

bit: a git-based tool for the management of code and data

Jorge Boucas¹

¹ Bioinformatics, Max Planck Institute for Biology of Ageing, Cologne, 50931, Germany

DOI: [10.21105/joss.00534](https://doi.org/10.21105/joss.00534)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 15 January 2018

Published: 26 January 2018

Licence

Authors of JOSS papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

bit, [b]ermuda [i]nformation [t]riangle, is a git-based tool for the management of code and data.

In the increasing world of data, the mingles of data, code, and project management have become crucial for a successful and agile processing of data into meaning.

While several proficient tools have emerged along the last years for each of the individual tasks, a unified tool is still in the need for many. Avoiding the costs of developing and deploying an unified tool, using existing tools in a microservices architecture is at the reach of any, allows refactoring and is infrastructure independent.

bit uses git for code versioning and for example ownCloud for storing and exchanging data. It avoids the costs of data versioning by logging in git the changes made on the data storage platform (eg. ownCloud). With commit messages being shared between data exchange logs (kept on the project's wiki) and code pushes, data and respective code are permanently linked.

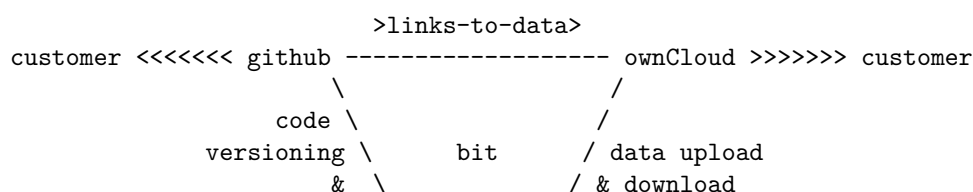
Data can be transferred over http requests using ownCloud's API and avoiding therefore the risks of mounted drives and the need for ftp access. Every time data is uploaded to ownCloud, the upload is registered on the respective wiki on github together with an hyperlink to the respective folder. Project metadata and data can like this be kept in different instances.

Using rsync, it allows seamless multiple rsync calls for efficient synchronisations of projects between different HPCs or local machines.

It is primarily developed to support multiple bioinformaticians/analysts working on multiple shared research projects on one or more HPC clusters but it can be easily used on local/cloud/non-HPC machines without changes to the code.

bit promotes increased reproducibility of scientific results by making use of commit messages to fuse code and data.

The open source bit package is freely available at <https://github.com/mpg-age-bioinformatics/bit>.



```
data upload logs \      /  
                  \      /      rsync  
                  HPC/local -----HPC2/local2
```