DHTech Education & Training Series: Reproducible Research

David Nelson, Jose Hernandez

Reviewed by: Julia Damerow, Fernanda Freire

DOI: 10.5281/zenodo.15870520

Contents

Reproducible Research

If we are going to take the time to build computational tools for humanistic research, we want others to be able to reproduce our work. This can either be for the purposes of verification of our results or for applying the tools to new datasets/research questions. This desire for reproducibility is motivated by the following questions:

- Will others be able to run my code?
- Will others know that my code works and does what I say it does?
- Will others be able to adapt, modify and contribute to my code?

Humanities-specific barriers

In the humanities, reproducible research is a moving target. While the social sciences have recently reckoned with a so-called “replication crisis,” the humanities are only beginning to think about how their research can be reproducible. As the humanities increasingly work with large data sets and computational tools that exceed what can be manually verified by a third-party observer, we need to agree upon best practices that will ensure our peers can trust the validity of our results. This problem is further aggrieved by the fact that most developers in the digital humanities are not software engineering professionals and may only be aware of some coding best practices but not all of the ones they need to increase the future sustainability of their code and long term digital projects.

Creating reproducibility

This reproducibility does not occur by happenstance and there are a couple of steadfast principles that will allow you to provide others with the capacity of reproducing your work, which are:

- Well-organized projects that others can understand (covered in our Project Organization lesson)
- Using a version control system to facilitate collaboration on the project (covered in our Git & GitHub lesson)
- Recording your environment so others can run your code on their machine (covered in our Environments lesson)

These three are not by any means all-encompassing but they provide a signifiant starting point for those in the humanities that are beginning to learn software development practices or are moving on to larger digital humanities projects. Each one of these, as stated above has their own lesson, but there is an underlying lesson that ties them together which is **documentation**.

Documentation as law

Documentation is the practice of recording what your code or tool does, how it should be used, and any necessary context so that others (including your future self) can understand, run, and extend it. Therefore, without correct and pervasive documenation all of our other reproducibility efforts fall short.

Documentation can be found at all levels, and it is important that we know what it looks like so that when you are looking at tools, you can find it—and when you are building, you can make sure that it is prioritized.

The most granular documentation consists of comments, which are short notes written alongside the code to explain specific lines or logic. Each programming language has its own way to add comments in the code. In Python, for example, a comment looks like this (starting with #):

```
print("Hello") # This line will print the string
```

For bigger projects with many files it is good practice to provide entire documentation files. These files are typically called README or `README.md` and part of a the repository. One step further than README files, are dedicated docs pages that contain tutorials, descriptions, and everything in between made available via a webpage. No matter the scale, documentation creates the foundation for reproducibility and readability for all levels of users from experienced programmers to first-time coders!

Resources

- Rik Peels, “Replicability and replication in the humanities.” *Research Integrity and Peer Review* 4 (2019). https://doi.org/10.1126/science.aac4716.
- Joseph Flanagan, “Reproducible research: Strategies, tools, and workflows.” *Studies in Variation, Contacts and Change in English*, eds. Turo Hiltunen, Joe McVeigh, Tanja Säily (Helsinki: Research Unit for Variation, Contacts and Change in English, 2017). https://varieng.helsinki.fi/series/volumes/19/flanagan/