

Archive your code: short-term and long-term solutions

By Sara Gonzales, Data Librarian

Many research projects rely heavily on code for data analysis. For some it's a relatively simple Python or R script. Others may be working with advanced algorithms, modeling, and homegrown databases. At a certain point in the life of a project, archiving the code may become necessary. Perhaps you want to freeze it in time, preventing further changes while still allowing others to view and download it. Or perhaps you wish to save a version of the code that can be released for data-sharing purposes as part of a funder's or a journal's requirements. To achieve these two goals, however, very different approaches and tools might be required.

Archiving in the short-term

If you're working with code in [GitHub](#), it's likely that you are already aware of GitHub's robust features for collaboration. But did you know that you can [archive](#) the code in your GitHub repositories? Archiving in GitHub makes all code, commits, pull requests, projects, issues, and wikis read-only. Others can still fork the repository, but no further changes can be made to the original. This can be a great solution for making completed project code available to colleagues in the short-term.

Though it may be tempting to use GitHub for permanent long-term storage after archiving your project, this is not recommended as a best practice. For an online repository to meet the requirements for long-term storage requested by most funders or journals requiring data-sharing, one very important criteria must be met: the repository must have a long-term plan for its preservation, and the preservation of the digital objects that it holds. In addition, once you have become an authenticated user of the repository, you should be granted access to it (and to your deposited materials) in perpetuity. GitHub's Terms of Service contain no explicit commitment to maintain the website into the distant future. And as the terms outline, "GitHub has the right to suspend or terminate your access to all or any part of the Website at any time, with or without cause, without or without notice, effective immediately. GitHub reserves the right to refuse service to anyone for any reason at any time." (GitHub Terms of Service, help.github.com/en/articles/github-terms-of-service)

Long-term solutions

If you have code, data, and/or files that currently reside in GitHub but which should be archived for the purposes of data sharing or long-term access, this can be done in a few simple steps. First, clone or download your repository onto secure storage with enough space to hold it, such as the FSM servers. Next, upload the contents of the repository to a trusted online resource designed for and dedicated to long-term storage of digital files. The repository you choose may be influenced by the type of materials you would like to deposit. If you have preprints to deposit, you might choose the popular repositories Arxiv or [Biorxiv](#) (pronounced "bio-archive). For data and code, you might choose generalist, data-friendly repositories like [Dryad](#), [Figshare](#), or [Zenodo](#) (whose maintainers have given some thought to long-term preservation and mention it in their policies; see [Zenodo's](#) and [Dryad's](#)). Additionally, Northwestern Medicine supports a robust institutional repository called [DigitalHub](#), which is a long-term storage solution for all types of research outputs, from papers and data to posters and presentations.

DigitalHub has several examples of datasets that have been uploaded to comply with a publisher's data sharing policies: see Marilyn Cornelis's [data from studies on the epidemiology of coffee](#), or Marta Perez's [pulmonary hypertension data in mice](#). DigitalHub also makes it easier for people to find, and depending on the license you've used, re-use your data for further work. The repository offers a variety of file and compression formats and can work with file sizes up to 2GB (or beyond, as needed). If you need assistance or have further questions about uploading datasets to DigitalHub, please contact DigitalHub@northwestern.edu.

Printed: Thursday, September 28, 2023 4:13 PM

Source: <https://vfmghslapps01.fsm.northwestern.edu/News/archive-your-code-short-term-and-long-term-solutions.pdf>