Center for Neuroengineering & Therapeutics (CNT)

http://cnt.upenn.edu/

2023 Summer Student Data Pipelines

Penn Engineering

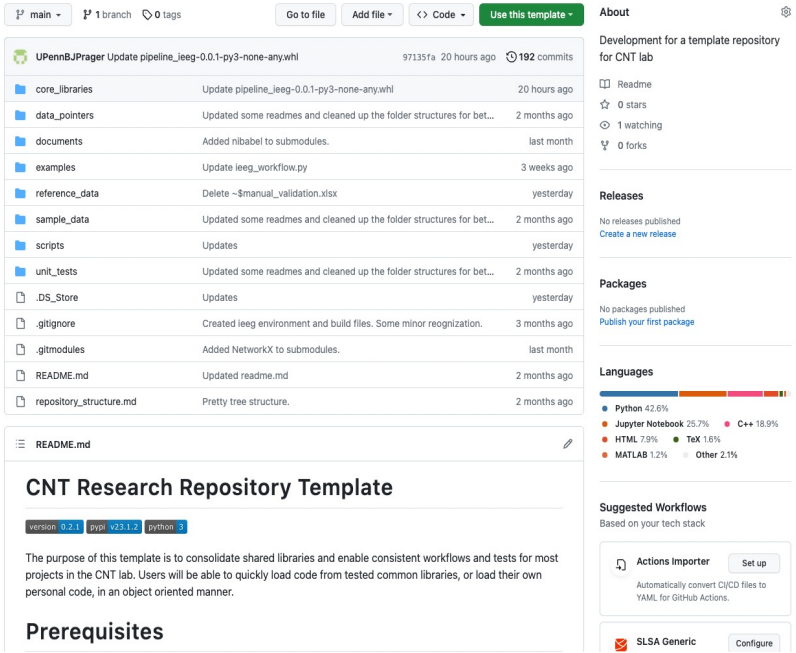
Github

“GitHub, Inc. is an Internet hosting service for software development and version control using Git. It provides the distributed version control of Git plus access control, bug tracking, software feature requests, task management, continuous integration, and wikis for every project.”

In essence, Git can provide the backbone to a successful and resilient data ecosystem.

Github

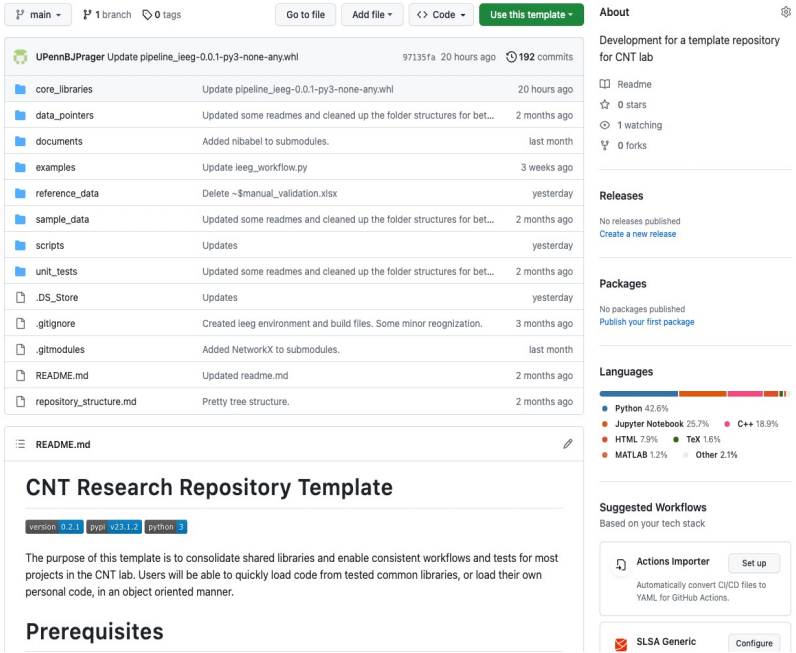
- Most advanced coding projects are stored in repositories.
- A repository contains code, documentation, libraries, etc.



Penn Engineering

Github

- Using git also acts as a file backup and versioning control.
- Multiple users can update a repository, and git will track these changes.



Penn Engineering

Github – CNT Repository

- The CNT manages a collection of repositories built up by various researches at the following URL:
  - https://github.com/penn-cnt
- We have been consolidating the various repositories into a single repository that allows us to make pipelines for research easily.
  - https://github.com/UPennBJPrager/CNT_Development
  - This is the repository you should be working with during your time here.

Github – CNT Pipeline

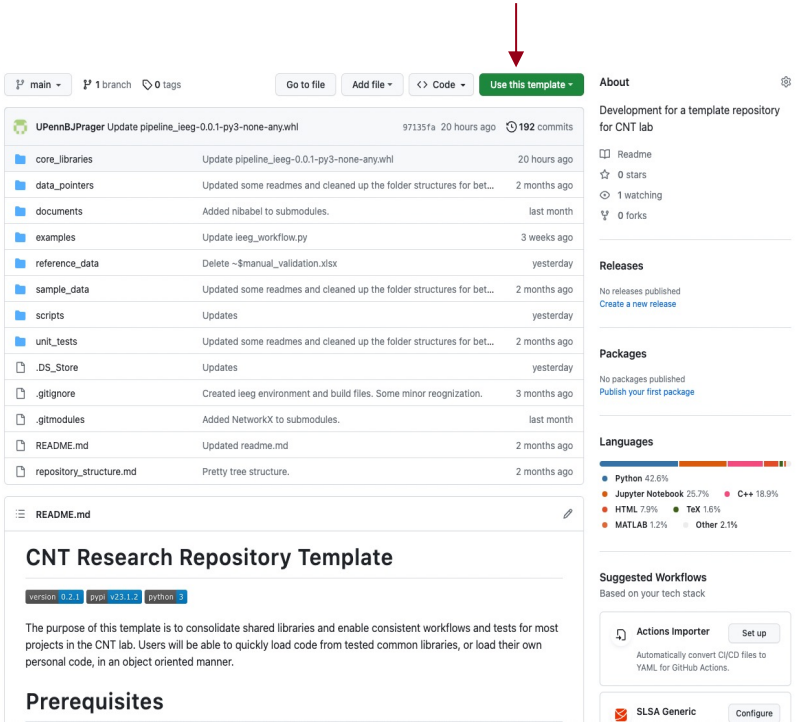
- The CNT pipeline repository is still under development, but we request you use this repository as your working environment.
- When you leave the lab, we will be able to retain and merge your work back into the lab much more readily if you work in this template.

Github – Installation

- Git can be run through a variety of clients or terminals, but we recommend using the Github desktop.
  - https://desktop.github.com/
- This client helps you download and update repositories with comments in one simple environment.
  - For more information, please visit: https://docs.github.com/en/desktop
- Once installed, please let Josh Asuncion know to be added to the CNT user group.

CNT Pipeline Installation

- Once you have a github account, click on the green "Use this template." at:
  - https://github.com/UPennBJPrager/CNT_Development
- Click on "Create a new repository" within the drop-down menu.



Penn Engineering

CNT Pipeline – Git Desktop

- Once you have linked your Git account to your github desktop, you can simply add the new repository to your desktop agent.
- In the upper left-hand corner of the app, click on the current repository for a complete list of actions.
- For more information, please visit: https://docs.github.com/en/desktop

CNT Pipeline – Layout

- These next few slides will provide a brief introduction to the current layout of the repository template and its core libraries.
- As this is a work in progress, your feedback about the organization and accessibility are greatly appreciated on behalf of the lab.

Core Libraries

- This is one of the most important folders in the repositories.
- We maintain commonly used tools for a wide variety of projects in this folder.
  - At present, we only support Python libraries.
  - Under the Python subfolder, you can find python wheels and environment files for different project types (eeg/imaging/etc.).
  - For more information about how to install these files, please refer to the README file in the repository top directory.

Data Pointers

- You will likely not need to interact with this folder often, unless you wish to manually locate a given file.
- This folder is meant to provide pathing to data we store locally to reduce time spent downloading data and to ensure we do not duplicate data needlessly.

Documents

- This folder provides documentation for various libraries and projects.
- Under the libraries subfolder, you can find a summary of documentation strings provided by code authors for the core libraries.
  - Not every code is provided with documentation. If you find yourself needing help, please reach out to your mentor.
  - These files are generated automatically, so once you have documentation, please let us know and we will work to include the documentation in future releases.

Examples

- This folder provides examples for how to use various libraries provided in this repository.
- At present, we only provide a sample eeg workflow.
  - This can be found under examples/pipeline/ieeg/iEEG_example.ipynb
- As researches complete their work, more examples will be made available.

Reference Data & Sample Data

- These two folders provide data to test ground truths against and to test the data pipeline, respectively.
- At present, we are actively encouraging researchers to provide more data for testing models and for unit tests.

Scripts

- This folder is one of the most important for your work.
- This folder is where project code should reside.
- You are free to make any subfolders, but please try to store your code and any associated files within this directory.
- If you wish to not make your code public for any reason, you have two options:
  - You can create a folder under scripts/python called ‘private’. This directory and any subdirectory will not be uploaded to the web-based repository.
  - You can edit the .gitignore file in the root directory to create new private folders.
- Note, any private data will still be merged with our lab-wide repository at the time of off-boarding.

user_data*

- This folder may not be visible on the public repository.
- This is on purpose, this folder is where your data should reside.,
- Anything within CNT_Development/user_data folder will not be uploaded by default.
- Please be aware, data with PHI or sensitive information should *NOT* go into your repository, even in this directory. Sharing is disabled to save space on our web account, not to protect patient information.

Questions?

- We appreciate your cooperation in using this template for your work.
- If you have any questions or comments, please send them to:
  - Brian Prager, bjprager@seas.upenn.edut

- Good luck with your research!

Penn Engineering