Code management for a reproducible science with Git, GitFlow, GitHub and Zenodo

Nicolas Barrier
nicolas.barrier@ird.fr

Criscely Luján
criscely.lujan@ird.fr

1IRD, UMR-MARBEC
2Université Paris-Sud

2021-06-23
To reproduce an ecosystem modelling study, you will need:

- The model parameters (time-step, mortality rates)
- The forcing files (temperature, oxygen, plankton, etc.)
- **The numerical model** (source or executable)
- **The analysis scripts** (R, Python, Matlab)
- The external libraries used (graphical libraries, mathematical ones, etc.)

Therefore, a proper management of codes (model + scripts) is necessary.

**This can be achieved by version control**
Version control

- Tracks changes (commits) over time
- Saves snapshots of a project (tags)
- Creates derivates of a project (branches)
- Facilitates collaboration among multiple users
Version control softwares

<table>
<thead>
<tr>
<th>Category</th>
<th>Version control software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Local only</strong></td>
<td>RCS (1982) · SCCS (1972)</td>
</tr>
<tr>
<td><strong>Proprietary</strong></td>
<td>PVCS (1985) · QVCS (1991)</td>
</tr>
<tr>
<td><strong>Concepts</strong></td>
<td>Baseline · Branch · Changeset · Commit · Data comparison · Delta compression · Fork (Gated commit) · Interleaved deltas · Merge · Repository · Tag · Trunk</td>
</tr>
</tbody>
</table>

Git

- **Popular and successful**
  - Active development + multiple extensions

- **Tracks any type of files**
  - Works best with ASCII files (.html, .tex, .R, .c, .f90)
  - Large binary files with Git-LFS (.nc, .Rdata, .mat, .csv)

- **Branching**
  - Light branch management
  - Smarter merges
  - Easy code management with Git-Flow
Several web-based hosting services are available:

- **GitHub**, Gitlab
- Institutional: GitLab Bioinfo, GitLab UM, SourceSup, Ifremer Forge

**Why use GitHub?**

- Popular (largest host of source code in the world)
- Facilitates worldwide collaboration
- Free (and improved accounts for education)
- Works with Git-LFS
- Continuous integration ([GitHub Actions](https://github.com))
- Association with **Zenodo**
Zenodo = a *catch-all repository for EC funded research*, based on FAIR principles (*Findable, Accessible, Interoperable, Reusable*).

**Automatic generation of DOIs when a new release is drafted.**
Good practices

Main branch (main/master)
- Not updated too often
- Each change is identified (tag and release)

Development branch (develop).
- Independent developments in separated branches

Bug fix:
- In master with a new version number
- In develop

Seems complicated? Not at all, this is GitFlow!
GitFlow

**Git extension** to manipulate these branches.

```plaintext
git flow TYPE start NAME
git flow TYPE publish NAME
git flow TYPE finish NAME
TYPE=branch type (feature, release, hotfix)
NAME=branch name (feature name, version number)
```
Example: Ichthyop

For an example, visit https://github.com/ichthyop/ichthyop

ICHTHYOP

License: GPL-3.0
DOI: 10.5281/zenodo.4707436
Release: v3.3.8
Issues: 8 open

Version 3.3.8
Apr 21, 2021
10.5281/zenodo.4707436

Version 3.3.7
Apr 19, 2021
10.5281/zenodo.4700363

Version 3.3.6
Nov 4, 2020
10.5281/zenodo.4244484

Version 3.3.5
Nov 4, 2020
10.5281/zenodo.4243817

Version 3.3.4
Nov 4, 2020
10.5281/zenodo.4243814

Cite all versions? You can cite all versions by using the DOI
10.5281/zenodo.4243813. This DOI represents all versions, and will always resolve to the latest one. Read more.

Nicolas Barrier & Criscely Luján
Code management for a reproducible science