Fundamentals for Collaborating on Research Projects Using GitHub

Rob Crystal-Ornelas

Madison Burrus, Shreyas Cholia, Joan Damerow, Valerie Hendrix, Emily Robles, Helen Weierbach, and Charuleka Varadharajan

ESS-DIVE BERKELEY LAB NERSC NCEAS DataONE U.S. DEPARTMENT OF ENERGY Office of Science

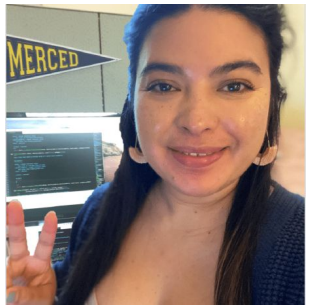
AGU 2021

Special thanks to GitHub collaborators

Rob Crystal-Ornelas
@rob_c_ornelas

Emma Hudgins
@emmajhudgins

Pedro Henrique P. Braga
@pedrohp_braga



Luna Sánchez Reyes
@LunaSare

Katherine Hébert
@hebert_kat

Brandon Edwards
brandonedwards.ca

Ciera Martinez
@cierareports

SORTEE
Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology

A history of GitHub in one slide

- First launched in April 2008
- Git (a version control system) + GitHub (a website) = collaboration on **computer code**
- As of 2021, over 56 million registered users

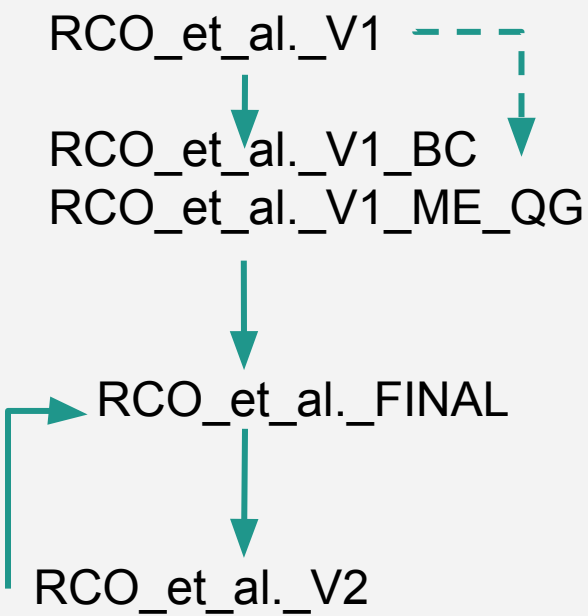
Especially within the science research community, GitHub has so many uses beyond collaborating on computer code

https://github.com/

GitHub's 'octocat'

GitHub is a place where **scientists** can collaborate on many things (docs, websites, etc.)

The ‘edit, save, attach’ method is inefficient

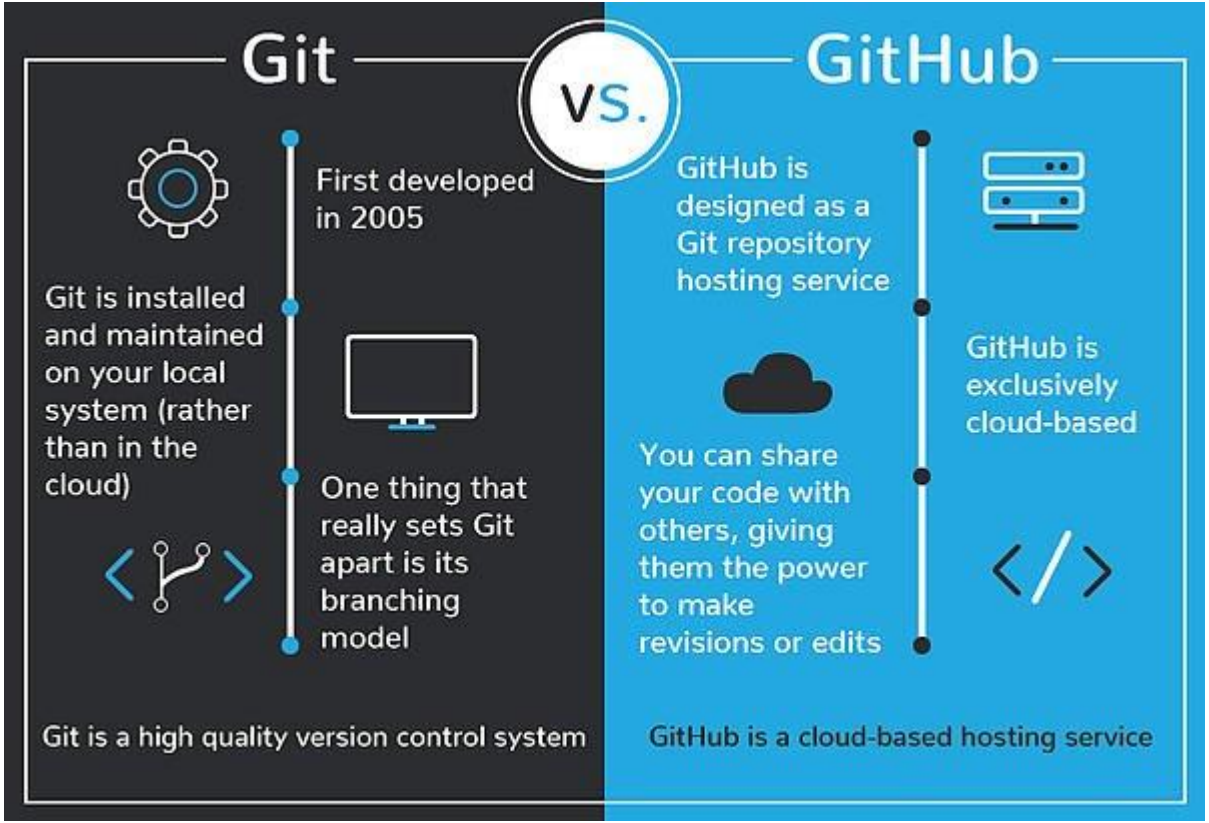


GitHub

Version control

“Like track changes [...] but more rigorous, powerful, and **scaled up to multiple files**." (Bryan, 2018)

What are Git and GitHub?

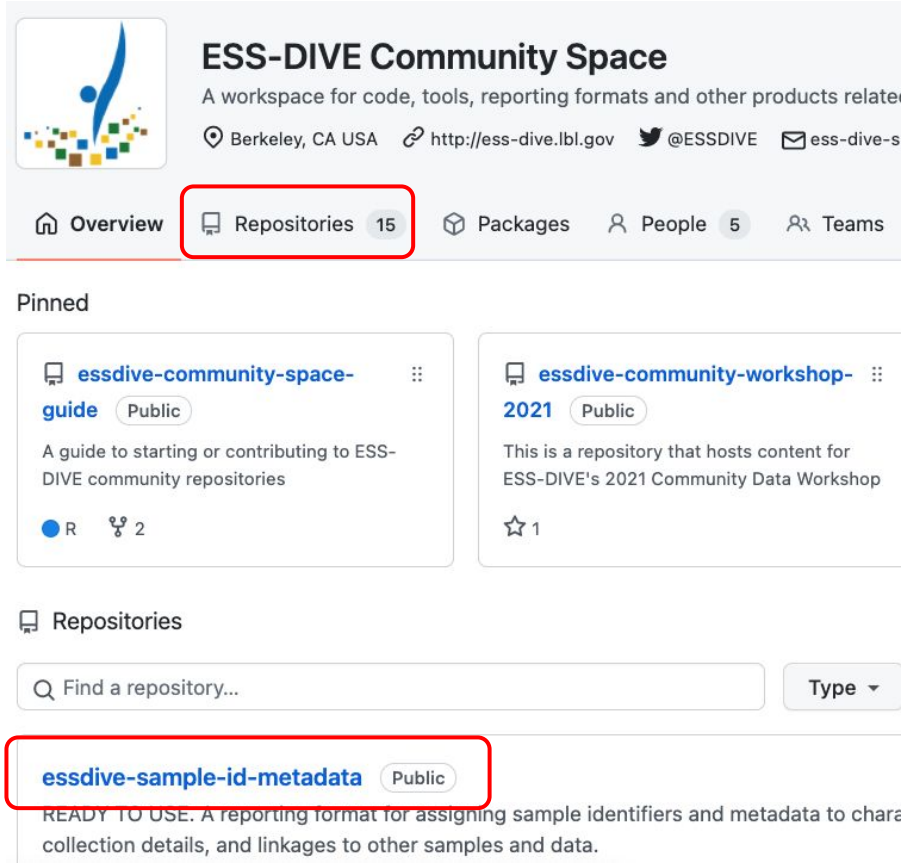


devmountain.com

Comparable to “Time Machine” on Mac, or “Version history” within Dropbox and OneDrive, with the added capacity to **“commit”** (provide an explanation of each change) and “**branch**”.

“**Branching**” helps avoid having to manually combine independently-updated versions (e.g., comparing 2 word documents independently updated by 2 people).

The GitHub user interface



GitHub repositories are the main ways we organize content on GitHub. Think of these as “folders” where you can keep and collaborate on your documents and code.

ESS-DIVE’s community space on GitHub: https://github.com/ess-dive-community

Questions?

Some caveats

GitHub is far from a perfect system in many ways

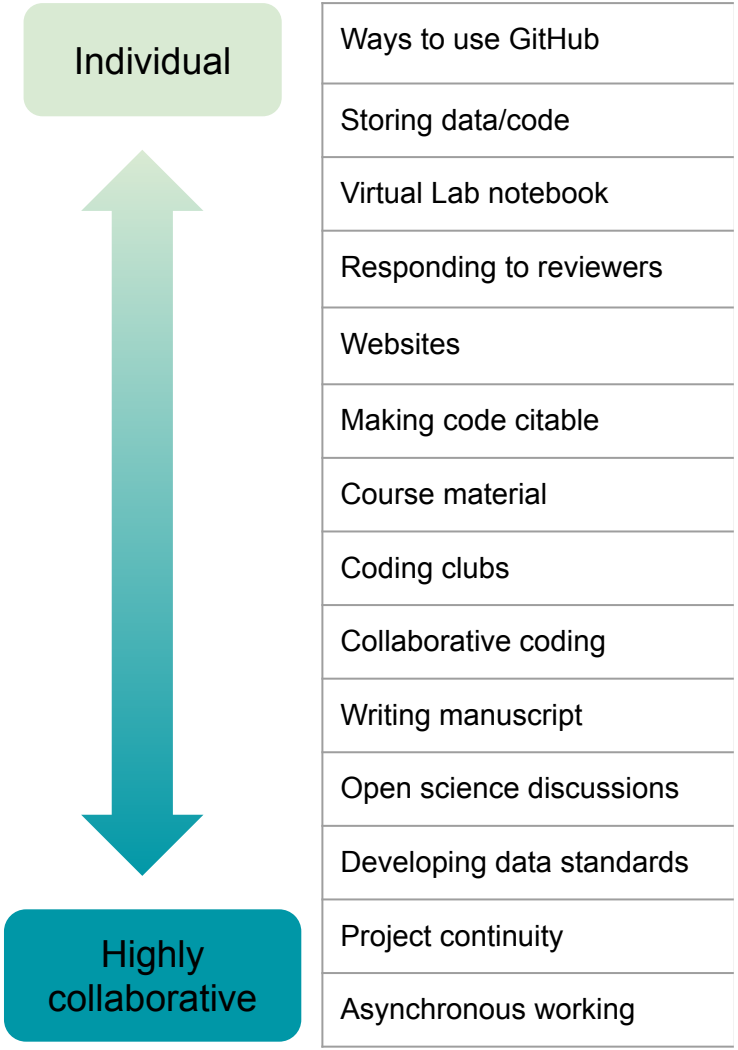
Acquired by microsoft in 2018

Putting code/data on GitHub? Be aware it's not a **long-term repository**

What else can GitHub do?

Many other potential uses:

- **Cloud storage** service for some types of files
- **Forking** allows people to use others' repositories as **templates** for their own work
- Provides a **hosting** service for web content
- Allows you to freeze your work at a given moment in time as a **release** which can be linked to a DOI (Required by many journals/funders)
- Provides **integration** with other tools (e.g. OSF, zenodo repository)



ESS-DIVE

Many more ways to use GitHub for science.

We can discuss any of these topics further during discussion!

Questions?

Ready to start working with GitHub?

1. Create your *free* github account: https://github.com/

2. Become familiar with a few GitHub basics

    A. Quick intro to github terms and features.

3. Set up your own **GitHub profile**.

    A. instructions from GitHub

1. Create a repository with the same name as your user name
2. Make sure the repository is public
3. In that new repository, add a README.md file and create repo.
4. Introduce yourself (using markdown language)!

Rob Crystal-Ornelas

robcrystalornelas

PhD in Ecology & Evolution / Postdoc at Lawrence Berkeley National Lab

Edit profile

9 followers · 11 following

robcrystalornelas / README.md

Hi there 👋

About me

My name is Rob Crystal-Ornelas and I'm currently a postdoc at Lawrence Berkeley National Lab. You can find me on or .

- 🌱I'm currently working on community data standards and meta-analysis of environmental data.
- 😄Pronouns: he/him/his

Collaboration

- 💻I'm working on a digital book about how to do meta-analysis in R, and always looking for collaborators on that project.
- ⌨️I'm collaborating on a manuscript summarizing how to use GitHub in Ecology and Evolution research. Follow along here

My skills

Code Python | Code R | Code LaTeX | Code Markdown | Code jupyter | Code git | Code GitHub

1. Create a repository with the same name as your user name

Overview | Repositories 17 | Projects | Packages | Stars 18

Find a repository... | Type | Language | Sort | New

meta-analysis_of_ecological_data Public

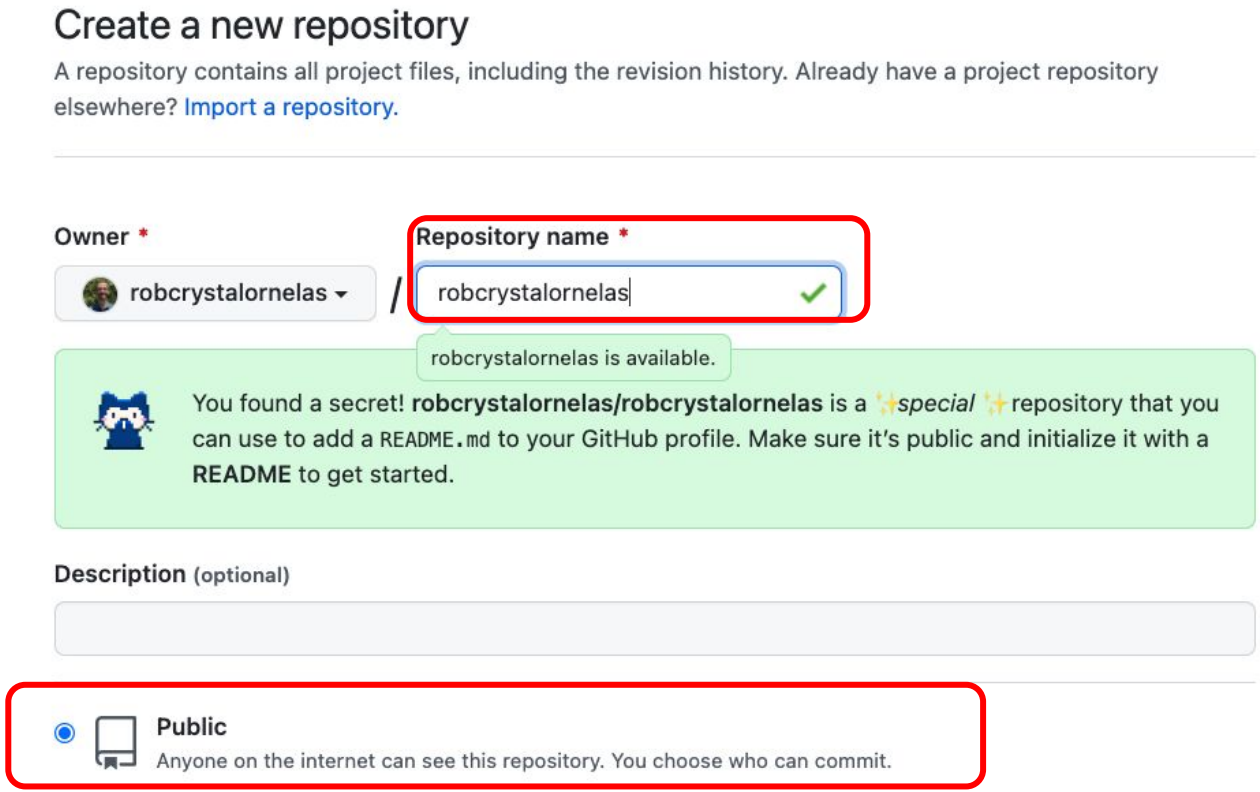
This is a repository for draft versions of a book called Meta-Analysis of Ecological Data in R (MaEDR)

Starred

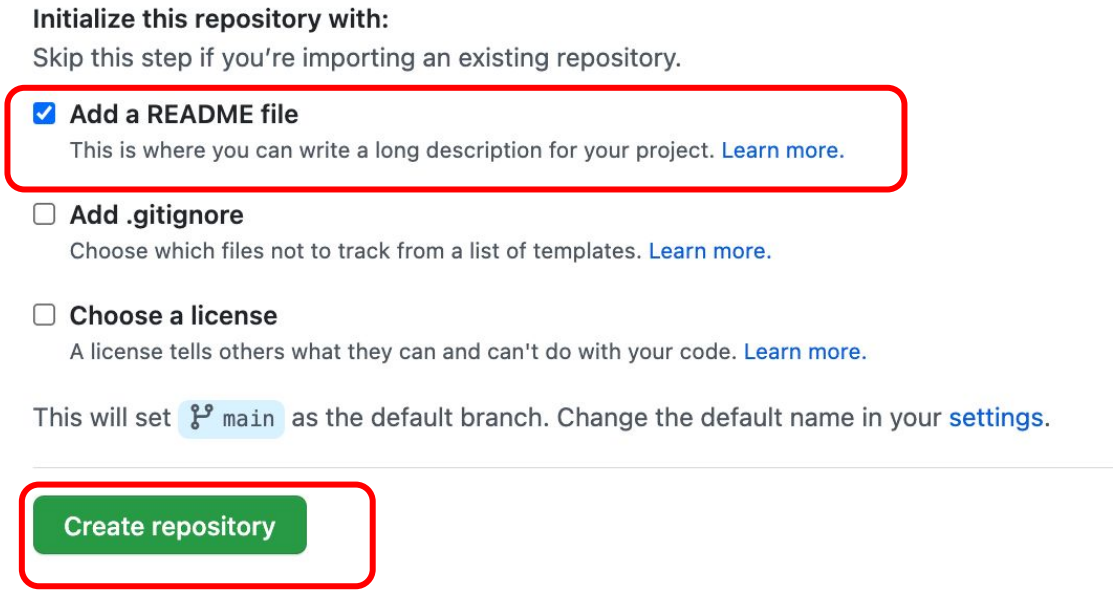
meta-analysis | ecology | systematic-review | meta-analyses | environmental-science

TeX | 4 | 2 | Creative Commons Attribution 4.0 International | Updated on Oct 28

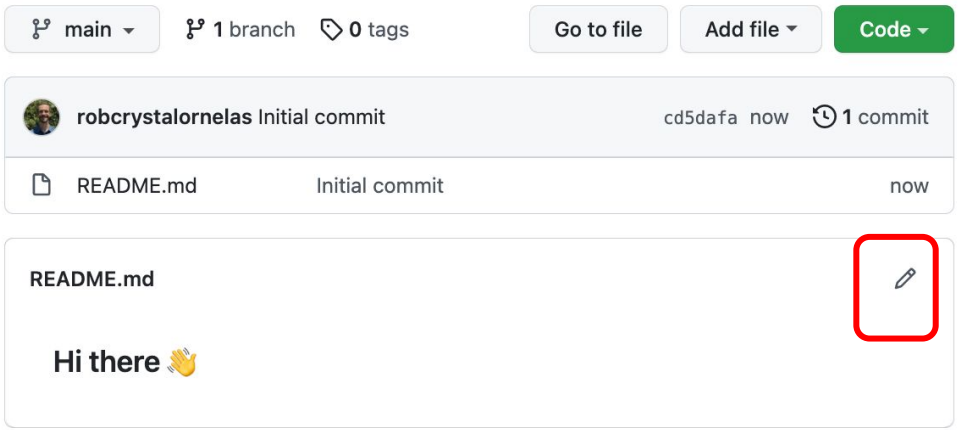
1. Create a repository with the same name as your user name
2. Make the repository public

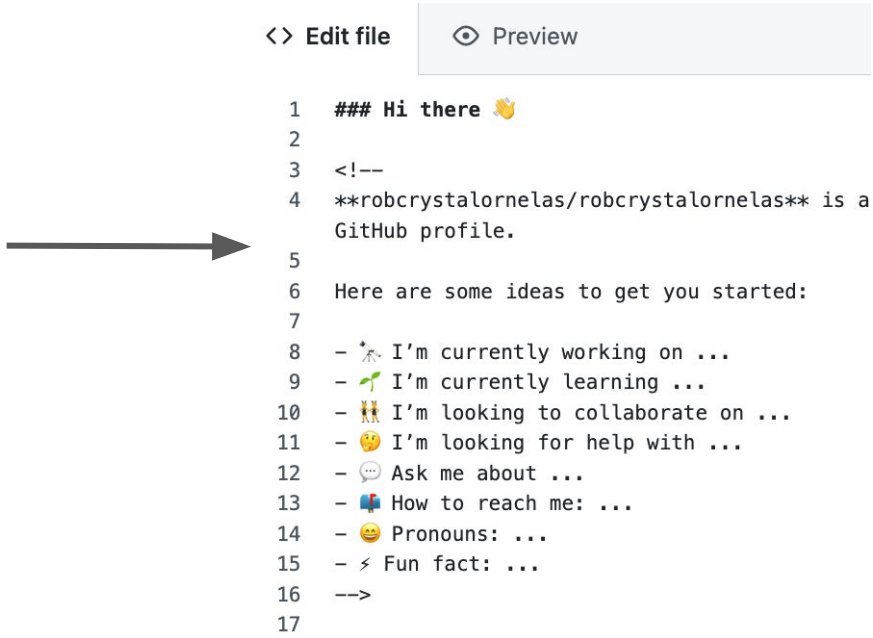


3. Add a README.md file and publish your repo



4. Open the README.md file, and introduce yourself!





Ways to learn more

GitHub tutorials	https://lab.github.com/
Markdown basics	https://commonmark.org/help/
Getting to know Git	https://github.com/dlab-berkeley/Bash-Git
Longer read: Git, GitHub, and R	https://happygitwithr.com/

Next steps

Check out: Paper on using GitHub for data standards
https://doi.org/10.1029/2021EA001797

Follow along: We're writing a paper (on GitHub :)) sharing ways of using GitHub in Enviro Sci / Ecology research:
https://github.com/SORTEE-Github-Hackathon/manuscript

Have fun on GitHub!

rcrystalornelas@lbl.gov
Twitter: @rob_c_ornelas
GitHub: robcrystalornelas

Thank you!