software carpentry

Version Control with Git

Reference

Setting Up Git

- Use `git config` to configure a user name, email address, editor, and other preferences once per machine.

Creating a Repository

- `git init` initializes a repository.

Tracking Changes

- `git status` shows the status of a repository.
- Files can be stored in a project's working directory (which users see), the staging area (where the next commit is being built up) and the local repository (where revisions are permanently recorded).
- `git add` puts files in the staging area.
- `git commit` saves the revisions in the staging area to the local repository.
- Always write a log message when committing changes.

Exploring History

- `git diff` displays differences between revisions.
- `git checkout` recovers old versions of files.

Ignoring Things

- The `.gitignore` file tells Git what files to ignore.

Remotes in GitHub

- A local Git repository can be connected to one or more remote repositories.
- Use the HTTPS protocol to connect to remote repositories until you have learned how to set up SSH.

- `git push` copies changes from a local repository to a remote repository.
- `git pull` copies changes from a remote repository to a local repository.

Collaborating

- `git clone` copies a remote repository to create a local repository with a remote called `origin` automatically set up.

Conflicts

- Conflicts occur when two or more people change the same file(s) at the same time.
- The version control system does not allow people to blindly overwrite each other's changes. Instead, it highlights conflicts so that they can be resolved.

Open Science

- Open scientific work is more useful and more highly cited than closed.

Licensing

- People who incorporate GPL'd software into their own software must make their software also open under the GPL license; most other open licenses do not require this.
- The Creative Commons family of licenses allow people to mix and match requirements and restrictions on attribution, creation of derivative works, further sharing, and commercialization.
- People who are not lawyers should not try to write licenses from scratch.

Hosting

- Projects can be hosted on university servers, on personal domains, or on public forges.
- Rules regarding intellectual property and storage of sensitive information apply no matter where code and data are hosted.

Glossary

change set
A group of changes to one or more files that are or will be added to a single commit in a version control repository.

commit
To record the current state of a set of files (a change set) in a version control repository. As a noun, the result of commiting, i.e. a recorded change set in a repository. If a commit contains changes to multiple files, all of the changes are recorded together.

conflict
A change made by one user of a version control system that is incompatible with changes made by other users. Helping users resolve conflicts is one of version control's major tasks.

HTTP
The Hypertext Transfer Protocol used for sharing web pages and other data on the World Wide Web.

infective license
A license, such as the GPL, that legally requires people who incorporate material under the infective license into their own work to also release under the same infective license (eg. under the GPL license).

merge
(a repository): To reconcile two sets of changes to a repository.

protocol
A set of rules that define how one computer communicates with another. Common protocols on the Internet include HTTP and SSH.

remote
(of a repository) A version control repository connected to another, in such way that both can be kept in sync exchanging commits.

repository
A storage area where a version control system stores the full history of commits of a project and information about who changed what, when.

resolve
To eliminate the conflicts between two or more incompatible changes to a file or set of files being managed by a version control system.

revision
A recorded change set of a version control repository. The same as a commit.

SSH
The Secure Shell protocol used for secure communication between computers.

timestamp
A record of when a particular event occurred.

version control
A tool for managing changes to a set of files. Each set of changes creates a new commit of the files; the version control system allows users to recover old commits reliably, and helps manage conflicting changes made by different users.
