

Git and GitHub

What is Version Control?

Version control is any of a variety of systems that have been developed to keep track of source code as it is being written. In some ways it can be thought of as similar to a track changes system in a word processing programme, but it is much more comprehensive. The goal of version control is to keep track of every change that has been made to a document or set of documents in a way that allows changes to be recorded, documented, and rolled back if necessary. As reproducible documents are written as code a version control system is a good way to keep track of the associated files containing the relevant documents and data. It also enables ways to make them available for public scrutiny.

A number of systems have been created to do version control for software and documents. One of the earliest modern systems was CVS which stood for Concurrent Versions System, this was the most popular system until Subversion (SVN) arrived, which then dominated until git arrived. Git has dominated ever since. There have also been more commercial systems that have been popular in enterprise settings such as BitKeeper and Perforce.

What is git?

[Git](#) is a version control system developed in 2005 by Linus Torvalds, the person who originally developed the Linux operating system. It was developed for work on the Linux operating system's kernel. It is designed to be a distributed system allowing thousands of coders to hold code on different computers and allow them to be merged together to form a coherent code base for a product or system. There can be many parallel branches of the code. The fact that it was designed for thousands of coders to use at once and yet maintain a coherent and functioning codebase makes it a good system for scientific collaboration. It works on all of the major operating systems and there should be minimal problems there. As it has origins in Linux development it probably works more easily in UNIX-style environments like Linux and MacOS, but, given the popularity of Windows, attention has been paid to making it work seamlessly with Windows products. This does lead to the issue of instructions being a little different depending on the computer you are using. Git is free and open source software.

Git operates by having a main repository that contains the code or documents, this typically sits on a main server. Users then clone copies of the full repository to their personal computers. Each of these copies can serve as a backup of the whole system. This main repository view is a centralized workflow and is one of the most common styles of development, but there are other more complicated workflows. Users can then make changes, commit them to their local repository and then push them to the main repository. Git allows multiple workers to work on software and ensures that they do not overwrite one another's work.

What is GitHub?

Git and GitHub are different entities and it is important that they are not confused although they are often used in parallel. GitHub is a cloud-based platform that serves as a host for git repositories, it also adds some additional features. In most use cases for Open Science GitHub serves as the main repository with individual personal computers serving as local repositories. Figure 1 demonstrates the typical workflow for most Open Science GitHub projects.

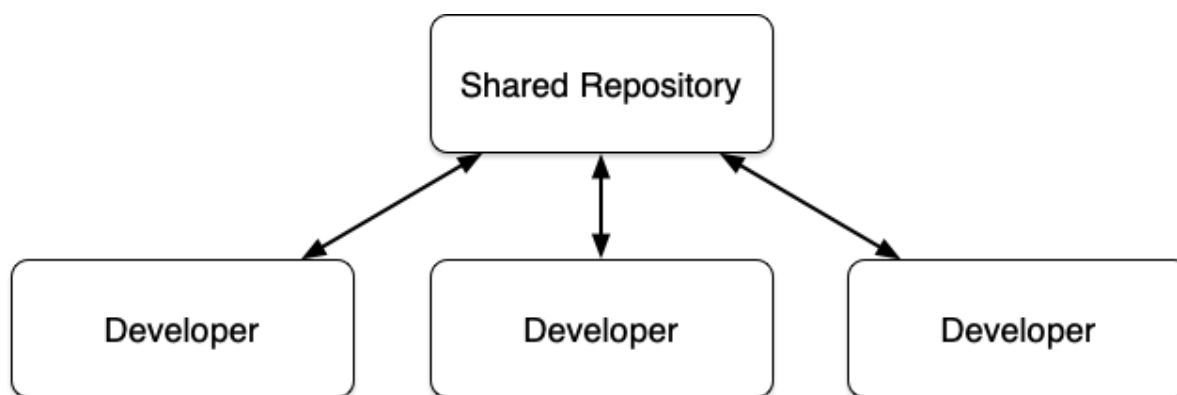


Figure 1: A Basic GitHub Workflow