Research Data Services

Data Literacy Series, n.09/2024

GitHubbing for Open Research

GitHub repositories serve as vital platforms in today's computational research, allowing scientists to store and share their project code. Although these repositories weren't specifically designed for data storage and preservation, their collaborative features and version control capabilities significantly aid project workflows. To boost their long-term value, we recommend the following best practices:

ENHANCE YOUR REPO FINDABILITY

- Make your repo public (whenever possible)
- Name the repo in such a way that clearly reflects its purpose, enhances discoverability, and is easy to read and remember MORE INFO
- Add keywords and programming languages as tags

ENSURE FUTURE UNDERSTANDABILITY

- Establish a clear directory structure with a logical hierarchy (e.g., src/, data/, docs/) to facilitate navigation MORE INFO
- Adopt consistent file naming conventions that reflect the content MORE INFO
- Create a master README.md file to outline the project's purpose, file structure, and interrelations MORE INFO
- Include and cross-reference additional README files for specific directories if needed
- Provide project context and goals in a docs/ directory or use the wiki pages feature

STREAMLINE COLLABORATION

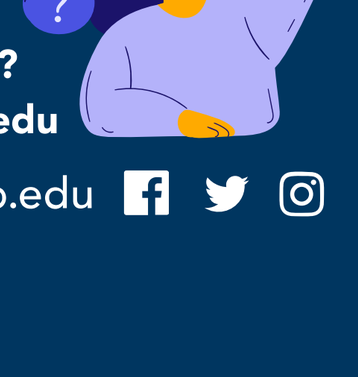
- Share expectations and preferred workflows for collaboration
- To encourage project contributors, add a contributing.md file MORE INFO
- Use labels and milestones to organize issues MORE INFO
- Consider setting up a GitHub organization for your lab or research group MORE INFO

MAXIMIZE REPRODUCIBILITY & REUSE

- Document all dependencies and consider including a requirements.txt for Python or an renv.lock file for R to ensure reproducibility and simplify environment setup MORE INFO MORE INFO
- Clearly specify a license for your code to clarify usage rights and protect your work, using common permissive options like MIT, GPL, or Apache
- Consider adding a license.md file at the root directory MORE INFO
- Indicate explicit licensing for data files with proper citation/attribution, ideally using a CC license MORE INFO
- Consider adding a citation.cff file at the root directory MORE INFO

ENRICH LONG-TERM PRESERVATION

- Certify you have opted into the GitHub Archive Program MORE INFO
- Create formal releases for completed cycles or versions to give users stable reference points MORE INFO
- Archive releases and issue DOIs for your repository using Zenodo or Figshare MORE INFO
- Consider archiving relevant project deliverables and data in certified and curated data repositories MORE INFO



UC SANTA BARBARA Library

Have questions?
rds@library.ucsb.edu

www.library.ucsb.edu