3 Discussion Points

Arni Magnusson

SPC Git/GitHub Workshop
Noumea, 13 April 2022

Overview

Good habits *commit early and often, keep commits small, commit messages, keep a light repo, pull before making changes*

Collaboration *multiple contributors, merge conflicts, plumbing analogy, communication, forks, branches, pull requests*

Repo management *separate folders vs. repos vs. forks vs. branches, main git folder*

SPC projects *github vs. penguin vs. onedrive vs. teams vs. c-drive vs. vm, enhanced team collaboration, open science, improved backup*

Pacific Community
Communauté du Pacifique

Good Habits

Good Habits

Pacific Community
Communauté du Pacifique

Commit early and often

Create repo and first commits as soon as you begin working on something new

Keep each commit small *so it can be described with a short commit message*

Write descriptive commit messages

Relatively short, around 60 chars

Describing purpose or specific things that were changed

Good Habits (cont)

Pacific Community
Communauté du Pacifique

Keep a light repo

Usually under 1 GB

Avoid including a large file that will change

General idea with repos is not to include output files

Release assets are a good approach to attach heavy files while keeping the repo light

Pull before making changes

To make sure you're adding to the current state of the repo

Avoiding merge conflicts

Pacific Community
Communauté du Pacifique

Collaboration

Collaboration

Pacific Community
Communauté du Pacifique

Multiple contributors

- Merge conflicts
- Plumbing analogy
- Communication

Pull requests

- Forks, branches
- Submitting and handling pull requests

Pacific Community
Communauté du Pacifique

Repo Management

Repo Management

Pacific Community
Communauté du Pacifique

Related workflows can be organized in

- Separate files
- Separate folders
- Separate repos

⇒ *Makes things easier to understand, extend, and maintain*

⇒ *Also reduces the risk of merge conflicts*

Repo Management (cont)

Main git folder

Consider storing all Git repos in one place

`c:/git` *in Windows*

`~/git` *in Linux*

For example, I store this workshop repo

`https://github.com/PacificCommunity/ofp-sam-git-workshop`

on my computer as

`~/git/PacificCommunity/ofp-sam/git-workshop`

Benefits

Easy to find repos

Repos are not repeated or forgotten in multiple places

Related repos are next to each other

The main Git folder can be excluded from regular backups

Pacific Community
Communauté du Pacifique

SPC Projects

SPC Projects

Where to work on a project

GitHub repo	
Shared drive	*penguin*
OneDrive	
Teams folder	
C-drive	*desktop*
Personal VM	*wsl*

No rules, but worth thinking about the pros and cons of each options

Git/GitHub skills are one factor to consider

GitHub repos are the default standard for open science and reproducible analyses

Practical for regional and international collaboration

Easy to migrate from GitHub to other platforms if the need arises

GitHub vs. Shared drive

	GitHub	Shared drive
Backups	Every change is saved as snapshot	Some
Collaboration	Pull requests, view contributions, issues	Difficult to see what others do
Open science	Can share with the world	Local network only
Large files	Each repo $<$ 1 GB plus assets	No limits
Expertise	Requires Git/GitHub skills	Easy to copy files
Used for	Analyses, software, data hub, information	Same
Expectation	Should (ideally) run on any computer	Can be anything
Style	Minimalistic, organized	Kitchen sink

Input and Result Files

Thursday, 02 September 2010 08:40

SPC routinely assesses bigeye, yellowfin, and skipjack tuna in the Western and Central Pacific, and South Pacific albacore tuna. New assessments are carried out each year and presented to the WCPFC Scientific Committee in August. The files associated with these new assessments will be placed on this website as soon as possible after 30 June of each year.

	Albacore	Bigeye	Skipjack	Yellowfin	Sharks	Striped Marlin	Swordfish
2006						View	
2007				View			
2008	View	View	View				
2009	View	View		View			
2010		View	View				
2011	View	View	View	View			
2012	View				View	View	
2013					View		View
2014		View	View	View	View		
2015	View						
2016			View				
2017		View		View			View
2018	View	View					
2019			View		View	View	
2020		View		View			

WCPO Assessment
Research
Workshops & Training
National Advice & Reports
Regional Advice

When to Use GitHub

Could we put everything on GitHub? *that we have traditionally put on the shared drive*

Yes

The webpage of zipped assessments could also be on GitHub

Each assessment could be organized in 4–5 repos:

1. skj-2022-stepwise
2. skj-2022-diagnostic
3. skj-2022-grid
4. skj-2022-retro
5. skj-2022-plots

The *data preparation* could also be organized in repos

When to Use GitHub

Should we put everything on GitHub? *that we have traditionally put on the shared drive*

+ The main benefits are enhanced team collaboration, open science, and improved backup
– The main cost is the required expertise and the need to run Git commands in every step

  The costs can be reduced by Git/GitHub training and efficient Git configuration

⇒ For OFP data preparation and stock assessments, using GitHub repositories should probably not be compulsory but **encouraged**

Summary

Pacific Community
Communauté du Pacifique

Good habits *commit early and often, keep commits small, commit messages, keep a light repo, pull before making changes*

Collaboration *multiple contributors, merge conflicts, plumbing analogy, communication, forks, branches, pull requests*

Repo management *separate folders vs. repos vs. forks vs. branches, main git folder*

SPC projects *github vs. penguin vs. onedrive vs. teams vs. c-drive vs. vm, enhanced team collaboration, open science, improved backup*