

# MAKING OPEN SCIENCE WORK FOR YOU: TOOLS FOR REPRODUCIBLE NEUROSCIENCE

STEPHANIE NOBLE, PHD  
NEUROPRISM LAB  
DEPARTMENT OF PSYCHOLOGY  
DEPARTMENT OF BIOENGINEERING  
CENTER FOR COGNITIVE & BRAIN SCIENCE  
NORTHEASTERN UNIVERSITY

# WHY DO OPEN SCIENCE?



# WHY DO OPEN SCIENCE?

ADVANCE SCIENCE



# WHY DO OPEN SCIENCE?

## ADVANCE SCIENCE

### Open data

Transparency through reproducible analysis, better outreach through exchange of data with partners like Encyclopedia of Life, and accelerated discovery through data reuse



### Open resources

Better training in Open Science methods and increasing access to resources for data collection and database construction



### Open Science Principles



### Open source

Reproducible analyses, accelerated synthesis through data and tool sharing, and improvement via shared data cleaning and checking



### Open access

Faster knowledge transfer as published works become more easily shareable



### Open methods

Standards development for collection protocols and metadata, and easier interpretation and decision-making scrutiny



### Open peer review

Greater scientific rigour through increased scrutiny of data and methods

adapted from Gallagher et al., 2020; Nature Ecology & Evolution

# WHY DO OPEN SCIENCE?

ADVANCE SCIENCE

## Open data

Transparency through reproducible analysis, better outreach through exchange of data with partners like Encyclopedia of Life, and accelerated discovery through data reuse



## Open resources

Better training in Open Science methods and increasing access to resources for data collection and database construction



Open  
Science  
Principles



## Open access

Faster knowledge transfer as published works become more easily shareable



## Open methods

Standards development for collection protocols and metadata, and easier interpretation and decision-making scrutiny

## Open source

Reproducible analyses, accelerated synthesis through data and tool sharing, and improvement via shared data cleaning and checking



## Open peer review

Greater scientific rigour through increased scrutiny of data and methods

ADVANCE YOUR CAREER

adapted from Gallagher et al., 2020; Nature Ecology & Evolution

# TOOLS FOR REPRODUCIBLE NEUROSCIENCE

- Quick & dirty onboarding for...

# TOOLS FOR REPRODUCIBLE NEUROSCIENCE

- Quick & dirty onboarding for...
  - *Open code*
    - Focus: Version control
    - See also: Code testing, Reproducible workflows

# TOOLS FOR REPRODUCIBLE NEUROSCIENCE

- Quick & dirty onboarding for...
  - *Open code*
    - Focus: Version control
    - See also: Code testing, Reproducible workflows
  - *Open data*
    - Focus: Public dataset sharing and reuse
    - See also: Data Management

# TOOLS FOR REPRODUCIBLE NEUROSCIENCE

- Quick & dirty onboarding for...
  - *Open code*
    - Focus: Version control
    - See also: Code testing, Reproducible workflows
  - *Open data*
    - Focus: Public dataset sharing and reuse
    - See also: Data Management
  - *Other topics in reproducible neuroscience*
    - Open publishing, Open community

# TOOLS FOR REPRODUCIBLE NEUROSCIENCE

- Quick & dirty onboarding for...
  - *Open code*
    - Focus: Version control
    - See also: Code testing, Reproducible workflows
  - *Open data*
    - Focus: Public dataset sharing and reuse
    - See also: Data Management
  - *Other topics in reproducible neuroscience*
    - Open publishing, Open community
- Guiding stars
  - Why bother?
  - Quick Start (Fastest + easiest way to get started)
  - Common bottlenecks

# OPEN CODE: VERSION CONTROL

# OPEN CODE: VERSION CONTROL

track + manage changes to code

Why bother?



# OPEN CODE: VERSION CONTROL

track + manage changes to code

## Why bother?

- Save time, dissemination, recognition, & collaboration



# OPEN CODE: VERSION CONTROL

track + manage changes to code

## Why bother?

- Save time, dissemination, recognition, & collaboration
- Most useful if your project involves...
  - Multiple “stages” of development
  - Multiple files
  - Code intended to be used by others (or future you, or you across computers, etc.)
  - Contributing to an existing project
  - Also to learn transferable skills: critical for many jobs

The screenshot shows a Reddit post on a dark-themed browser window. The title of the post is "Should I use GitHub?". The post was made 2 years ago by a user named Weary\_Mango\_113. The content of the post discusses the challenges of managing a codebase solo over 9 months, mentioning that it has become a mess and difficult to clean up. It asks if GitHub is suitable for such a purpose instead of renaming files on localhost. The post has 170 upvotes and 51 comments. A note at the bottom indicates it is an archived post.

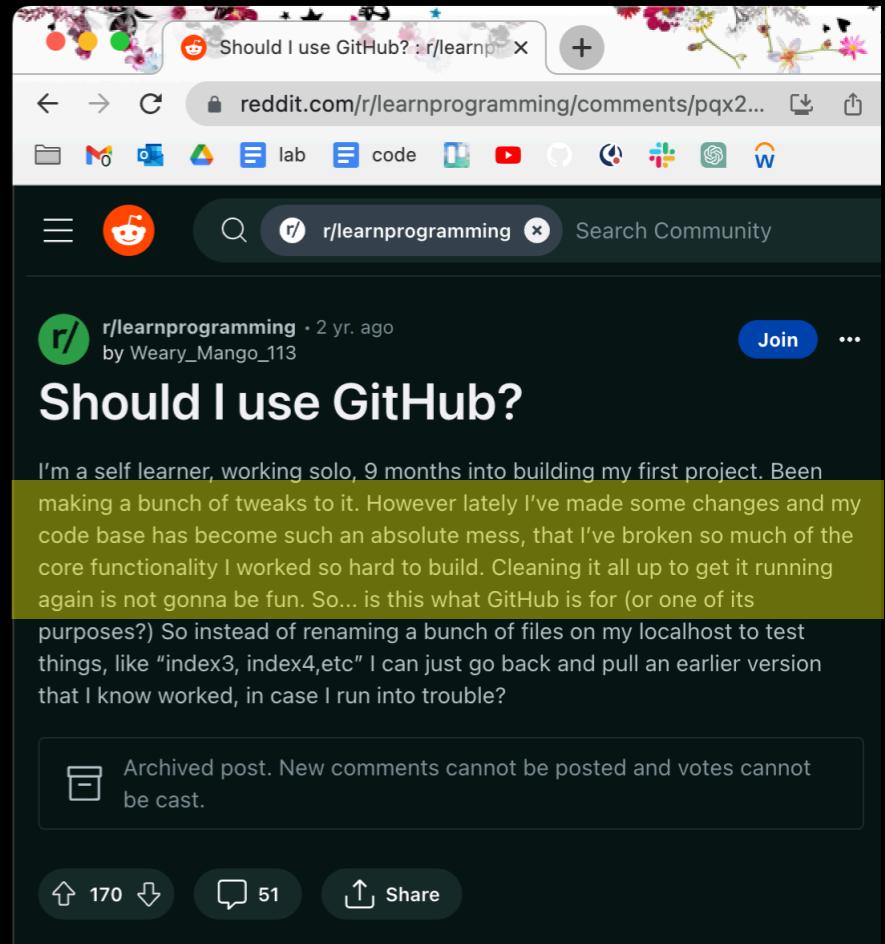


# OPEN CODE: VERSION CONTROL

track + manage changes to code

## Why bother?

- Save time, dissemination, recognition, & collaboration
- Most useful if your project involves...
  - Multiple “stages” of development
  - Multiple files
  - Code intended to be used by others (or future you, or you across computers, etc.)
  - Contributing to an existing project
  - Also to learn transferable skills: critical for many jobs
- Less critical if your project involves...
  - A single file
  - A one-off quick & dirty test (but you’d be surprised...)

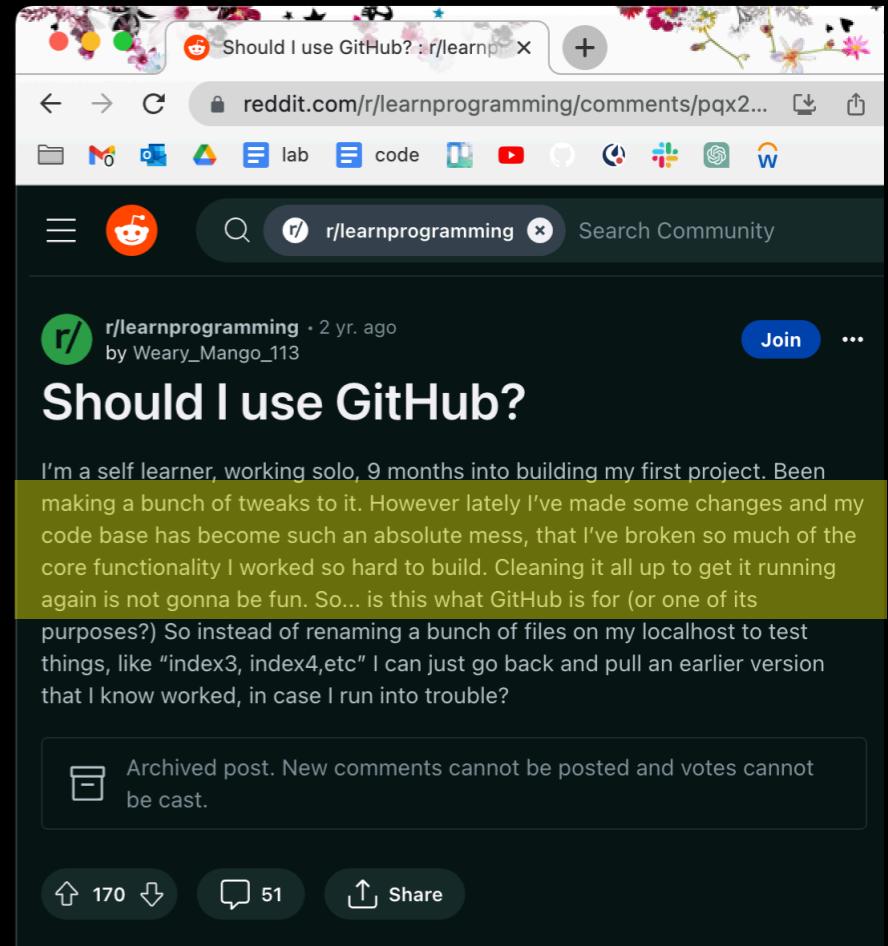


# OPEN CODE: VERSION CONTROL

track + manage changes to code

## Why bother?

- Save time, dissemination, recognition, & collaboration
- Most useful if your project involves...
  - Multiple “stages” of development
  - Multiple files
  - Code intended to be used by others (or future you, or you across computers, etc.)
  - Contributing to an existing project
  - Also to learn transferable skills: critical for many jobs
- Less critical if your project involves...
  - A single file
  - A one-off quick & dirty test (but you’d be surprised...)
- TLDR: Put in some time at the beginning to save lots of time / effort / heartache later



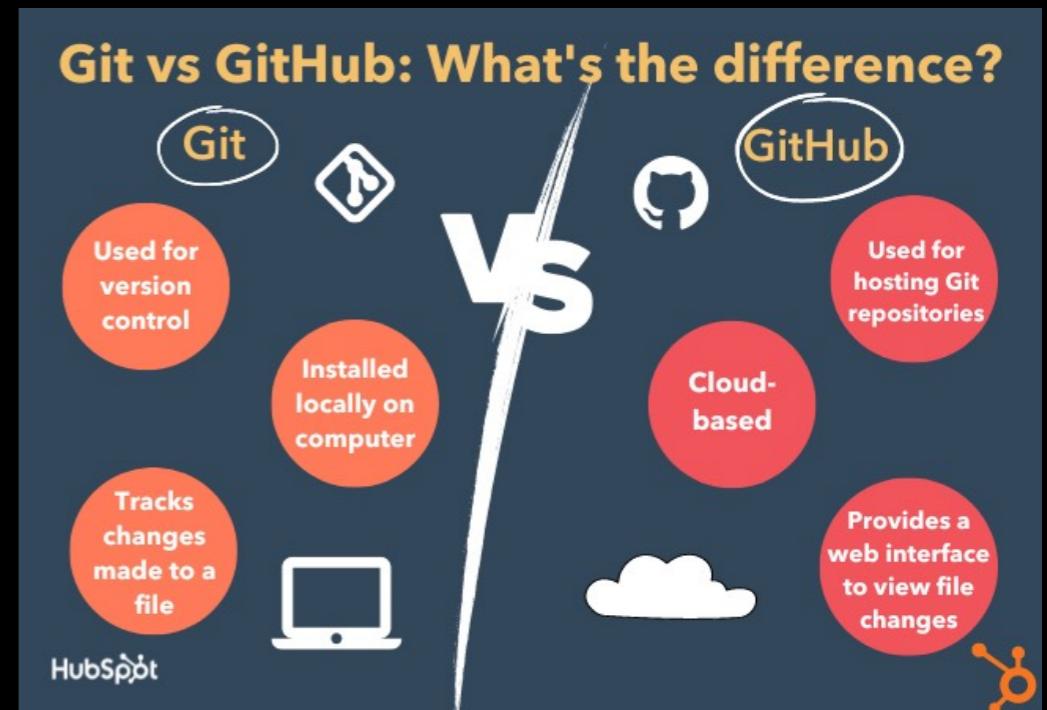
# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Quick Start

# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Quick Start

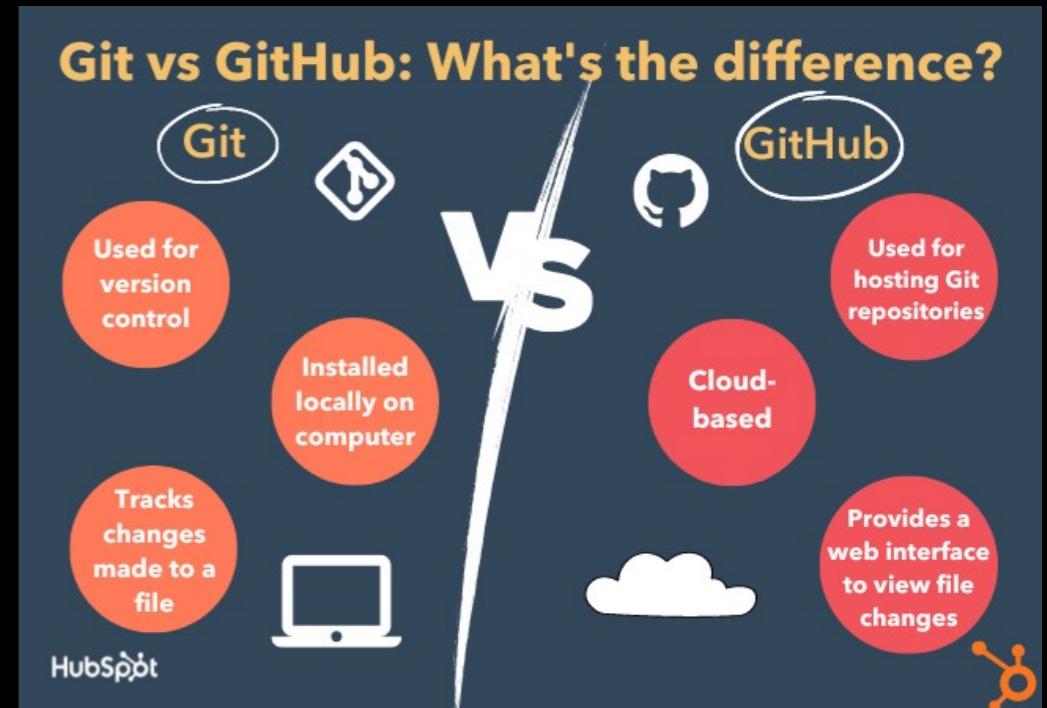
- First, some terminology:
  - **Git** ~ commands
    - lots of Git-specific terminology
  - **GitHub** ~ online storage



# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Quick Start

- First, some terminology:
  - **Git** ~ commands
    - lots of Git-specific terminology
  - **GitHub** ~ online storage
- Demo platform: **GitHub website**
  - Great for quick & dirty but has limited functionality
  - Other options (easy->complex): GitHub Desktop, IDE (editor) + extensions, command line, etc.



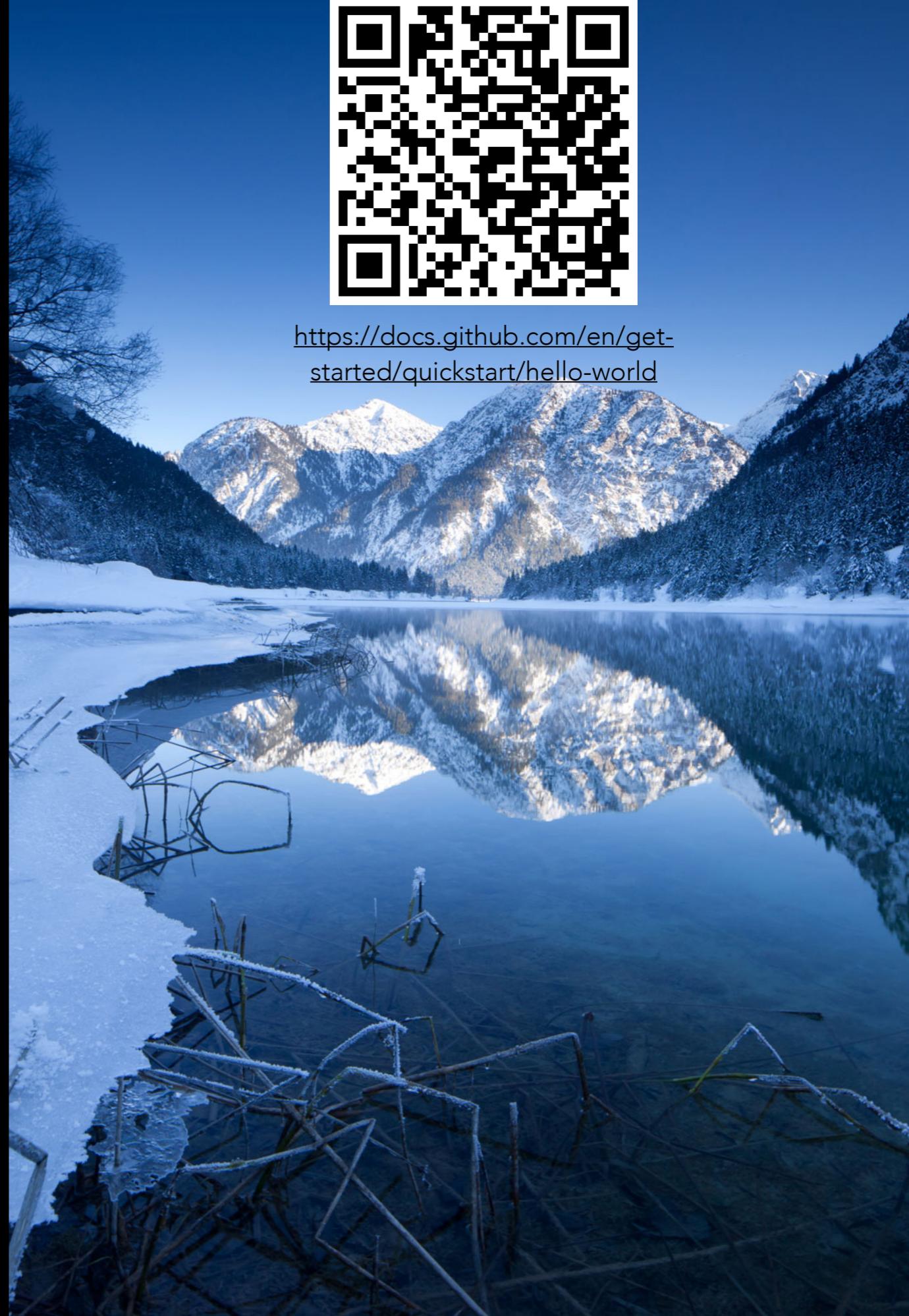
# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

- Step 1. Make a GitHub account
- Step 2. Make a new repository
- Step 3. Create a branch
- Step 4. Make + commit changes



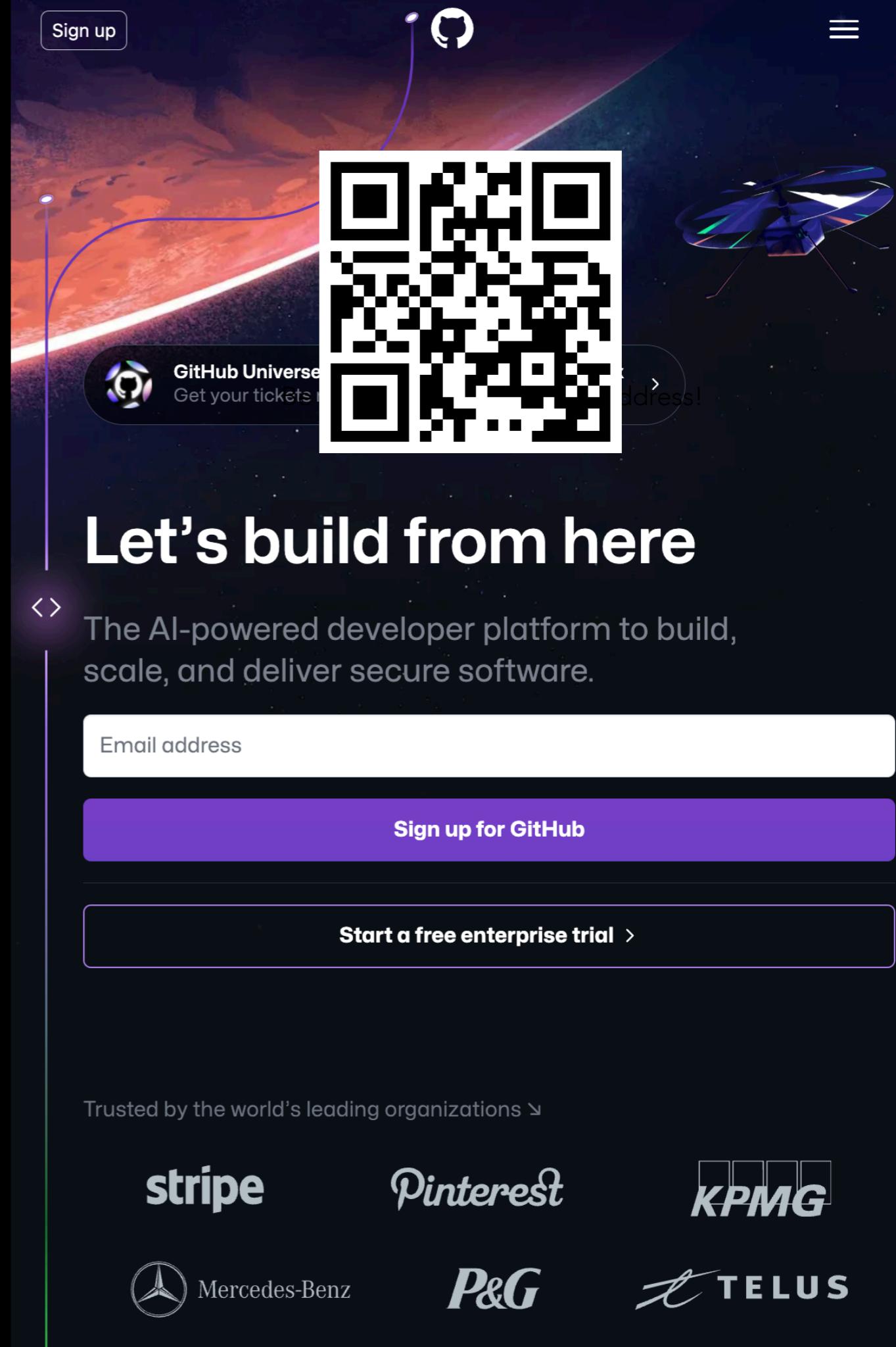
<https://docs.github.com/en/get-started/quickstart/hello-world>



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Your first project

- Step 1. Make a GitHub account  
(surprise, surprise)
- Step 2. Make a new repository
- Step 3. Create a branch
- Step 4. Make + commit changes



The image shows a screenshot of the GitHub sign-up page. At the top right is a "Sign up" button and a GitHub logo. Below it is a large QR code. The background features a dark, futuristic scene with a road on Mars and a satellite in space. A banner at the top says "GitHub Universe Get your tickets now". The main heading "Let's build from here" is prominently displayed in white. Below it, a subtext reads "The AI-powered developer platform to build, scale, and deliver secure software." There is a text input field for an email address and a purple "Sign up for GitHub" button. At the bottom, there is a link "Start a free enterprise trial >" and a section titled "Trusted by the world's leading organizations" featuring logos for Stripe, Pinterest, KPMG, Mercedes-Benz, P&G, and TELUS.

Sign up

GitHub Universe  
Get your tickets now

Let's build from here

The AI-powered developer platform to build, scale, and deliver secure software.

Email address

Sign up for GitHub

Start a free enterprise trial >

Trusted by the world's leading organizations

stripe

Pinterest

KPMG

Mercedes-Benz

P&G

TELUS

# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

- Step 1. Make a GitHub account  
(surprise, surprise)

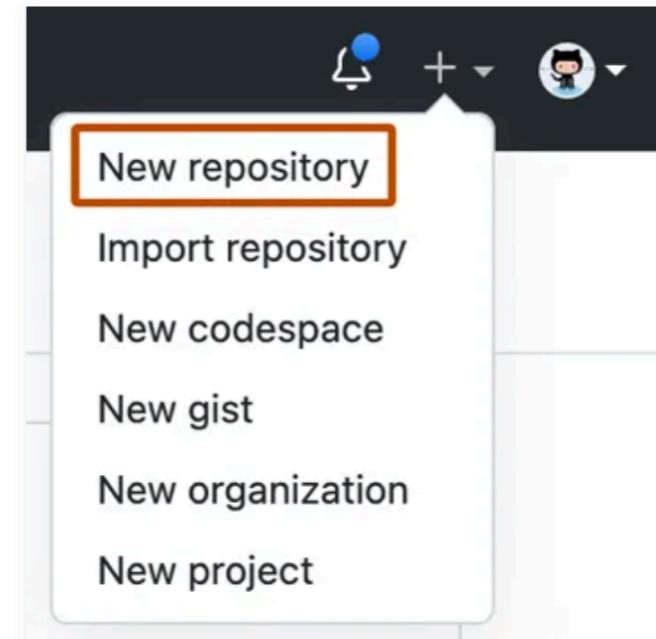
### • Step 2. Make a new repository

- Step 3. Create a branch

- Step 4. Make + commit changes

project

- 1 In the upper-right corner of any page, use the + drop-down menu, and select **New repository**.



- 2 In the "Repository name" box, type `hello-world`.

- 3 In the "Description" box, type a short description.

- 4 Select whether your repository will be **Public** or **Private**.

- 5 Select **Add a README file**.

- 6 Click **Create repository**.

# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

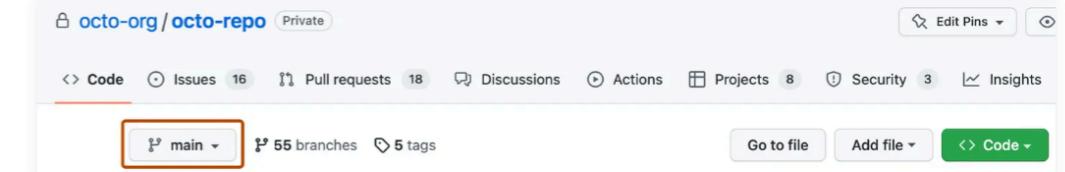
- Step 1. Make a GitHub account (surprise, surprise)

- Step 2. Make a new repository

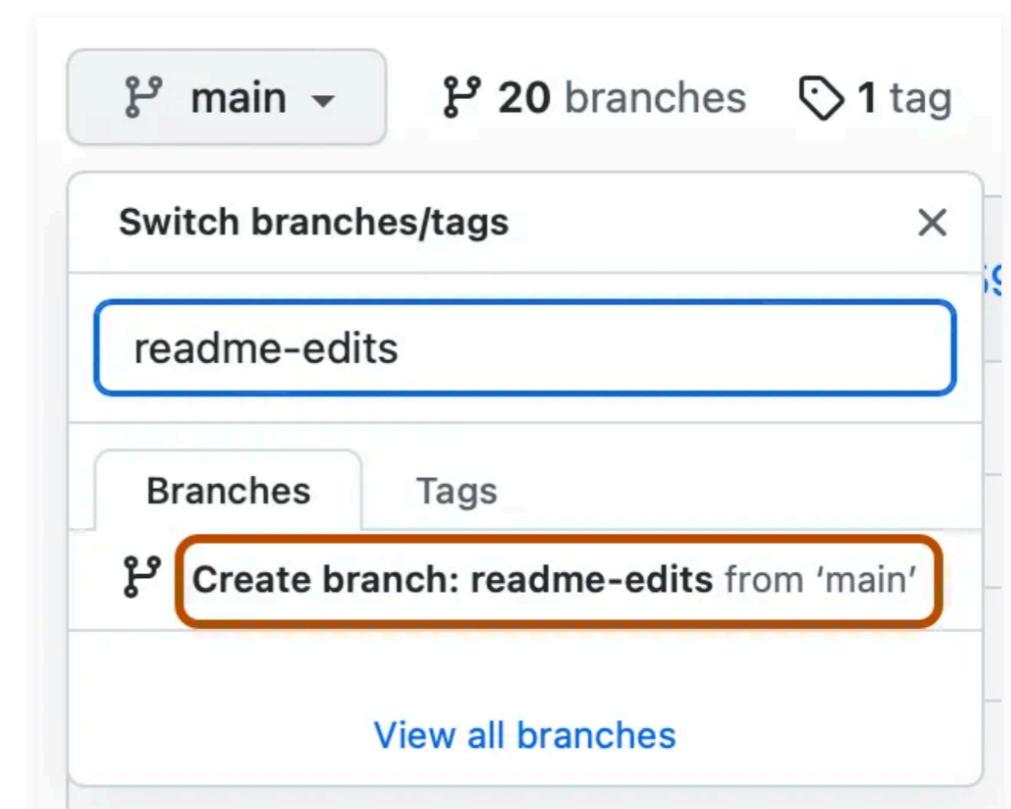
- **Step 3. Create a branch** ↪  
version; useful to separate "development"  
from "main" (production-ready) branches
- Step 4. Make + commit changes



- ① Click the **Code** tab of your `hello-world` repository.
- ② Above the file list, click the dropdown menu that says **main**.



- ③ Type a branch name, `readme-edits`, into the text box.
- ④ Click **Create branch: readme-edits from main**.



Now you have two branches, `main` and `readme-edits`. Right now, they look exactly the same. Next you'll add changes to the new branch.

# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

- Step 1. Make a GitHub account  
(surprise, surprise)
- Step 2. Make a new repository  

- Step 3. Create a branch  
version; useful to separate "development"  
from "main" (production-ready) branches  

- Step 4. Make + commit changes  
save + explain changes,  
like a checkpoint  




- ➊ Under the `readme-edits` branch you created, click the *README.md* file.
- ➋ To edit the file, click .
- ➌ In the editor, write a bit about yourself. Try using different Markdown elements.
- ➍ Click **Commit changes....**
- ➎ In the "Propose changes" box, write a commit message that describes your changes.
- ➏ Click **Propose changes**.



# VERSION CONTROL: A QUICK & DIRTY GUIDE

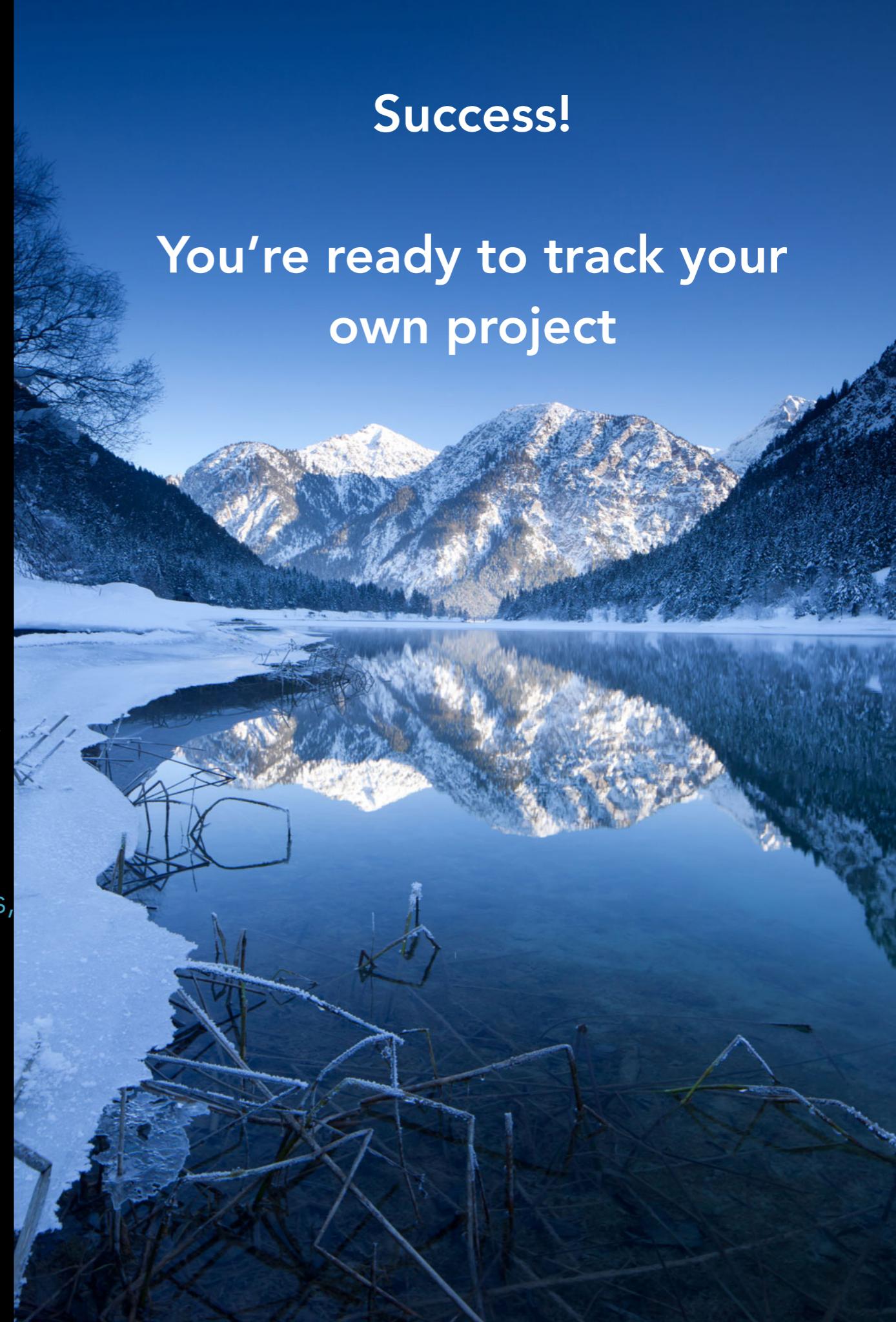
## Quick Start: Your first project

- Step 1. Make a GitHub account  
*(surprise, surprise)*
  - Step 2. Make a new repository  
*↗ project*
  - Step 3. Create a branch  
*↗ version; useful to separate "development" from "main" (production-ready) branches*
  - Step 4. Make + commit changes  
*↗ save + explain changes, like a checkpoint*
- 

*For your very first project, you can definitely stop here. Get comfortable getting your code out there!*

**Success!**

**You're ready to track your own project**



# VERSION CONTROL: A QUICK & DIRTY GUIDE

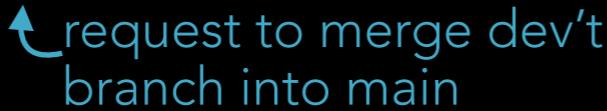
## Quick Start: Your first project

- Step 1. Make a GitHub account  
(surprise, surprise)

- Step 2. Make a new repository  


- Step 3. Create a branch  
version; useful to separate "development"  
from "main" (production-ready) branches

- Step 4. Make + commit changes  
  
save + explain changes,  
like a checkpoint

- 
- Step 5. Make a pull request  
  
request to merge dev't  
branch into main

- Step 6. Merge pull request



- 1 Click the Pull requests tab of your `hello-world` repository.
- 2 Click New pull request
- 3 In the Example Comparisons box, select the branch you made, `readme-edits`, to compare with `main` (the original).
- 4 Look over your changes in the diffs on the Compare page, make sure they're what you want to submit.



Showing 1 changed file with 3 additions and 3 deletions.

6 README.md

```
@@ -1,3 +1,3 @@
 1 - # test-area-2
 2 - edit1
 3 - edit2
 1 + # About me
 2 +
 3 + My name is Mona Lisa.
```

- 5 Click Create pull request.
- 6 Give your pull request a title and write a brief description of your changes. You can include emojis and drag and drop images and gifs.
- 7 Optionally, to the right of your title and description, click the  next to **Reviewers**, **Assignees**, **Labels**, **Projects**, or **Milestone** to add any of these options to your pull request. You do not need to add any yet, but these options offer different ways to collaborate using pull requests. For more information, see "[About pull requests](#)".
- 8 Click Create pull request.

# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

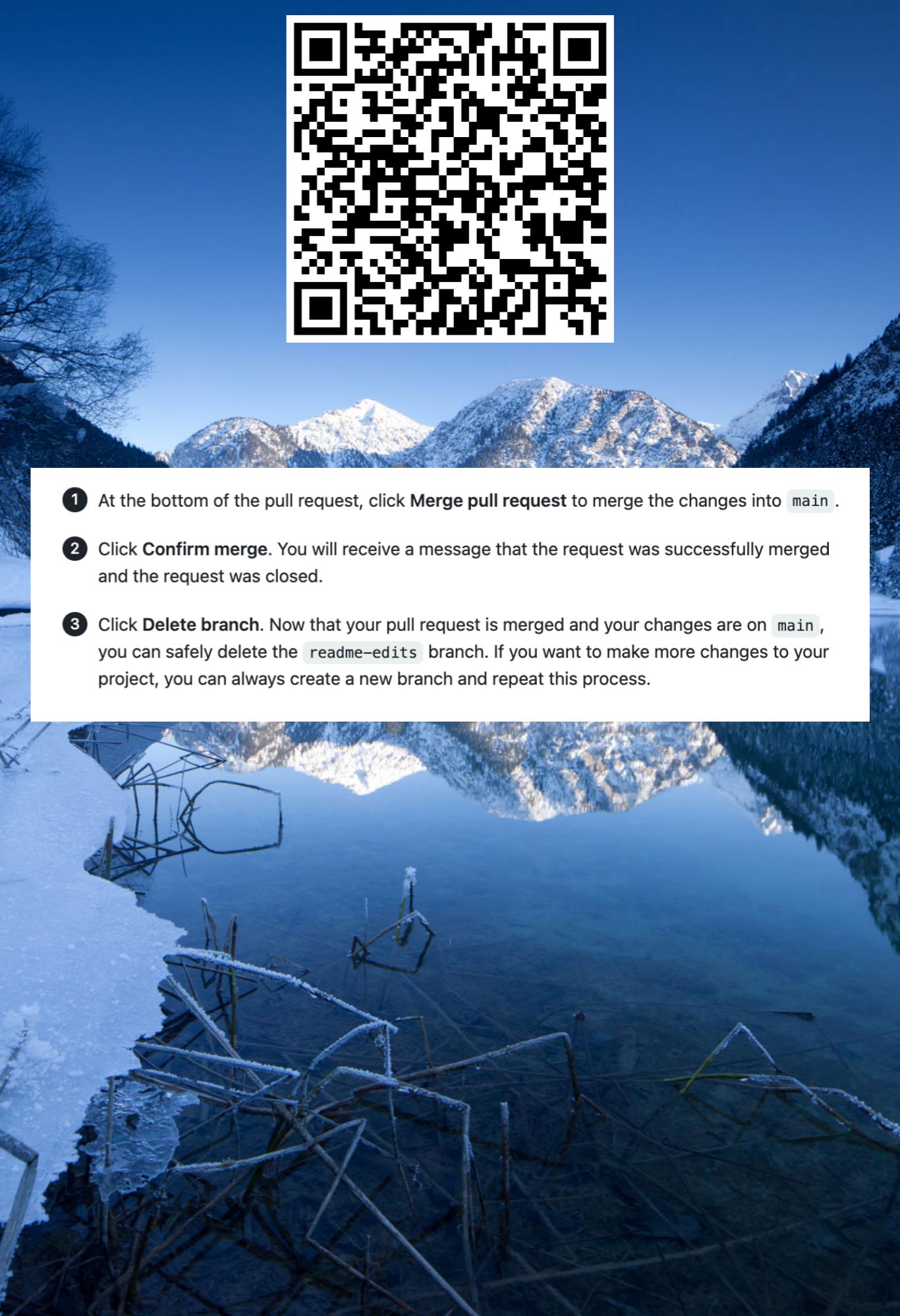
- Step 1. Make a GitHub account  
(surprise, surprise)
- Step 2. Make a new repository  

- Step 3. Create a branch  
version; useful to separate "development" from "main" (production-ready) branches  

- Step 4. Make + commit changes  
save + explain changes, like a checkpoint  

- Step 5. Make a pull request  
request to merge dev't branch into main  

- Step 6. Merge pull request  
accept request, resolving conflicts if necessary  

- ① At the bottom of the pull request, click **Merge pull request** to merge the changes into `main`.
- ② Click **Confirm merge**. You will receive a message that the request was successfully merged and the request was closed.
- ③ Click **Delete branch**. Now that your pull request is merged and your changes are on `main`, you can safely delete the `readme-edits` branch. If you want to make more changes to your project, you can always create a new branch and repeat this process.

# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

- Step 1. Make a GitHub account  
(surprise, surprise)
- Step 2. Make a new repository  
↗ project
- Step 3. Create a branch  
↗ version; useful to separate "development" from "main" (production-ready) branches
- Step 4. Make + commit changes  
↗ save + explain changes, like a checkpoint

---

- Step 5. Make a pull request  
↗ request to merge dev't branch into main
- **Step 6. Merge pull request**  
↗ accept request, resolving conflicts if necessary

Success!

You're ready to manage multiple versions



# VERSION CONTROL: A QUICK & DIRTY GUIDE

## Quick Start: Your first project

- Step 1. Make a GitHub account
  - Step 2. Make a new repository
  - Step 3. Create a branch
  - Step 4. Make + commit changes
- 
- Step 5. Make a pull request
  - Step 6. Merge pull request



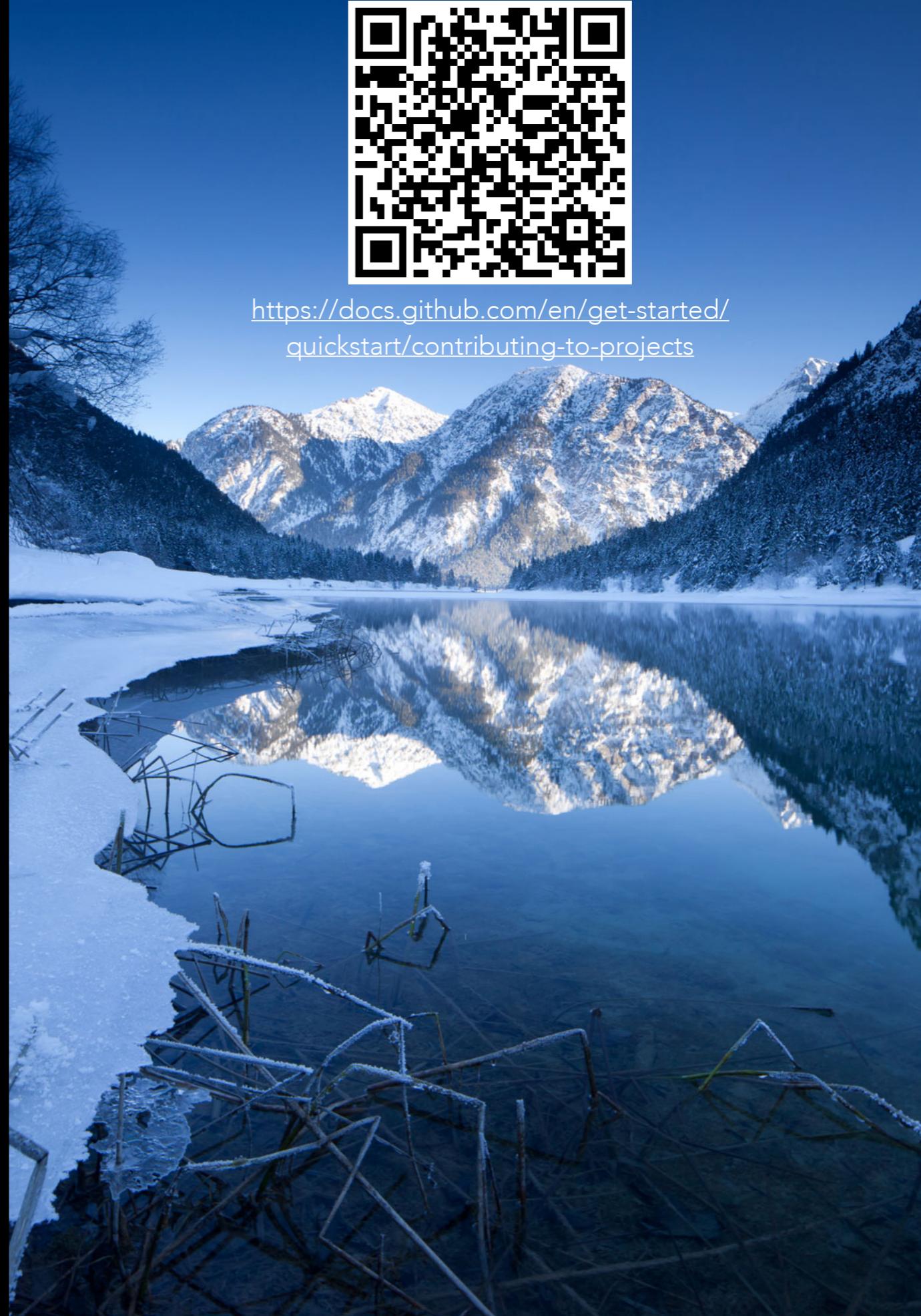
# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account
- Step 2. Make a new repository
- Step 3. Create a branch
- Step 4. Make + commit changes
- Step 5. Make a pull request
- Step 6. Merge pull request



[https://docs.github.com/en/get-started/  
quickstart/contributing-to-projects](https://docs.github.com/en/get-started/quickstart/contributing-to-projects)



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

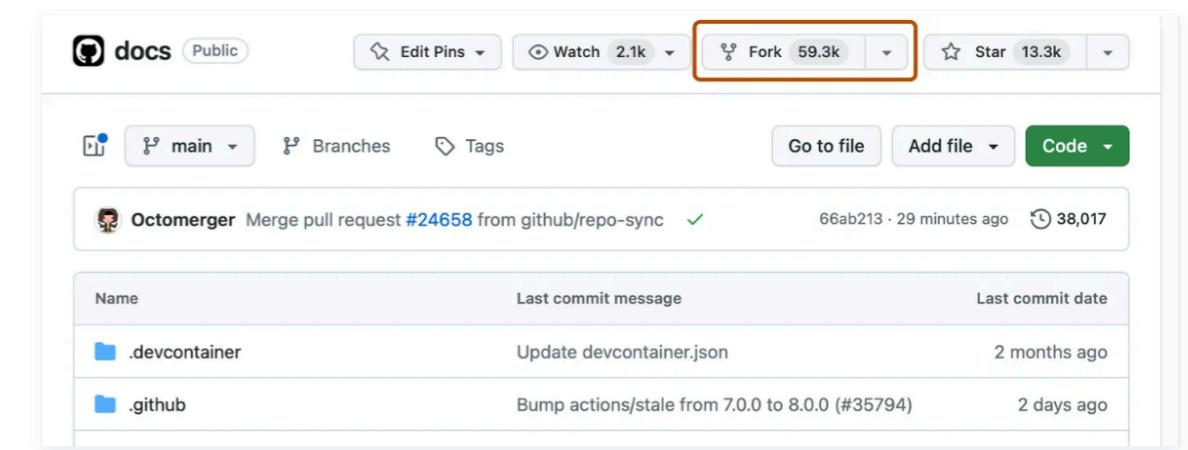
- Step 1. Make a GitHub account
  - copy to your account
  - Fork an existing
- **Step 2. ~~Make a new repository~~**

- Step 3. Create a branch
- Step 4. Make + commit changes

- 
- Step 5. Make a pull request
  - Step 6. Merge pull request



- 1 Navigate to the `Spoon-Knife` project at <https://github.com/octocat/Spoon-Knife>.
- 2 In the top-right corner of the page, click **Fork**.



Name	Last commit message	Last commit date
.devcontainer	Update devcontainer.json	2 months ago
.github	Bump actions/stale from 7.0.0 to 8.0.0 (#35794)	2 days ago

- 3 Under "Owner," select the dropdown menu and click an owner for the forked repository.
- 4 By default, forks are named the same as their upstream repositories. Optionally, to further distinguish your fork, in the "Repository name" field, type a name.
- 5 Optionally, in the "Description" field, type a description of your fork.
- 6 Optionally, select **Copy the DEFAULT branch only**.

# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account
  - copy to your account
  - Fork an existing
- Step 2. ~~Make a new repository~~
- Step 3. Create a branch
- Step 4. Make + commit changes

---

- Step 5. Make a pull request
- Step 6. Merge pull request



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account
  - copy to your account
  - Fork an existing
- Step 2. ~~Make a new repository~~
- Step 3. Create a branch
- Step 4. **Make + commit changes**

---

- Step 5. Make a pull request
- Step 6. Merge pull request



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account

copy to your account

Fork an existing

- Step 2. ~~Make a new repository~~

- Step 3. Create a branch

- Step 4. Make + commit changes

- 
- Step 5. Make a pull request ↪

This time, between your fork and the  
original repo

- Step 6. Merge pull request



To do so, head on over to the repository on GitHub where your project lives. For this example, it would be at [https://github.com/<your\\_username>/Spoon-Knife](https://github.com/<your_username>/Spoon-Knife). You'll see a banner indicating that your branch is one commit ahead of `octocat:main`. Click **Contribute** and then **Open a pull request**.

GitHub will bring you to a page that shows the differences between your fork and the `octocat/Spoon-Knife` repository. Click **Create pull request**.

GitHub will bring you to a page where you can enter a title and a description of your changes. It's important to provide as much useful information and a rationale for why you're making this pull request in the first place. The project owner needs to be able to determine whether your change is as useful to everyone as you think it is. Finally, click **Create pull request**.



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account

copy to your account  


Fork an existing

- Step 2. ~~Make a new repository~~

- Step 3. Create a branch

- Step 4. Make + commit changes

- 
- Step 5. Make a pull request ↪

This time, between your fork and the  
original repo

- Step 6. ~~Merge pull request~~

Manage feedback ↪

The ball is now in their court, but you can help!



Pull Requests are an area for discussion. In this case, the Octocat is very busy, and probably won't merge your changes. For other projects, don't be offended if the project owner rejects your pull request, or asks for more information on why it's been made. It may even be that the project owner chooses not to merge your pull request, and that's totally okay. Your changes exist in your fork. And who knows--maybe someone you've never met will find your changes much more valuable than the original project.



# VERSION CONTROL: A QUICK & DIRTY GUIDE

Quick Start: Contributing to a project  
when you're not on the core team

- Step 1. Make a GitHub account
  - copy to your account
  - Fork an existing
- Step 2. ~~Make a new repository~~

- Step 3. Create a branch
- Step 4. Make + commit changes

- 
- Step 5. Make a pull request ↪  
This time, between your fork and the  
original repo
  - Step 6. ~~Merge pull request~~  
Manage feedback ↪  
The ball is now in their court, but you can help!

Success!

You're ready to contribute to  
the world



# VERSION CONTROL: THE QUICK & DIRTY GUIDE

Common bottlenecks

# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Common bottlenecks

- Using commands you don't understand can be dangerous
  - Example dangerous commands: `git reset`, `git checkout -f` (force), `git push -f`
    - Overwrite files you're working on locally, mess up the repo, destroy records, etc.
  - **Solution:** Learn about a command before you use it and think twice if it's described as dangerous. If you are desperate and no one else is working with you, manually make a local copy of your work first. Don't do this at all on team projects.

# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Common bottlenecks

- Using commands you don't understand can be dangerous
  - Example dangerous commands: `git reset`, `git checkout -f` (force), `git push -f`
    - Overwrite files you're working on locally, mess up the repo, destroy records, etc.
  - **Solution:** Learn about a command before you use it and think twice if it's described as dangerous. If you are desperate and no one else is working with you, manually make a local copy of your work first. Don't do this at all on team projects.
- Merging can be messy as your project becomes more complicated
  - **Solution:** integrating branches is an art. For an actively developing project, try to regularly keep development branches up-to-date with the latest version of the master

# VERSION CONTROL: THE QUICK & DIRTY GUIDE

## Common bottlenecks

- Using commands you don't understand can be dangerous
  - Example dangerous commands: `git reset`, `git checkout -f` (force), `git push -f`
    - Overwrite files you're working on locally, mess up the repo, destroy records, etc.
  - **Solution:** Learn about a command before you use it and think twice if it's described as dangerous. If you are desperate and no one else is working with you, manually make a local copy of your work first. Don't do this at all on team projects.
- Merging can be messy as your project becomes more complicated
  - **Solution:** integrating branches is an art. For an actively developing project, try to regularly keep development branches up-to-date with the latest version of the master
- **General advice:** Don't worry too much. Commit often, learn as you go along, and be encouraged that using git means you have lots of new (safe) options to undo mistakes (<https://ohshitgit.com/>)

# OPEN CODE: OTHER RESOURCES

[More Git Resources](#)

# OPEN CODE: OTHER RESOURCES

## More Git Resources

- 5 minute Git (GitHub Desktop): <https://alicelepisier.com/git-tutorial/slides.html#1>
- Interactive intro to command line Git: <https://learngitbranching.js.org/>
- Deeper dive into command line Git (OHBM OSSIG): <https://ohbm.github.io/hackathon2021/traintrack/>

# OPEN CODE: OTHER RESOURCES

## More Git Resources

- 5 minute Git (GitHub Desktop): <https://alicelepisier.com/git-tutorial/slides.html#1>
- Interactive intro to command line Git: <https://learngitbranching.js.org/>
- Deeper dive into command line Git (OHBM OSSIG): <https://ohbm.github.io/hackathon2021/traintrack/>

## Related topics

- **Code testing (OHBM OSSIG):** [https://www.youtube.com/watch?v=gfPP2pQ8Rms&ab\\_channel=OHBMOpenScienceSIG](https://www.youtube.com/watch?v=gfPP2pQ8Rms&ab_channel=OHBMOpenScienceSIG)
  - Benefit: quickly know whether code is accurate or if you broke something (an expensive lesson)
  - Quick recommendation: Can be useful for most people. Take a look.
- **Reproducible workflows (OHBM OSSIG):** [https://www.youtube.com/watch?v=tk2eZSrM8oA&ab\\_channel=OHBMOpenScienceSIG](https://www.youtube.com/watch?v=tk2eZSrM8oA&ab_channel=OHBMOpenScienceSIG) ; <https://brainlife.io/about/>
  - Benefit: Maximize others' ability to reproduce (version control, testing, coding best practices, environments/containers, etc.)
  - Quick recommendation: More useful for intermediate / expert applications

# OPEN DATA: PUBLIC DATASET SHARING & REUSE

# OPEN DATA: PUBLIC DATASET SHARING & REUSE

# SHARING DATA PUBLICLY

Why bother?

# SHARING DATA PUBLICLY

## Why bother?

- Dissemination, recognition (citations), & collaboration
  - You won't get scooped, but many people release with their first publication
- NIH wants you to (cf. Data Management & Sharing Plan)
- Trend towards sharing full statistical maps
- Reusable for others and you, too
- Bonus: BIDS apps work with BIDS data (fMRIPrep, MRIQC)

# SHARING DATA PUBLICLY

## Why bother?

- Dissemination, recognition (citations), & collaboration
  - You won't get scooped, but many people release with their first publication
- NIH wants you to (cf. Data Management & Sharing Plan)
- Trend towards sharing full statistical maps
- Reusable for others and you, too
- Bonus: BIDS apps work with BIDS data (fMRIPrep, MRIQC)
- Common bottlenecks:
  - Some medical institutions won't let you
  - As usual, involves some effort (but worth the benefits)

# SHARING DATA PUBLICLY

nature human behaviour

Perspective | Published: 07 December 2020

## A hitchhiker's guide to working with large, open-source neuroimaging datasets

[Corey Horien](#) , [Stephanie Noble](#), [Abigail S. Greene](#), [Kangjoo Lee](#), [Daniel S. Barron](#), [Siyuan Gao](#), [David O'Connor](#), [Mehraveh Salehi](#), [Javid Dadashkarimi](#), [Xilin Shen](#), [Evelyn M. R. Lake](#), [R. Todd Constable](#) & [Dustin Scheinost](#) 

*Nature Human Behaviour* 5, 185–193 (2021) | [Cite this article](#)

4328 Accesses | 23 Citations | 83 Altmetric | [Metrics](#)

### Abstract

Large datasets that enable researchers to perform investigations with unprecedented rigor are growing increasingly common in neuroimaging. Due to the simultaneous increasing popularity of open science, these state-of-the-art datasets are more accessible than ever to researchers around the world. While analysis of these samples has pushed the field forward, they pose a new set of challenges that might cause difficulties for novice users. Here we offer practical tips for working with large datasets from the end-user's perspective. We cover all aspects of the data lifecycle: from what to consider when downloading and storing the data to

# SHARING DATA PUBLICLY



nature human behaviour

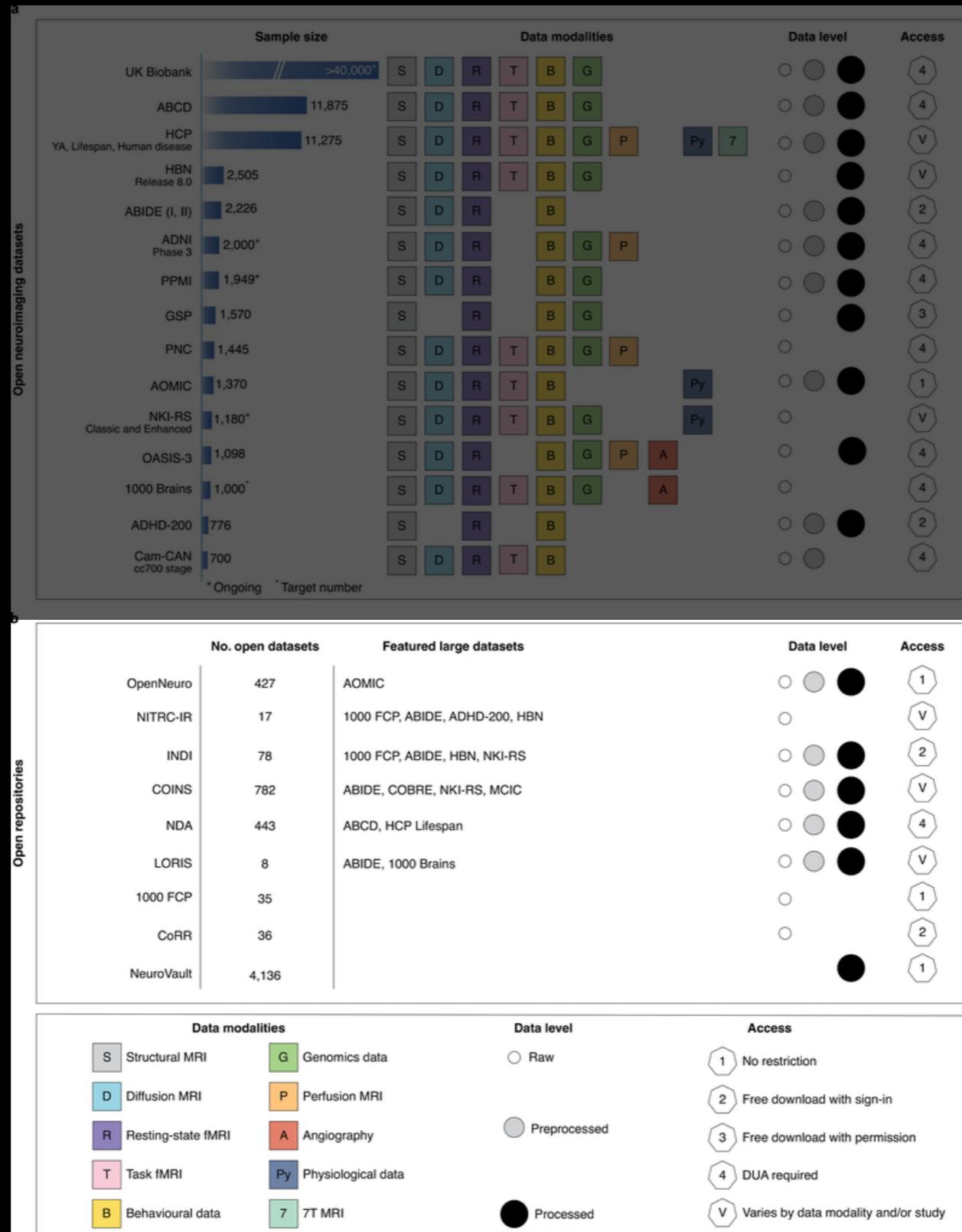
Perspective | Published: 07 December 2020

**A hitchhiker's guide to working with large, open-source neuroimaging datasets**

Corey Horien Stephanie Noble, Abigail S. Greene, Kangjoo Lee, Daniel S. Barron, Siyuan Gao, David O'Connor, Mehraveh Salehi, Javid Dadashkarimi, Xilin Shen, Evelyn M. R. Lake, R. Todd Constable & Dustin Scheinost



# SHARING DATA PUBLICLY



**nature human behaviour**

Perspective | Published: 07 December 2020

**A hitchhiker's guide to working with large, open-source neuroimaging datasets**

Corey Horien , Stephanie Noble, Abigail S. Greene, Kangjoo Lee, Daniel S. Barron, Siyuan Gao, David O'Connor, Mehraveh Salehi, Javid Dadashkarimi, Xilin Shen, Evelyn M. R. Lake, R. Todd Constable & Dustin Scheinost



# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- **Step 1. Determine what to share**

Raw or preprocessed?

- Step 2. Confirm you're allowed to share

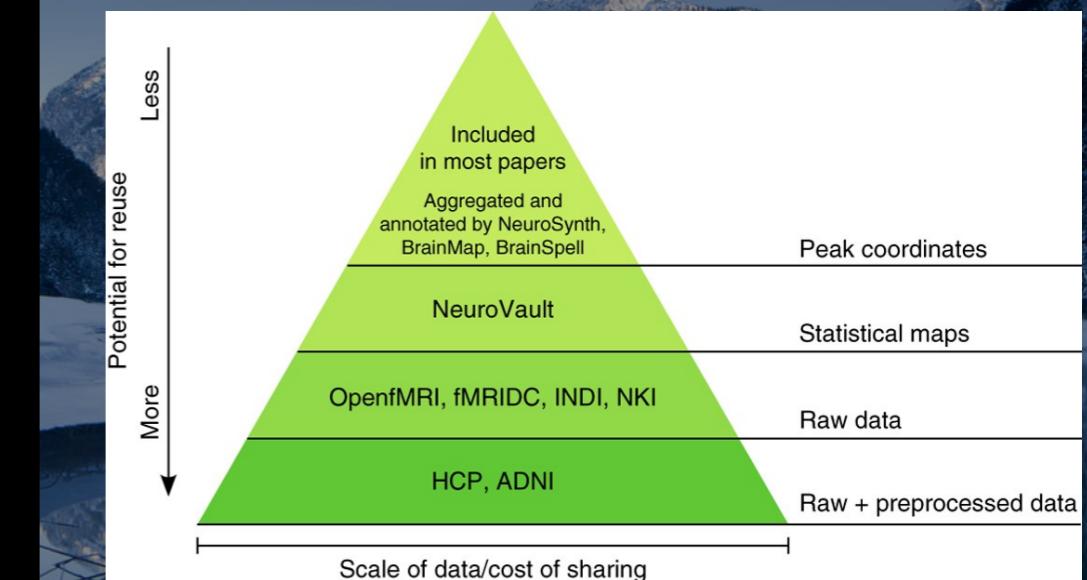
- Step 3. Find a database and sign up

- Step 5. Anonymize

- Step 6. Confirm other requirements

- Step 7. Upload!

Imagine a dataset you  
could contribute



<https://www.nature.com/articles/nn.3818>

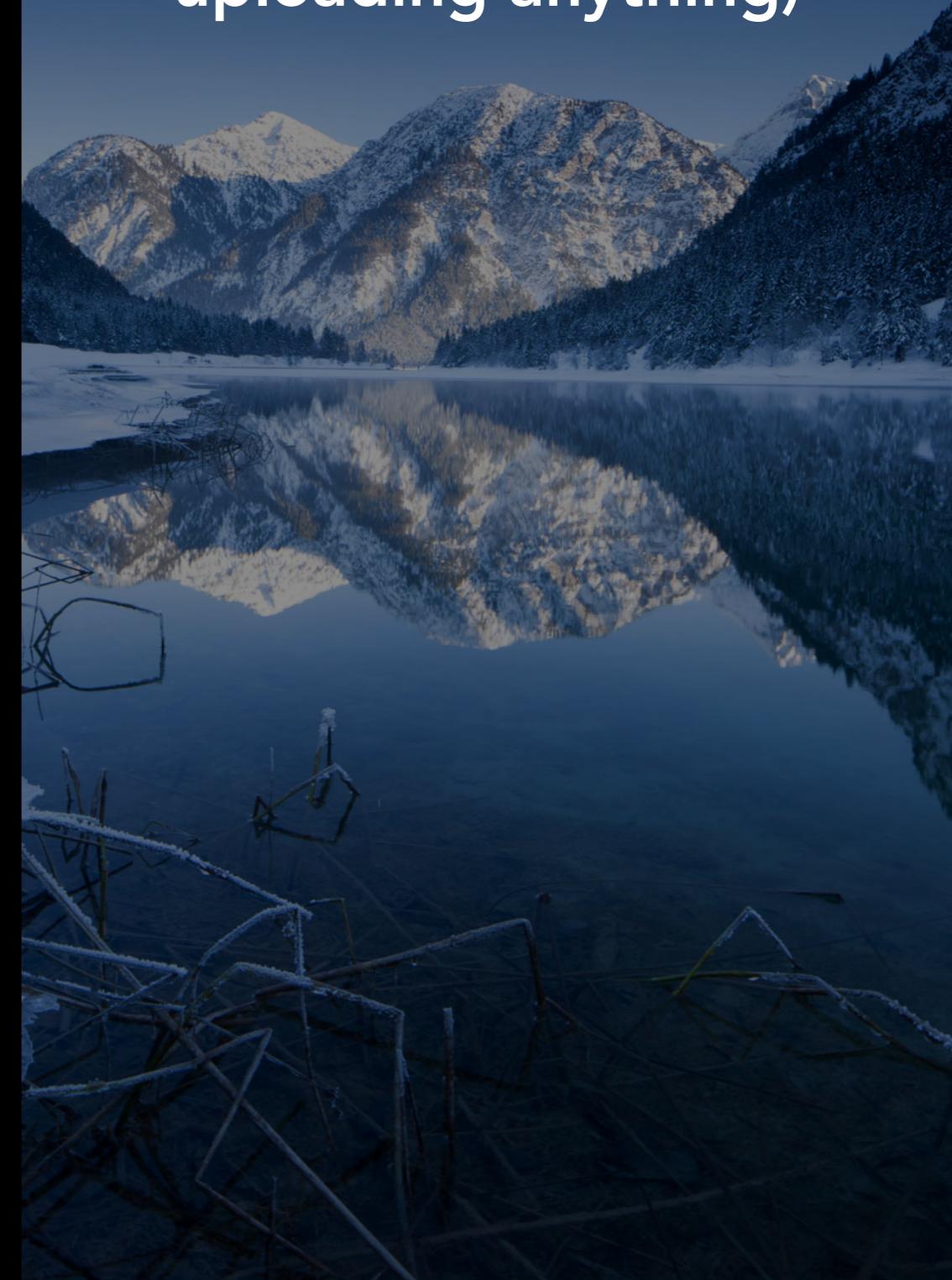
How much could you reasonably share?

# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what to share  
Raw or preprocessed?
- **Step 2. Confirm you're allowed to share**  
Consent, IRB, etc
- Step 3. Find a database and sign up
- Step 5. Anonymize
- Step 6. Confirm other requirements
- Step 7. Upload!

For today you can pretend  
(we're not actually  
uploading anything)



# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what to share  
Raw or preprocessed?
- Step 2. Confirm you're allowed to share  
Consent, IRB, etc
- **Step 3. Find a database and sign up**  
Review contribution guide to check: 1) you can share the level of data you want, & 2) you have all required data
- Step 5. Anonymize
- Step 6. Confirm other requirements
- Step 7. Upload!

## Guides for Two Potential Databases:

### INDI guide



[http://fcon\\_1000.projects.nitrc.org/indi/  
indi\\_data\\_contribution\\_guide.pdf](http://fcon_1000.projects.nitrc.org/indi/indi_data_contribution_guide.pdf)

Requires (1) T1, (1) resting state scan, age, sex, & handedness

### OpenNeuro guide



[https://docs.openneuro.org/user\\_guide.html](https://docs.openneuro.org/user_guide.html)

# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what to share  
Raw or preprocessed?
- Step 2. Confirm you're allowed to share  
Consent, IRB, etc
- Step 3. Find a database and sign up  
Review contribution guide to check: 1) you can share the level of data you want, & 2) you have all required data
- **Step 5. Anonymize**  
Deface structural scan, rename, remove name and DOB, etc. (see contribution guide)
- Step 6. Confirm other requirements
- Step 7. Upload!

poldracklab/  
**pydeface**

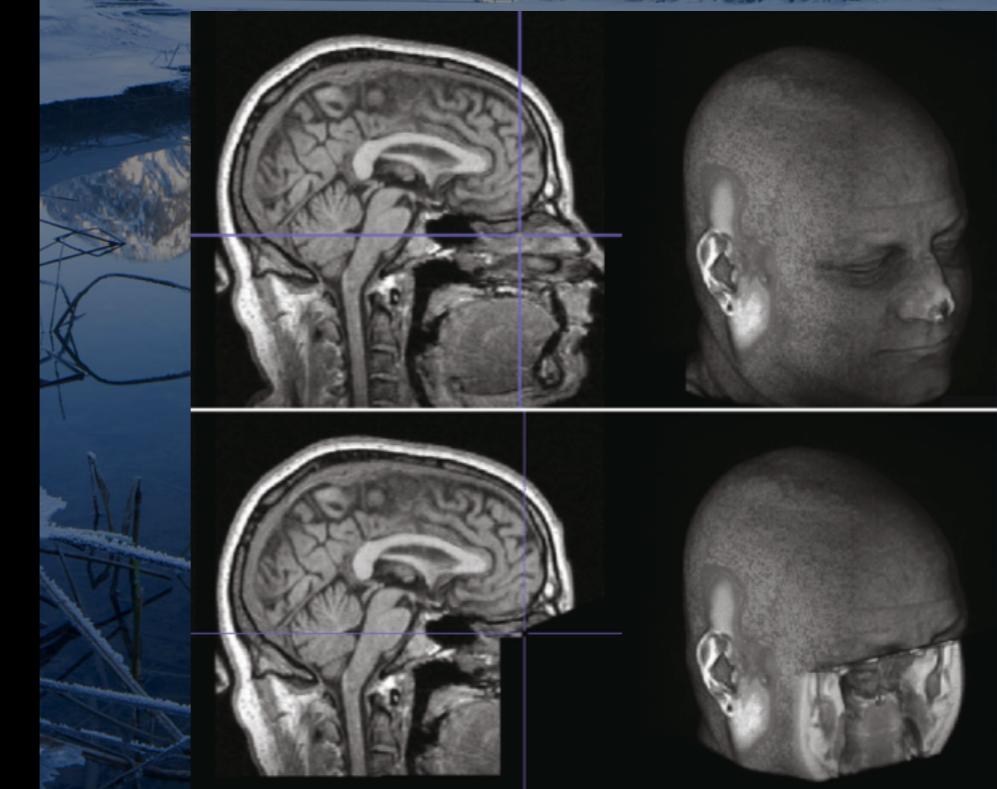
defacing utility for MRI images

9  
Contributors

21  
Used by

98  
Stars

37  
Forks



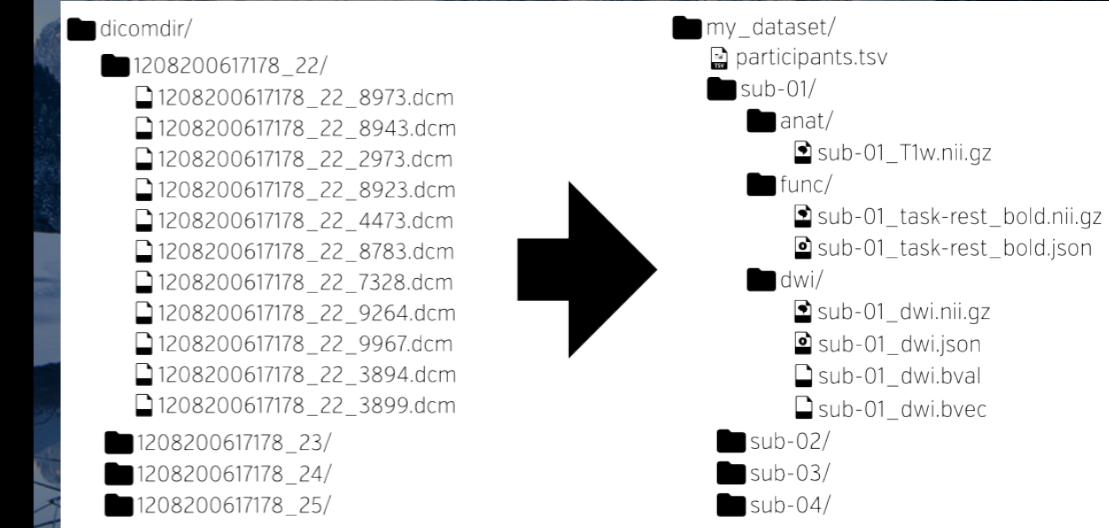
[https://www.physicamedica.com/article/  
S1120-1797\(21\)00095-8/fulltext](https://www.physicamedica.com/article/S1120-1797(21)00095-8/fulltext)

# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what to share  
Raw or preprocessed?
- Step 2. Confirm you're allowed to share  
Consent, IRB, etc
- Step 3. Find a database and sign up  
Review contribution guide to check: 1) you can share the level of data you want, & 2) you have all required data
- Step 5. Anonymize  
Deface structural scan, rename, remove name and DOB, etc. (see contribution guide)
- **Step 6. Confirm other requirements**  
E.g., BIDS format (OpenNeuro), reorient (optional but desirable for INDI), metadata, description, etc.
- Step 7. Upload!

## BIDS: Brain Imaging Data Structure



<https://hackmd.io/@effigies/bids-derivatives-readme>

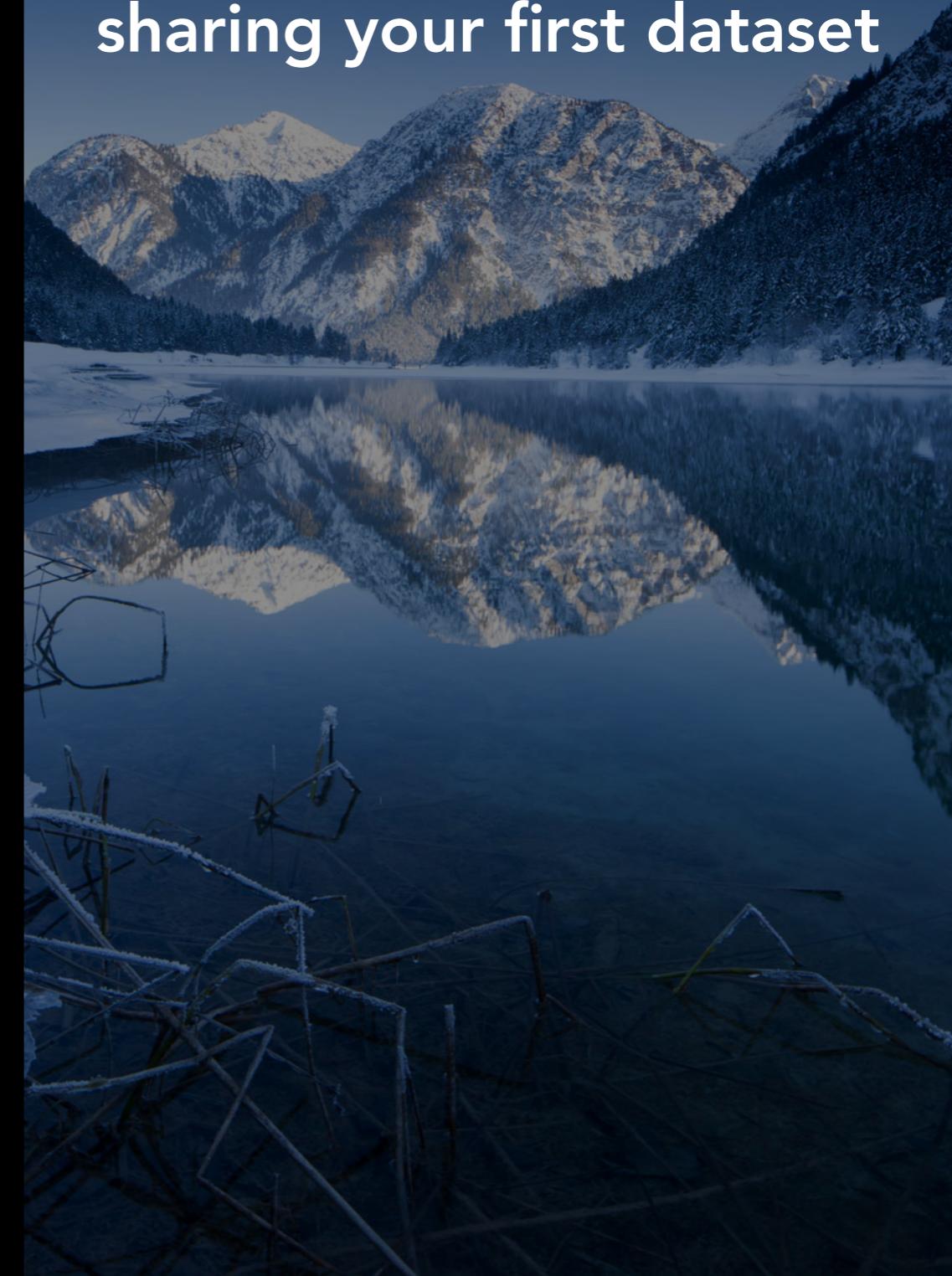
# SHARING DATA PUBLICLY: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what to share  
Raw or preprocessed?
- Step 2. Confirm you're allowed to share  
Consent, IRB, etc
- Step 3. Find a database and sign up  
Review contribution guide to check: 1) you can share the level of data you want, & 2) you have all required data
- Step 5. Anonymize  
Deface structural scan, rename, remove name and DOB, etc. (see contribution guide)
- Step 6. Confirm other requirements  
E.g., BIDS format (OpenNeuro), reorient (optional but desirable for INDI), metadata, description, etc.
- **Step 7. Upload!**

Success!

You are on your way to  
sharing your first dataset



# SHARING DATA PUBLICLY

Common bottlenecks

# SHARING DATA PUBLICLY

## Common bottlenecks

- Code for preparing data can sometimes be finicky, and reorganization can get annoying

# SHARING DATA PUBLICLY

## Common bottlenecks

- Code for preparing data can sometimes be finicky, and reorganization can get annoying
  - Not a deal-breaker, just adjust time expectations accordingly

# SHARING DATA PUBLICLY

## Common bottlenecks

- Code for preparing data can sometimes be finicky, and reorganization can get annoying
  - Not a deal-breaker, just adjust time expectations accordingly
- May be prohibited by some medical institutions, especially VA

# SHARING DATA PUBLICLY

More Data Sharing & Reuse Resources

# SHARING DATA PUBLICLY

## More Data Sharing & Reuse Resources

- Where to share data and sharing results - see Horien et al., 2021: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7992920/>
- More considerations regarding data sharing (Open Research at the Wellcome Centre for Integrative Neuroimaging): <https://open.win.ox.ac.uk/pages/open-science/community/Open-WIN-Community/docs/data/>
- BIDS & BIDS Apps
  - Great guide to both: Markiewicz et al. <https://hackmd.io/@effigies/bids-derivatives-readme>
  - <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005209>
  - <https://psy6983.brainhackmtl.org/modules/bids/>

# SHARING DATA PUBLICLY

## More Data Sharing & Reuse Resources

- Where to share data and sharing results - see Horien et al., 2021: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7992920/>
- More considerations regarding data sharing (Open Research at the Wellcome Centre for Integrative Neuroimaging): <https://open.win.ox.ac.uk/pages/open-science/community/Open-WIN-Community/docs/data/>
- BIDS & BIDS Apps
  - Great guide to both: Markiewicz et al. <https://hackmd.io/@effigies/bids-derivatives-readme>
  - <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005209>
  - <https://psy6983.brainhackmtl.org/modules/bids/>

## Related topics

- Data Management (DataLad): <https://www.datalad.org/>
  - Benefit: record how you've changed your data
  - Expert level: moderate/high. Recommended for large projects and tracking processing.

# OPEN DATA: PUBLIC DATASET SHARING & REUSE

# OPEN DATA: PUBLIC DATASET SHARING & REUSE

# USING PUBLICLY AVAILABLE DATA

Why bother?

# USING PUBLICLY AVAILABLE DATA

## Why bother?

- Save time, \$\$, effort
  - Feasibility in grants
- Larger samples than most labs can collect
  - Facilitates rigorous & reproducible science
  - Many of the most important recent papers in NI have relied on open datasets
- Learn transferable skills

# USING PUBLICLY AVAILABLE DATA

## Why bother?

- Save time, \$\$, effort
  - Feasibility in grants
- Larger samples than most labs can collect
  - Facilitates rigorous & reproducible science
  - Many of the most important recent papers in NI have relied on open datasets
- Learn transferable skills
- Common bottlenecks:
  - For large ( $n>500$ ) datasets, need moderately large computational resources (most of you have access to HPC, though)
  - May not include specialized populations (though more and more populations are being released)

# USING PUBLICLY AVAILABLE DATA



nature human behaviour

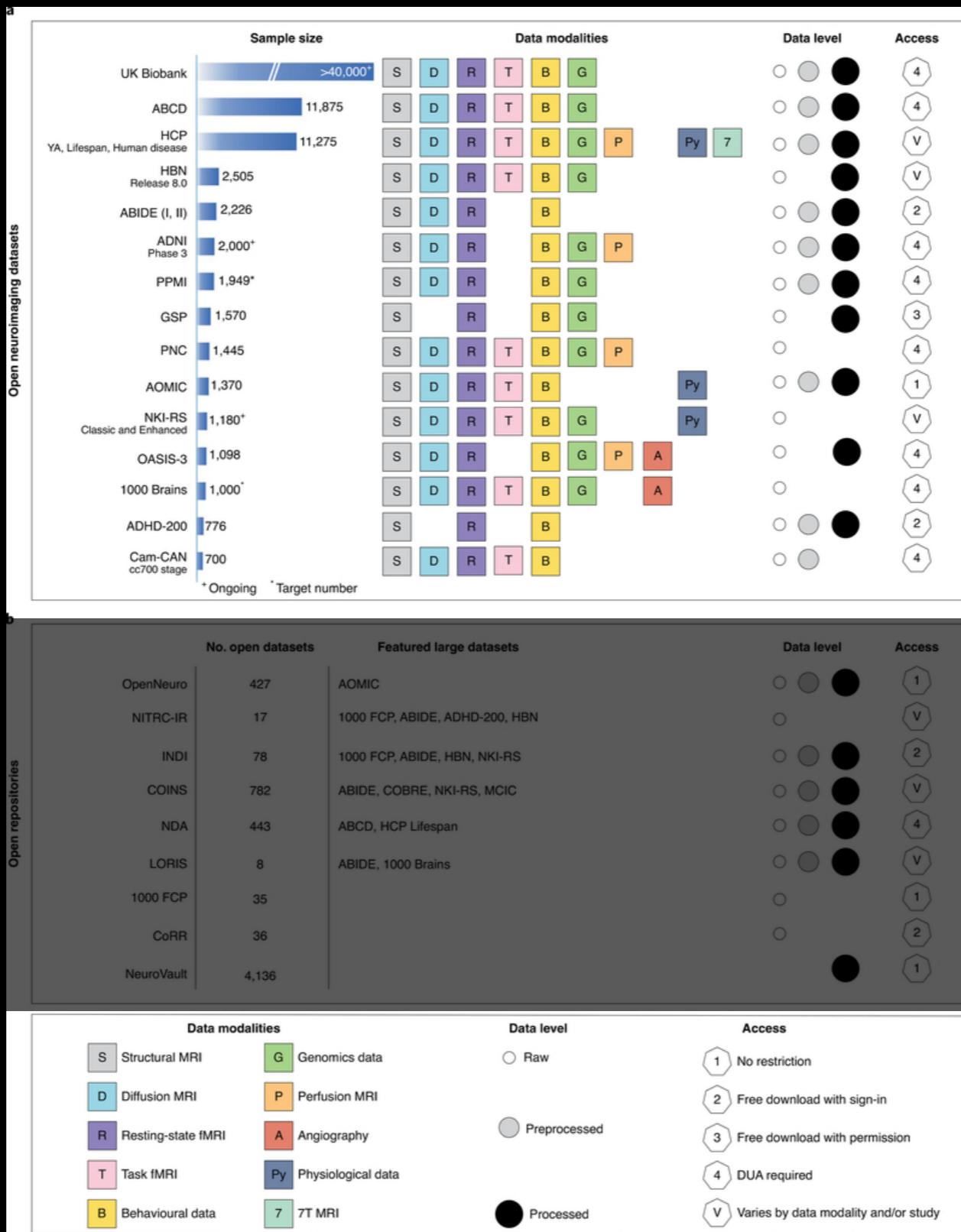
Perspective | [Published: 07 December 2020](#)

# A hitchhiker's guide to working with large, open-source neuroimaging datasets

[Corey Horien](#)  [Stephanie Noble](#), [Abigail S. Greene](#), [Kangjoo Lee](#), [Daniel S. Barron](#), [Siyuan Gao](#), [David O'Connor](#), [Mehraveh Salehi](#), [Javid Dadashkarimi](#), [Xilin Shen](#), [Evelyn M. R. Lake](#), [R. Todd Constable](#) & [Dustin Scheinost](#) 



# USING PUBLICLY AVAILABLE DATA



**nature human behaviour**

Perspective | Published: 07 December 2020

**A hitchhiker's guide to working with large, open-source neuroimaging datasets**

Corey Horien , Stephanie Noble, Abigail S. Greene, Kangjoo Lee, Daniel S. Barron, Siyuan Gao, David O'Connor, Mehraveh Salehi, Javid Dadashkarimi, Xilin Shen, Evelyn M. R. Lake, R. Todd Constable & Dustin Scheinost



# PUBLIC DATA REUSE: A QUICK & DIRTY GUIDE

## Quick Start

- **Step 1. Determine what data you want**  
What do you want to do? Which databases have that data?
- Step 2. Confirm ethical approval
- Step 3. Confirm storage & computing
- Step 4. If required: create account, sign data use agreement, etc.
- Step 5. Download some data!

NeuroVault



A public repository of unthresholded statistical maps, parcellations, and atlases of the brain.

What is it?	Why use it?	Supported by
A place where researchers can publicly store and share unthresholded statistical maps, parcellations, and atlases produced by MRI and PET studies.	<ul style="list-style-type: none"><li>• Interactive visualization</li><li>• A permanent URL</li><li>• Publicly shareable</li><li>• Improves meta-analyses</li></ul>	  

[Get started and upload an image!](#)

Recently added collections of images from published papers

Name	Number of images
Odour-imagery ability is associated with change in humans	5
Auditory cortical microstructure during voice and speech processing	7
Affective speech modulates the neural mechanisms underlying perceived aesthetic value	7
Consciously Feeling the Connection between the Brain and the Body: Neural Mechanisms Underlying the Sense of Agency	6
Reward Learning over Weeks versus Minutes Increases the Neural Representation of Value in the Human Brain	75
Reactive vs. Proactive Executive Decision Making in the Human Brain	14
Episodic memory retrieval success is associated with rapid replay of hippocampal episodes	20
Distinct replay signatures for prospective decision-making and memory preservation	9

<https://neurovault.org/>

**Let's get some processed data from NeuroVault!**

# PUBLIC DATA REUSE: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what data you want  
What do you want to do? Which databases have that data?
- **Step 2. Confirm ethical approval**  
Generally straightforward to seek IRB Exemption 4 unless data identifiable in combination
- Step 3. Confirm storage & computing
- Step 4. If required: create account, sign data use agreement, etc.
- Step 5. Download some data!



# PUBLIC DATA REUSE: A QUICK & DIRTY GUIDE

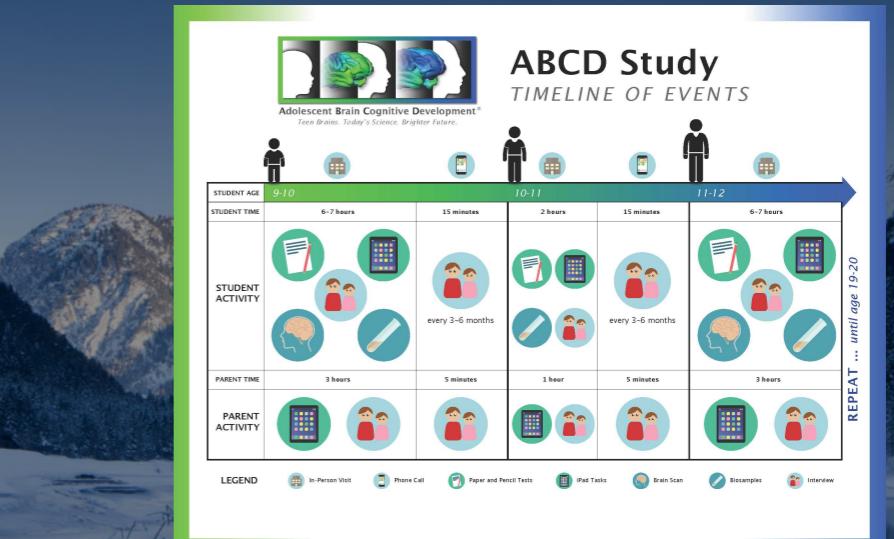
## Quick Start

- Step 1. Determine what data you want  
What do you want to do? Which databases have that data?
- Step 2. Confirm ethical approval  
Generally straightforward to seek IRB Exemption 4 unless data identifiable in combination
- **Step 3. Confirm storage & computing**
  - Storage: Plan for additions, backups, + any intermediates
  - Computing: Time your pipeline + anticipate how to scale
- Step 4. If required: create account, sign data use agreement, etc.
- Step 5. Download some data!

### Example: ABCD release 1 (~10,000 individuals)

#### Storage

- Raw NifTI data: ~13.5 TB
- Preprocessed connectivity matrices: ~25.6 MB
- More data released over the years



#### Computing

- Connectome-behavior regression <1 day on laptop
- Processing raw data >1 mo on HPC or AWS
  - Scaling up: time a couple subjects + account for parallelization
  - Special considerations, e.g., HCP already on AWS so credits, fewer transfer limits, etc.

In-lab cluster < university cluster < AWS

\$\$

user expertise

In-lab maintenance

# PUBLIC DATA REUSE: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what data you want  
What do you want to do? Which databases have that data?
- Step 2. Confirm ethical approval  
Generally straightforward to seek IRB Exemption 4 unless data identifiable in combination
- Step 3. Confirm storage & computing
  - Storage: Plan for additions, backups, + any intermediates
  - Computing: Time your pipeline + anticipate how to scale
- **Step 4. If required: create account, sign data use agreement, etc.**
- Step 5. Download some data!



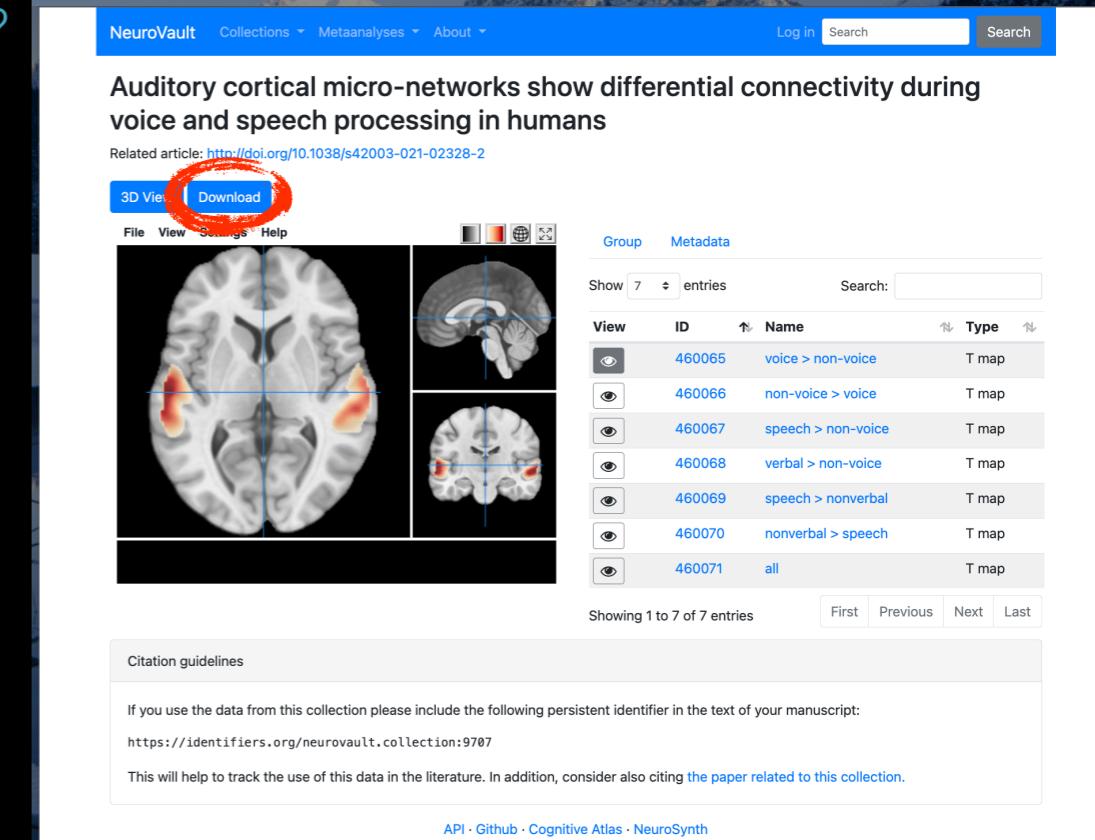
# PUBLIC DATA REUSE: A QUICK & DIRTY GUIDE

## Quick Start

- Step 1. Determine what data you want  
What do you want to do? Which databases have that data?
- Step 2. Confirm ethical approval  
Generally straightforward to seek IRB Exemption 4 unless data identifiable in combination
- Step 3. Confirm storage & computing
  - Storage: Plan for additions, backups, + any intermediates
  - Computing: Time your pipeline + anticipate how to scale
- Step 4. If required: create account, sign data use agreement, etc.
- Step 5. Download some data!



<https://neurovault.org/collections/9810/>



Auditory cortical micro-networks show differential connectivity during voice and speech processing in humans

Related article: <http://doi.org/10.1038/s42003-021-02328-2>

View	ID	Name	Type
eye	460065	voice > non-voice	T map
eye	460066	non-voice > voice	T map
eye	460067	speech > non-voice	T map
eye	460068	verbal > non-voice	T map
eye	460069	speech > nonverbal	T map
eye	460070	nonverbal > speech	T map
eye	460071	all	T map



# PUBLIC DATA REUSE

Common bottlenecks

# PUBLIC DATA REUSE

## Common bottlenecks

- Expertise + setup for getting very large datasets (>5,000; e.g., UKB) can be nontrivial

# PUBLIC DATA REUSE

## Common bottlenecks

- Expertise + setup for getting very large datasets (>5,000; e.g., UKB) can be nontrivial
- Processing large datasets (>1,000; e.g., HCP) is time-intensive (6-9 months for ABCD release 1)
  - if you're completely new, may be worth starting with preprocessed data to get a feel for it

# PUBLIC DATA REUSE

## Common bottlenecks

- Expertise + setup for getting very large datasets (>5,000; e.g., UKB) can be nontrivial
- Processing large datasets (>1,000; e.g., HCP) is time-intensive (6-9 months for ABCD release 1)
  - if you're completely new, may be worth starting with preprocessed data to get a feel for it
- AWS can be (accidentally) expensive—best to get an expert involved
  - Before you run anything: always test on a small scale first
  - During: check outputs periodically
  - After: double check you've closed your instance

# PUBLIC DATA REUSE

# PUBLIC DATA REUSE

## Other Resources

- Large, publicly available datasets (Horien et al., 2021): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7992920/>
- In addition to official guides, many datasets have third-party guides, e.g.:
  - HCP Tutorial (Andy's Brain Blog): [https://andysbrainbook.readthedocs.io/en/latest/HCP/HCP\\_1\\_DownloadData.html](https://andysbrainbook.readthedocs.io/en/latest/HCP/HCP_1_DownloadData.html)
  - A Hitchhiker's Guide to OpenNeuro: Secondary Analysis on the Web's Largest Repository of Open Neuroimaging Data (Buckser, 2022): <http://d-scholarship.pitt.edu/42175/>
- High performance computing (Barch et al., 2019) [https://training.incf.org/sites/default/files/CloudBasedComputer\\_MATRIX.pdf](https://training.incf.org/sites/default/files/CloudBasedComputer_MATRIX.pdf)

# PUBLIC DATA REUSE

## Other Resources

- Large, publicly available datasets (Horien et al., 2021): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7992920/>
- In addition to official guides, many datasets have third-party guides, e.g.:
  - HCP Tutorial (Andy's Brain Blog): [https://andysbrainbook.readthedocs.io/en/latest/HCP/HCP\\_1\\_DownloadData.html](https://andysbrainbook.readthedocs.io/en/latest/HCP/HCP_1_DownloadData.html)
  - A Hitchhiker's Guide to OpenNeuro: Secondary Analysis on the Web's Largest Repository of Open Neuroimaging Data (Buckser, 2022): <http://d-scholarship.pitt.edu/42175/>
- High performance computing (Barch et al., 2019) [https://training.incf.org/sites/default/files/CloudBasedComputer\\_MATRIX.pdf](https://training.incf.org/sites/default/files/CloudBasedComputer_MATRIX.pdf)

## Related topics

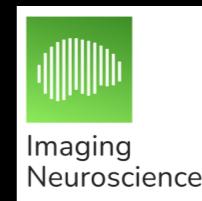
- Data Management (DataLad): <https://www.datalad.org/>
  - Scaling up: Managing 80TB and 15 million files from the HCP release: [https://handbook.datalad.org/en/latest/usecases/HCP\\_dataset.html](https://handbook.datalad.org/en/latest/usecases/HCP_dataset.html)

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN PUBLISHING

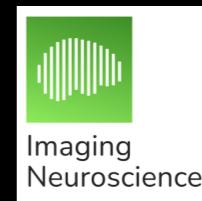
# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN PUBLISHING

- What is it
  - Open access, non-profit journals (Imaging Neuroscience! Aperture!), preprints, registered reports, non-traditional research objects, etc.



# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN PUBLISHING

- What is it
  - Open access, non-profit journals (Imaging Neuroscience! Aperture!), preprints, registered reports, non-traditional research objects, etc.
- Why bother?
  - Again, dissemination, recognition (citations), & collaboration
  - Often faster, can be lower cost



# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN PUBLISHING

- What is it
  - Open access, non-profit journals (Imaging Neuroscience! Aperture!), preprints, registered reports, non-traditional research objects, etc.
- Why bother?
  - Again, dissemination, recognition (citations), & collaboration
  - Often faster, can be lower cost
- Fastest + easiest way to get started
  - Post on arXiv / OSF, find community-led low fee open-access journals (Aperture, Imaging Neuroscience - fee waivers!)



# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN PUBLISHING

- What is it
  - Open access, non-profit journals (Imaging Neuroscience! Aperture!), preprints, registered reports, non-traditional research objects, etc.
- Why bother?
  - Again, dissemination, recognition (citations), & collaboration
  - Often faster, can be lower cost
- Fastest + easiest way to get started
  - Post on arXiv / OSF, find community-led low fee open-access journals (Aperture, Imaging Neuroscience - fee waivers!)
- Common bottlenecks
  - Growing recognition, competing against “glam” journals
    - ...but there is lots of community momentum + worth the benefits
  - Cost (depending)



# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN COMMUNITY

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN COMMUNITY

- What it is
  - Actively inclusive & collaborative communities, e.g., Brainhack, OHBM Open Science Room, Neuromatch

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN COMMUNITY

- What it is
  - Actively inclusive & collaborative communities, e.g., Brainhack, OHBM Open Science Room, Neuromatch
- Why bother?
  - Network, support, training, future collaborations, etc.

## Neuron

CellPress

### NeuroView

#### Brainhack: Developing a culture of open, inclusive, community-driven neuroscience

Rémi Gau,<sup>1,7\*</sup> Stephanie Noble,<sup>2</sup> Katja Heuer,<sup>3,4,7\*</sup> Katherine L. Bottnerom,<sup>5,7\*</sup> Işıl P. Bilgin,<sup>5,7,7\*</sup> Yu-Fang Yang,<sup>5,7\*</sup> Julia M. Hunttenberg,<sup>5,7\*</sup> Johanna M.M. Bayer,<sup>1,6,1,7\*</sup> Richard A.J. Bethlehem,<sup>1,2,3,7,7\*</sup> Christoph Vogelbacher,<sup>1,5</sup> Valentina Borghesani,<sup>1,6</sup> Elizabeth Levitt,<sup>1,7,18</sup> Hao-Ting Wang,<sup>19,20,21</sup> Sofie Van Den Bossche,<sup>22</sup> Xenia Koleleva,<sup>23,24</sup> Jon Haltz Legarreta,<sup>25</sup> Samuel Guay,<sup>26</sup> Selim Melvin Atay,<sup>27</sup> Gael P. Varoquaux,<sup>28,29</sup> Dorian C. Huijser,<sup>30,31</sup> Malin S. Sandström,<sup>32</sup> Peer Herholz,<sup>33</sup> Samuel A. Nastase,<sup>34</sup> Amanpreet Badhwar,<sup>1,5,35,36</sup> Guillaume Dumas,<sup>37,38</sup> Simon Schwab,<sup>39</sup> Stefano Moia,<sup>40,41</sup> Michael Dayan,<sup>42</sup> Yasmine Bassil,<sup>43</sup> Paula P. Brooks,<sup>34</sup> Matteo Marzini,<sup>32,44,45</sup> James M. Shine,<sup>46</sup> David O'Connor,<sup>47</sup> Xihé Xie,<sup>48</sup> Davide Pogglial<sup>49</sup> Patrick Friedrich,<sup>50</sup>

### CONSORTIA

The Brainhack Community includes: Nasim Anousheh, Aurora Arakereenute, Giacomo Azzone, Dipesh Barua, Daniel Bautista, Ruperto Beran, Arshitha Basavareddy, Marco Bedini, Pierre Bellec, R. Austin Bern, Kathryn Berluti, Steffen Bollmann, Saskia Bollmann, Claire Bradley, Jesse Brown, Augustus Buchwitz, Patrick Callahan, Theresia Caneila Y. Chan, Bramish Q. Chandio, Theresa Cheung, Siddarth Chopra, Al Weisnburg, Thomas G. Cleary, Etienne Coquelin, Gérard Corneille, R. Todd Constable, Claire Cury, Kamalaker Dadi, Pablo F. Damasceno, Samir Das, Fabrizio De Vico Fallani, Krista DeSasio, Erin W. Dickie, Lena Dorfschmidt, Eugene P. Duff, Elizabeth DuPre, Scott Duan, Daniel Eason, Odilia Esteban, Shreyas Fadnavis, Guillaume Flandrin, Jessica E. Flannery, John Flournoy, Stephanie J. Forkel, Alexandre R. Franco, Saamprasad Ganesh, Siyuan Gao, José C. García Alainis, Eleftherios Garyfallidis, Tristan Glärd, Enrico Glerean, Javier González-Carrasco, Camilo Gómez van Praet, Raul Gómez-Sanz, O. Helchenko, Daniel Heindl, Stephan Heunis, Felix Hoffstaedter, Daniela M. Hohmann, Corey Horien, Horea-Ioan Ioanis, Alexander D. Jordan, Chao Jiang, Michael J. Kennedy, Karthik Karthik, D. Michael J. Kennedy, Anisha Keshkoven, Ali R. Khan, Gregory Kiar, P. Christian Klink, Vincent Koppelman, Serge Koudoro, Angela R. Laird, Georg Lange, Marissa L. Laws, Roxane Licard, Sogol-Laleh Lorestani, Tomislav Lukić, Klaszana Lutza, Daniel J. Lurie, Déborah Lüssier, Christopher R. Madan, Lea-Theresa Mais, Sina Mansour L., J.P. Manzano-Patron, Dimitra Maoutsa, Mathieu Marcon, Daniel S. Margulies, Giorgio Marinato, Daniele Marinazzo, Christos Marinakis, Mathilde Marinier, Michael P. Mihm, Mengzhuo David Minzner, Michael P. Mitham, Kathryn L. Mills, Devide Morni, Clara A. Moreau, Aysha Motala, Iska Moxon-Erme, Thomas E. Nichols, Dylan M. Nilsson, Gustav Nilssonne, Lisa Novello, Caroline O'Donnell, Emily O'Farrell, Linda O'Keeffe, John A. Ondrus, Michael P. O'Neill, Kendra Ouchi, Patrick J. Park, Mahboobeh Paraspour, Lorenzo Pasquini, Scott Peltier, Cyril R. Pernet, Rudolph Pieparni, Pedro Pinheiro-Chagas, Jean-Baptiste Poline, Anqi Qiu, Tiago Quendera, Laura C. Rice, Joscelyn Sanchez-Hidalgo, Seige Rutherford, Matias Scherzer, Barbara Schreiner, Daniel Schreiter, Thomas B. Shaw, Ivanna Sitese, Melly Simantirik, Nikolicz Simplicatz, Hayli Spence, Julia Spranger, Andrija Stajduhar, Martin Szinte, Sylvain Takerkart, Angela Tam, Link Tejalibulya, Michel Thiebaut de Schotten, Ina Thomé, Laura Tomaz da Silva, Nicolas Tardieu, Daniel O. Ullrich, Michael Urban, W. VanKester, Namrita Vijayakumar, Matteo Visconti di Oleggio Castello, Jakub Vohnýzek, Jákša Vuković, Kirstie Jane Whitaker, Lucy Whitmore, Steve Wideman, Suzanne T. Witt, Hua Xie, Ting Xu, Chao-Gan Yan, Fang-Cheng Yeh, B.T. Thomas Yeo, and Xi-Nian Zuo.

Brainhack is an innovative meeting format that promotes scientific collaboration and education in an open, inclusive environment. This NeuroView describes the myriad benefits for participants and the research community and how Brainhacks complement conventional formats to augment scientific progress.

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN COMMUNITY

- What it is
  - Actively inclusive & collaborative communities, e.g., Brainhack, OHBM Open Science Room, Neuromatch
- Why bother?
  - Network, support, training, future collaborations, etc.
- Fastest + easiest way to get started
  - Join an activity or volunteer! So many opportunities that all usually hybrid: Brainhack Global, Brainhack OHBM, OHBM Open Science Room, Neurohackademy

**Neuron**



**NeuroView**  
**Brainhack: Developing a culture of open, inclusive, community-driven neuroscience**

Rémi Gau,<sup>1,7\*</sup> Stephanie Noble,<sup>2,7</sup> Katja Heuer,<sup>3,4,7</sup> Katherine L. Bottnerom,<sup>5,7</sup> Işıl P. Bilgin,<sup>5,7,7B</sup> Yu-Fang Yang,<sup>5,7B</sup> Julia M. Huntzman,<sup>9,7B</sup> Johanna M.M. Bayer,<sup>10,1,7B</sup> Richard A.J. Bethlehem,<sup>12,13,17B</sup> Shawn A. Hodges,<sup>14</sup> Christoph Vogelbacher,<sup>15</sup> Valentina Borghesani,<sup>16</sup> Elizabeth Levitt,<sup>17,18</sup> Hao-Ting Wang,<sup>19,20,21</sup> Sofie Van Den Bossche,<sup>22</sup> Xenia Koleleva,<sup>23,24</sup> Jon Haltz Legarreta,<sup>25</sup> Samuel Guay,<sup>26</sup> Selim Melvin Atay,<sup>27</sup> Gael P. Varoquaux,<sup>28,29</sup> Dorian C. Huijser,<sup>30,31</sup> Malin S. Sandström,<sup>32</sup> Peer Herholz,<sup>33</sup> Samuel A. Nastase,<sup>34</sup> Amanpreet Badhwar,<sup>1,35,36</sup> Guillermo Dumas,<sup>37,38</sup> Simon Schwab,<sup>39</sup> Stefano Moia,<sup>40,41</sup> Michael Dayan,<sup>42</sup> Yasmine Bassil,<sup>43</sup> Patrick P. Brooks,<sup>34</sup> Matteo Marzini,<sup>34,44,45</sup> James M. Shine,<sup>46</sup> David O'Connor,<sup>47</sup> Xiehe Xie,<sup>48</sup> Davide Poggliali,<sup>49</sup> Patrick Friedrich,<sup>50</sup> Georg Lange, Marissa L. Laws, Roxane Licard, Sogukalp Lisan, Tuncay Lükc, Klaszana Lutza, Daniel J. Lurie, Déborah Lüssier, Christopher R. Madan, Lea-Theresa Mais, Sina Mansour L., J.P. Manzano-Patron, Dimitra Maousta, Mathieu Marcon, Daniel S. Margulies, Giorgio Marinato, Daniele Marinazzo, Christopher Marinazzo, Camille Marquet, Farah Menguzzu, David Mähnner, Michael P. Milham, Kathryn L. Mills, Davide Momi, Clara A. Moreau, Aysha Motala, Iska Moxon-Erme, Thomas E. Nichols, Dylan M. Nilsson, Gustav Nilssonne, Lisa Novello, Caroline O'Doherty, Emily O'Farrell, Linda O'Keeffe, John A. Ondracek, Michael P. O'Neill, Kendra Outcalt, Patrick J. Park, Mahboobeh Paraspour, Lorenzo Pasquini, Scott Pettier, Cyril R. Pernet, Rudolph Pievani, Pedro Pinheiro-Chagas, Jean-Baptiste Poline, Anqi Qiu, Tiago Quendera, Laura C. Rice, Joscelyn Sanchez-Hidalgo, Seige Rutherford, Matias Scherzer, Barbara Schmid, Daniel Schmitt, Thomas B. Shaw, Ivana Sitese, Molly Simmonite, Nikoloz Simplitatz, Hayli Spence, Julia Spranger, Andrija Stajduhar, Martin Szinte, Sylvain Takerkart, Angela Tam, Link Tejajivilya, Michel Thiebaud de Schotten, Ina Thomé, Laura Tomaz da Silva, Nicolas Van Der Steene, Q. Umar, Daniel Vaidyanathan, W. VanKester, Narodita Vijayakumar, Matteo Visconti di Oleggio Castello, Jakub Vohnýk, Jákša Vuković, Kirstie Jane Whitaker, Lucy Whithrone, Steve Wideman, Suzanne T. Witt, Hua Xie, Ting Xu, Chao-Gan Yan, Fang-Cheng Yeh, B.T. Thomas Yeo, and Xi-Nian Zhu.

Brainhack is an innovative meeting format that promotes scientific collaboration and education in an open, inclusive environment. This NeuroView describes the myriad benefits for participants and the research community and how Brainhacks complement conventional formats to augment scientific progress.

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: OPEN COMMUNITY

- What it is
  - Actively inclusive & collaborative communities, e.g., Brainhack, OHBM Open Science Room, Neuromatch
- Why bother?
  - Network, support, training, future collaborations, etc.
- Fastest + easiest way to get started
  - Join an activity or volunteer! So many opportunities that all usually hybrid: Brainhack Global, Brainhack OHBM, OHBM Open Science Room, Neurohackademy
- Common bottlenecks
  - You (or your mentor) may not have the funds or need to justify
    - Many virtual options
    - Justification: provides training and networking (Gau et al., 2021)

**Neuron**



**NeuroView**  
**Brainhack: Developing a culture of open, inclusive, community-driven neuroscience**

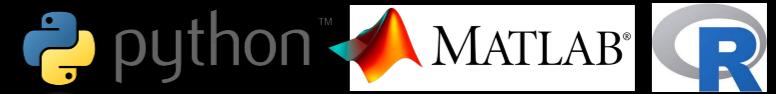
Rémi Gau,<sup>1,7\*</sup> Stephanie Noble,<sup>2,7</sup> Katja Heuer,<sup>3,7,8</sup> Katherine L. Bottnerom,<sup>3,7</sup> Ieill P. Bilgin,<sup>3,7,7b</sup> Yu-Fang Yang,<sup>3,7b</sup> Julia M. Huntenburg,<sup>3,7b</sup> Johanna M.M. Bayer,<sup>1,6,7b</sup> Richard A.J. Bethlehem,<sup>1,2,3,7b</sup> Shawn A. Rohrads,<sup>1,4</sup> Christoph Vogelbacher,<sup>1,5</sup> Valentina Borghesani,<sup>1,6</sup> Elizabeth Levitt,<sup>1,7,8</sup> Hao-Ting Wang,<sup>1,9,20,21</sup> Sofie Van Den Bossche,<sup>22</sup> Xenia Koleleva,<sup>23,24</sup> Jon Haltz Legarreta,<sup>25</sup> Samuel Guay,<sup>26</sup> Selim Melvin Atay,<sup>27</sup> Gael P. Varouquaux,<sup>28,29</sup> Dorian C. Huijser,<sup>30,31</sup> Malin S. Sandström,<sup>32</sup> Peer Herholz,<sup>33</sup> Samuel A. Nastase,<sup>34</sup> Amanpreet Badhwar,<sup>1,6,35,36</sup> Guillermo Dumas,<sup>37,38</sup> Simon Schwab,<sup>39</sup> Stefano Moia,<sup>40,41</sup> Michael Dayan,<sup>42</sup> Yasmine Bassil,<sup>43</sup> Michaela P. Brooks,<sup>34</sup> Matteo Marzini,<sup>32,44,45</sup> James M. Shine,<sup>46</sup> David O'Connor,<sup>47</sup> Xiehe Xie,<sup>48</sup> Davide Poggiali,<sup>49</sup> Patrick Friedrich,<sup>30</sup> Georg Lange, Marissa L. Laws, Roxane Licard, Sogukalp Lisan, Tuncay Lükc, Klaszana Lutza, Daniel J. Lurie, Déborah Lüssier, Christopher R. Madan, Lea-Theresa Mais, Sina Mansour L., J.P. Manzano-Patron, Dimitra Maoutsa, Mathieu Marcon, Daniela S. Margulies, Giorgio Marinato, Daniele Marinazzo, Christopher Marinazzo, Camille Marquet, Fabrizio Menguzzo, David Mähner, Michael P. Milham, Aysha Motala, Iska Moxon-Erme, Thomas E. Nichols, Dylan M. Nilsson, Gustav Nilsson, Lisa Novello, Caroline O'Doherty, Emily O'Farrell, Linda O'Keeffe, John A. Ondracek, Daniel P. Olszewski, Kendra Outcalt, Patrick J. Park, Mahboobeh Paraspour, Lorenzo Pasquini, Scott Peltier, Cyril R. Pernet, Rudolph Pieznik, Pedro Pinheiro-Chagas, Jean-Baptiste Poline, Anqi Qiu, Tiago Quendera, Laura C. Rice, Joscelyn Sanchez-Hidalgo, Seige Rutherford, Matias Scherzer, Viviana Silveira, Daniel Sternig, Thomas B. Stewart, Ivana Stiles, Melly Simmonds, Nikolicz Simplicatz, Hayli Spence, Julia Spranger, Andrija Stajduhar, Martin Szinte, Sylvain Takekart, Angela Tam, Link Teljevisibulya, Michel Thiebaud de Schotten, Ina Thomé, Laura Tomaz da Silva, Nicolas Tardieu, Daniel O. Uebel, and Michael W. VanKester, Narinda Vijayakumar, Matteo Visconti di Oleggio Castello, Jakub Vohnyček, Jákša Vuković, Kirstie Jane Whitaker, Lucy Whitmore, Steve Wideman, Suzanne T. Witt, Hua Xie, Ting Xu, Chao-Gan Yan, Fang-Cheng Yeh, B.T. Thomas Yeo, and Xi-Nian Zhu.

Brainhack is an innovative meeting format that promotes scientific collaboration and education in an open, inclusive environment. This NeuroView describes the myriad benefits for participants and the research community and how Brainhacks complement conventional formats to augment scientific progress.

# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: MODERN SCIENTIFIC PROGRAMMING

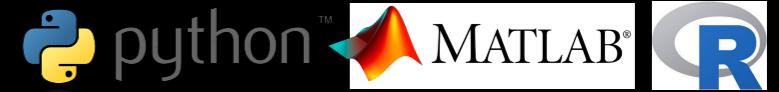
# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: MODERN SCIENTIFIC PROGRAMMING

- Highlights
  - Python (historically software dev), Matlab (historically engineering), R (historically stats)
    - “Historically”, but times are changing
    - Neuroimaging-specific libraries: Nipype, Nilearn
  - Programming assistance: GitHub copilot, ChatGPT (but be careful)
  - Containers (e.g., Docker)



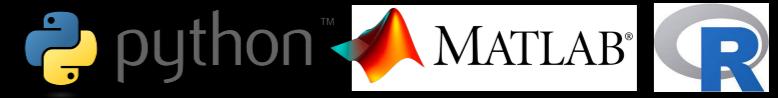
# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: MODERN SCIENTIFIC PROGRAMMING

- Highlights
  - Python (historically software dev), Matlab (historically engineering), R (historically stats)
    - “Historically”, but times are changing
    - Neuroimaging-specific libraries: Nipype, Nilearn
  - Programming assistance: GitHub copilot, ChatGPT (but be careful)
  - Containers (e.g., Docker)
- Why bother?
  - Building blocks of modern neuroscience = more rigorous projects, transferrable skills



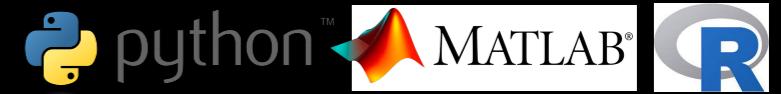
# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: MODERN SCIENTIFIC PROGRAMMING

- Highlights
  - Python (historically software dev), Matlab (historically engineering), R (historically stats)
    - “Historically”, but times are changing
    - Neuroimaging-specific libraries: Nipype, Nilearn
  - Programming assistance: GitHub copilot, ChatGPT (but be careful)
  - Containers (e.g., Docker)
- Why bother?
  - Building blocks of modern neuroscience = more rigorous projects, transferrable skills
- Fastest + easiest way to get started
  - Take on a project or project-oriented course that uses that language (no, seriously)
  - Datacamp, codecademy, etc is more for secondary support
  - Programming assistance = double-edged sword

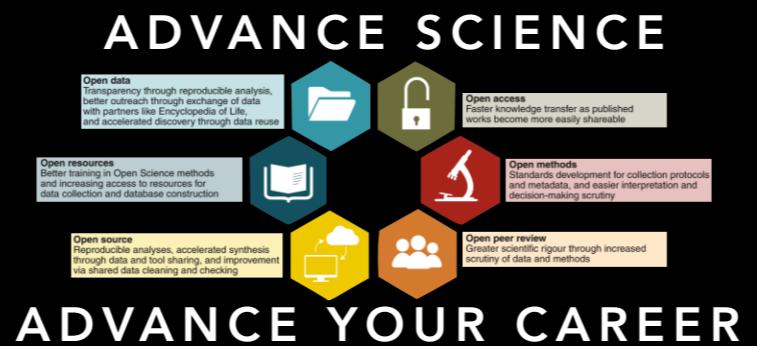


# OTHER TOPICS IN REPRODUCIBLE NEUROSCIENCE: MODERN SCIENTIFIC PROGRAMMING

- Highlights
  - Python (historically software dev), Matlab (historically engineering), R (historically stats)
    - “Historically”, but times are changing
    - Neuroimaging-specific libraries: Nipype, Nilearn
  - Programming assistance: GitHub copilot, ChatGPT (but be careful)
  - Containers (e.g., Docker)
- Why bother?
  - Building blocks of modern neuroscience = more rigorous projects, transferrable skills
- Fastest + easiest way to get started
  - Take on a project or project-oriented course that uses that language (no, seriously)
  - Datacamp, codecademy, etc is more for secondary support
  - Programming assistance = double-edged sword
- Common bottlenecks
  - Where to start?
    - Python is easiest + most flexible
      - but good to learn what your community uses most
    - Start with a real but manageable project (e.g., containers are more of an intermediate topic)

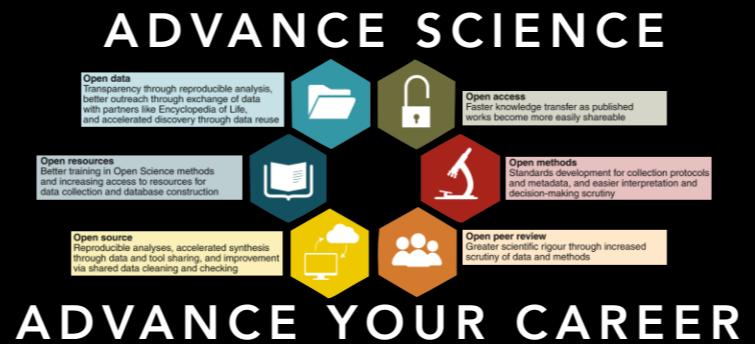


# CONCLUSION



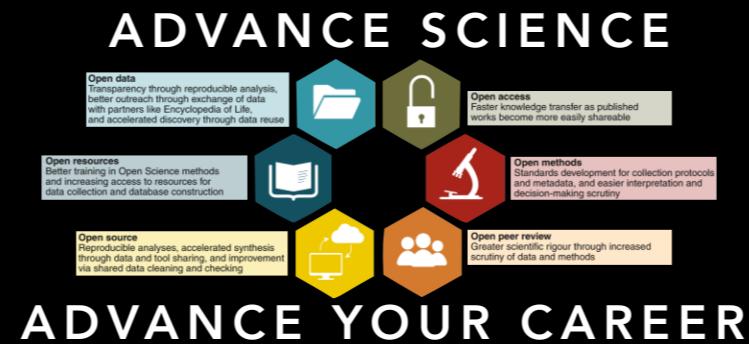
# CONCLUSION

- Open neuroscience practices not only benefit the community, but also:
  - Save you \$\$, time, & effort
  - Facilitate dissemination & recognition of your work
  - Teach you transferrable skills



# CONCLUSION

- Open neuroscience practices not only benefit the community, but also:
  - Save you \$\$, time, & effort
  - Facilitate dissemination & recognition of your work
  - Teach you transferrable skills



- Baby steps
  - **Open code:** Put one piece of code on GitHub
  - **Open data:** Get some preprocessed HCP data
  - **Open publishing:** Upload a manuscript to arXiv / OSF
  - **Open community:** Join a virtual Brainhack, OSR, or Neurohackademy event
  - **Modern scientific programming:** Follow a NiLearn tutorial

# ACKNOWLEDGMENTS

Now hiring—join the Neuro-PRISM team!

Dustin Scheinost &  
MINDS Lab (Yale)

Joshua Curtiss (MGH)

Amanda Mejia  
(Indiana)

Thomas Nichols  
(Oxford)

Todd Constable &  
Connectivity Lab (Yale)



Andrew Zalesky & Sina  
Mansour (Melbourne)

Joshua Vogelstein & Eric  
Bridgeford (Johns Hopkins)

Marisa Spann (Columbia)

Argyris Stringaris (UCL) &  
Dylan Nielson (NIMH)

Jia-Hong Gao (Peking) &  
Guoyuan Yang (SJTU)

Daniel Barron (Yale)

Chris Benjamin (Yale)

Karim Ibrahim (Yale)



NIH DSPAN F99/K00  
F99NS108557  
K00MH122372

NIH BRAIN K99/R00  
K99MH130894

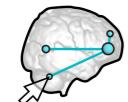


# THANK YOU! QUESTIONS?

 sNeuroble

 @sNeuroble

 sNeuroble



Biomage Suite **Web**

fast & portable image analysis

