immediately merge it into the local copy.
git clean -n	Shows which files would be removed from working directory. Use	git push	Push the branch to <remote>, along with necessary commits and</remote>

<remote> <branch>



the -f flag in place of the -n flag to execute the clean.

objects. Creates named branch in the remote repo if it doesn't exist.



Program

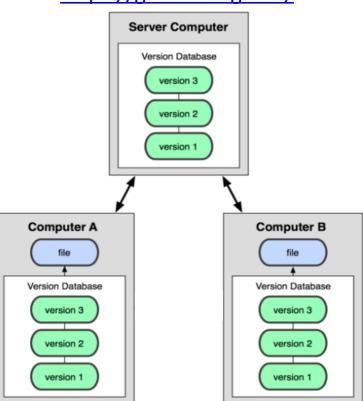
- ▶ Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



Git & GitHub/GitLab

- ▶ Git
 - "version control system" software
 - language with its commands
- ► GitHub.com (& GitLab.com)
 - web-based Git repository hosting system
 - servers from a private company
- GitLab.uliege.be
 - web-based Git repository hosting system
 - hosted at ULiège. 🙂

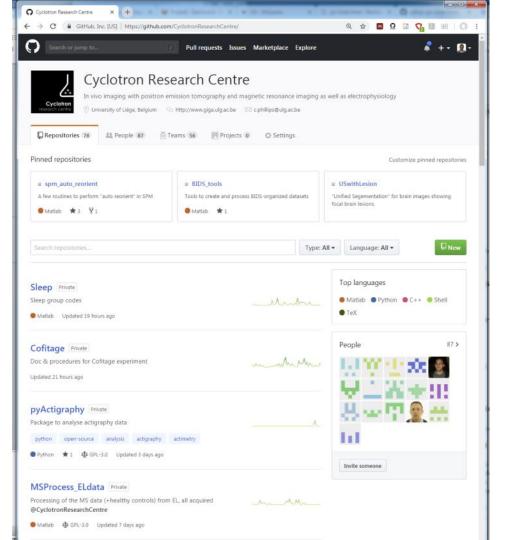
https://gitbub.com
https://gitlab.uliege.be/

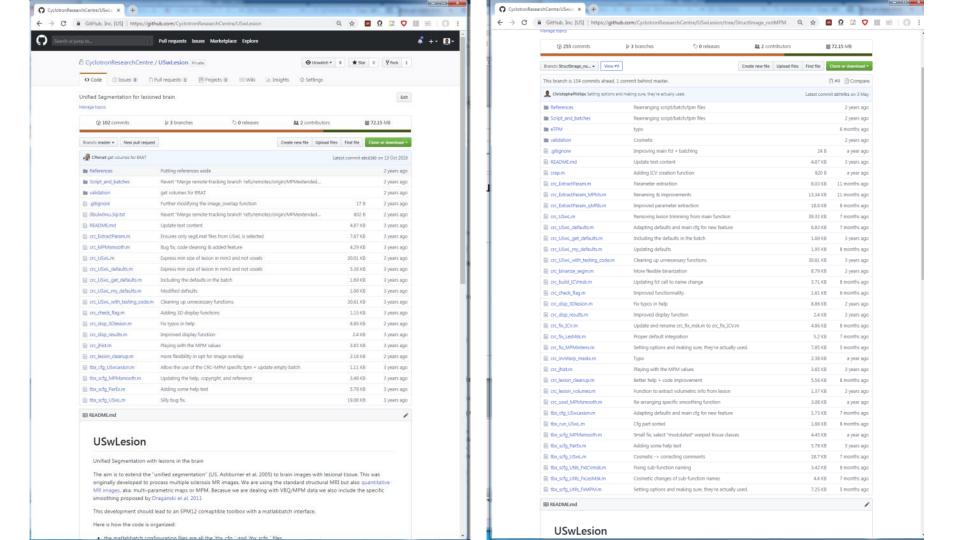


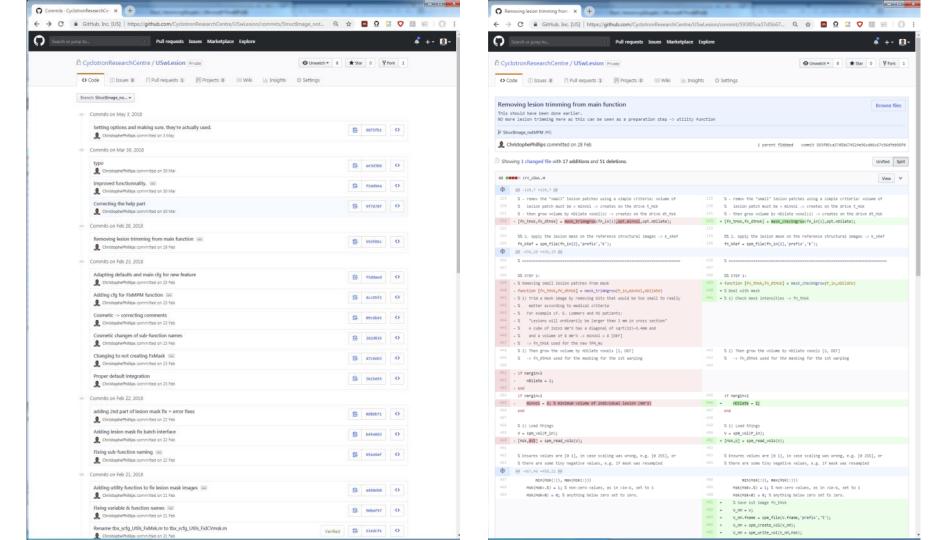


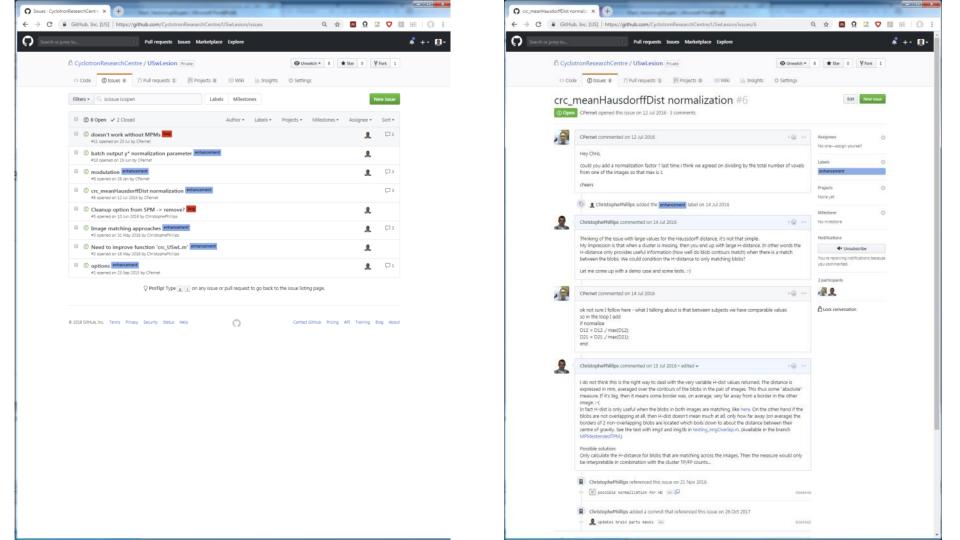
GitHub & GitLab features

- Code versioning
 - + branching, merging, releases
- Code documentation and Wiki
 - → build knowledge for the team
- Issue tracking
 - → discuss problems & requests in a forum
- Management
 - → access rights, visibility, groups/teams, ...











Program

- ▶ Why "Version Control" ?
- Basics of Version Control (VC)
- Git as a VC solution
- Being Git practical with GitHub/GitLab
- Conclusions & reference



"Reasons" not to VC?

- "It's only a small bit of code to try out an idea on my data..."
 - → This how breakthroughs happen and papers follow!
- "Nobody else will ever be interested in this..."
 - \rightarrow If you are, someone else will necessarily be!
- "My code is not ready yet..."
 - → The ULTIMATE reason to actually version your code!

Major hurdle is **psychological** or **carelessness**.



Some wisdom

"Writing software as if we are the only person that ever has to comprehend it is one of the biggest mistakes and false assumptions that can be made."

- Karolina Szczur



Code Versioning conclusion

- Absolutely necessary to manage any project that relies on code, script, batch,...
- Useful to keep track of changes, improvements & bug fixes over time
- Even more so with multiple developers/users
 - \rightarrow start alone \rightarrow team interest \rightarrow available to the community
- ▶ Open science → paper + code + data accessible



References

- J. D. Blischak et al., A Quick Introduction to Version Control with Git and GitHub, PLOS Computational Biology, 12(1): e1004668, 2016
 http://dx.doi.org/10.1371/journal.pcbi.1004668
- https://en.wikipedia.org/wiki/Version control
- https://en.wikipedia.org/wiki/Git
- https://git-scm.com/docs
- https://services.github.com/on-demand/downloads/github-git-cheat-sheet.pdf
- http://github.com/
- http://gitlab.com
- https://gitlab.uliege.be/
- Git GUI: https://desktop.github.com/ https://desktop.github.com/ https://desktop.github.com/ https://www.sourcetreeapp.com/
- https://www.campus.uliege.be/cms/c_9096862/fr/services-internet-intranet-offerts



Finally

"Programming is like pinball.

The reward for doing it is
the opportunity of doing it again."

- Unknown



Thank you for your attention!



