Version Control with Git

(index.html)

Hosting

> ✴ Learning Objectives
>
> - Explain different options for hosting scientific work.

The second big question for groups that want to open up their work is where to host their code and data. One option is for the lab, the department, or the university to provide a server, manage accounts and backups, and so on. The main benefit of this is that it clarifies who owns what, which is particularly important if any of the material is sensitive (i.e., relates to experiments involving human subjects or may be used in a patent application). The main drawbacks are the cost of providing the service and its longevity: a scientist who has spent ten years collecting data would like to be sure that data will still be available ten years from now, but that's well beyond the lifespan of most of the grants that fund academic infrastructure.

Another option is to purchase a domain and pay an Internet service provider (ISP) to host it. This gives the individual or group more control, and sidesteps problems that can arise when moving from one institution to another, but requires more time and effort to set up than either the option above or the option below.

The third option is to use a public hosting service like GitHub (http://github.com), BitBucket (http://bitbucket.org), or SourceForge (http://sourceforge.net). Each of these services provides a web interface that enables people to create, view, and edit their code repositories. These services also provide communication and project management tools including issue tracking, wiki pages, email notifications, and code reviews. These services benefit from economies of scale and network effects: it's easier to run one large service well than to run many smaller services to the same standard. It's also easier for people to collaborate. Using a popular service can help connect your project with communities already using the same service.

As an example, Software Carpentry is on GitHub (https://github.com/swcarpentry/) where you can find the source for this page (https://github.com/swcarpentry/git-novice/blob/gh-pages/12-hosting.md). Anyone with a GitHub account can suggest changes to this text.

Using large, well-established services can also help you quickly take advantage of powerful tools. One such tool, continuous integration (CI), can automatically run software builds and tests whenever code is committed or pull requests are submitted. Direct integration of CI with an online hosting service means this information is present in any pull request, and helps maintain code integrity and quality standards. While CI is still available in self-hosted situations, there is much less setup and maintenance involved with using an online service. Furthermore, such tools are often provided free of charge to open source projects, and are also available for private repositories for a fee.

📌 Institutional Barriers

Sharing is the ideal for science, but many institutions place restrictions on sharing, for example to protect potentially patentable intellectual property. If you encounter such restrictions, it can be productive to inquire about the underlying motivations either to request an exception for a specific project or domain, or to push more broadly for institutional reform to support more open science.

✏️ Can My Work Be Public?

Find out whether you are allowed to host your work openly on a public forge. Can you do this unilaterally, or do you need permission from someone in your institution? If so, who?

Software Carpentry (http://software-carpentry.org)
Source (https://github.com/swcarpentry/git-novice)
Contact (mailto:admin@software-carpentry.org)   License (LICENSE.html)