

Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution

This manuscript ([permalink](#)) was automatically generated from [SORTEE-Github-Hackathon/manuscript@1871a55](#) on July 14, 2022.

Authors

- **Robert Crystal-Ornelas**

 [0000-0002-6339-1139](#) ·  [robcrystalornelas](#) ·  [rob_c_ornelas](#)

Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

- **Brandon P.M. Edwards**

 [0000-0003-0865-3076](#) ·  [BrandonEdwards](#)

Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada

- **Katherine Hébert**

 [0000-0001-7866-6775](#) ·  [katherinehebert](#) ·  [hebert_kat](#)

Département de biologie, Université de Sherbrooke, Québec, Canada

- **Emma J. Hudgins**

 [0000-0002-8402-5111](#) ·  [emmajhudgins](#)

Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada

- **Luna L. Sánchez Reyes**

 [0000-0001-7668-2528](#) ·  [LunaSare](#) ·  [LunaSare](#)

School of Natural Sciences, University of California, Merced, USA

- **Eric R. Scott**

 [0000-0002-7430-7879](#) ·  [Aariq](#)

Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA

- **Matthew J. Grainger**

 [0000-0001-8426-6495](#) ·  [DrMattG](#) ·  [Ed_pheasant](#)

Terrestrial Biodiversity, Norwegian Institute for Nature Research - NINA, Postbox 5685 Torgarden, 7485 Trondheim, Norway

- **Vivienne Foroughirad**

 [0000-0002-8656-7440](#) ·  [yjf2](#) ·  [vforoughirad](#)

Department of Biology, Georgetown University, Washington, DC, USA

- **Allison D. Binley**

 [0000-0001-8790-9935](#) ·  [adbinley](#) ·  [AllisonBinley](#)

Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada

- **Cole B. Brookson**

 [0000-0003-1237-4096](#) ·  [colebrookson](#)

Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

- **Kaitlyn M. Gaynor**

 [0000-0002-5747-0543](#) ·  [kaitlyngaynor](#) ·  [kaitlyngaynor](#)

Departments of Zoology and Botany, University of British Columbia, Vancouver, BC, Canada; National Center for Ecological Analysis and Synthesis, Santa Barbara, CA 93101, USA

- **Saeed Shafiei Sabet**

 [0000-0001-5919-2527](#) ·  [shafieisabets](#) ·  [SaeedSHSABET](#)

Fisheries Department, Faculty of Natural Resources, University of Guilan, Sowmeh Sara, Iran

- **Ali Güncan**

 [0000-0003-1765-648X](#) ·  [Aguncan](#) ·  [aliguncan](#)

Department of Plant Protection, Faculty of Agriculture, Ordu University, 52200, Ordu, Turkey

- **Friederike Hillemann**

 [0000-0002-8992-0676](#) ·  [fhillemann](#)

Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

- **Helen Weierbach**

 [0000-0001-6348-9120](#) ·  [helenweierbach](#) ·  [HWeierbach](#)

Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

- **Dylan G. E. Gomes**

 [0000-0002-2642-3728](#) ·  [dylangomes](#)

(Current) National Academy of Sciences NRC Research Associateship Program, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA 98112; (Former) Cooperative Institute for Marine Resources Studies, Hatfield Marine Science Center, Oregon State University, Newport, OR, United States

- **Pedro Henrique Pereira Braga**

 [0000-0002-1308-1562](#) ·  [pedrohbraga](#) ·  [pedrohp_braga](#)

Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal, QC H4B 1R6, Canada

Abstract

Researchers in ecology and evolutionary biology are increasingly dependent on computational code to conduct research, and the use of efficient methods to share, reproduce, and collaborate on code as well as any research-related documentation has become fundamental. GitHub is an online, cloud-based service that can help researchers track, organize, discuss, share, and collaborate on software and other materials related to research production, including data, code for analyses, and protocols.

Despite these benefits, the use of GitHub by EEB researchers is not widespread due to the lack of domain-specific information and guidelines. To help EEB researchers adopt useful features from GitHub in their own workflows, we review twelve practical ways to use the platform. We outline features ranging from low to high technical difficulty: storing code, managing projects, coding collaboratively, conducting peer review, and writing a manuscript. Given that members of a research team may have different technical skills and responsibilities, we describe how the optimal use of GitHub features may vary among members of a research collaboration. As more ecologists and evolutionary biologists establish their workflows using GitHub, the field can continue to push the boundaries of collaborative, transparent, and open research.

Introduction

Most scientists, including ecologists and evolutionary biologists, are increasingly dependent on computational tools in their research¹. Researchers write and use software packages or data analysis code (hereafter, code) to perform scientific tasks ranging from data management, data analysis, and study replication, to the application and the development of tools for hypothesis testing. Maintaining code for scientific collaboration requires an efficient and well-documented work-flow². To facilitate this process, scientists have been adopting tools from information and systems technology, such as cloud-based services for documentation and version control (e.g., from the Google Suite, the Microsoft Suite, and GitHub³). However, most researchers lack exposure to adequate code practices and thus dedicate valuable time and effort to self-teaching research-facilitating tools. Thus, researchers may not adhere to standards of code quality and maintenance^{1,4,5}. Here, we review and discuss one of the most used web-based platforms for computational version control and collaboration, GitHub, and provide researchers in ecology and evolutionary biology (EEB) with practical workflows to facilitate and improve their code and its management.

With over 73 million registered users as of 2022, GitHub is the most widely-used web platform for collaborating on computer code⁶. GitHub provides a simple but powerful web interface that allows users to participate in projects, contribute code, report and discuss code bugs, discover existing code and data, and publish new code. Through version control, users have a detailed, chronological record of the files and directories stored in their repositories⁷ (see [Box 1](#)). This workflow provides a strong and clear advantage over sending files back-and-forth (e.g. via email), a process that can become challenging and time-consuming in long-term and collaborative projects⁸. Through its combination of version control and collaborative features, GitHub facilitates open source code alongside collaborative development⁹.

Git is the version control system that enables all the collaborative tools available on GitHub. Although the understanding of basic concepts of Git (such as commit, push, pull, checkout; see [Box 1](#)) is necessary, the GitHub web-based platform and its integrated development environments (such as the GitHub Desktop) allow users to perform most repository and data management operations without opening Git command-line sessions.

The expansive GitHub user-community and numerous GitHub resources have boosted its popularity^{7,9-11}. Nevertheless, although multiple articles have encouraged researchers in EEB to adopt GitHub as part of their research process^{3,12}, its use is still not widespread. First-time users without formal training in information technology may face steep learning curves because GitHub and its features have been centered on collaboration for code development in information systems¹³. Moreover, domain-specific resources providing tractable examples and practical guidance for researchers in EEB on GitHub are scarce (but see <https://ourcodingclub.github.io>; <https://www.openscapes.org>). Widespread adoption of GitHub for collaborating on research tasks can ultimately enable EEB researchers to spend less time on creating novel processes for collaboration and more time on their research¹⁴. More importantly, expanding the availability of data and code management standards – of which GitHub is one increasingly important component – makes research more reproducible and collaborative^{15,16}.

This paper is the result of an academic hackathon held during the 2021 conference for the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology (SORTEE, <https://www.sortee.org>). We convened a group of around thirty researchers in EEB with varying levels of familiarity with using GitHub as part of their research projects to showcase and discuss how existing features can contribute to documentation and collaboration in EEB research. During the hackathon, we identified the need for a formal discussion on how EEB researchers can benefit from GitHub and its features. Here, we outline twelve practical ways that EEB researchers can use GitHub features for more

collaborative, transparent, and reproducible science. We also provide critical perspectives on features that could be improved and catered towards research development.

Box 1

Glossary
repository: Commonly shortened to “repo”, a repository is a collection of files (e.g., a directory) tracked by Git. Repositories are managed by an owner and can be listed as “public” where the repository will be visible to all GitHub users or “private”, where the repository is visible only to authorized users. Repositories can be either “local” and saved on an individual’s computer or “remote” and stored on the cloud via GitHub’s web platform.
fork: A fork is a copy of a repository hosted on GitHub. If a repository is public, then anyone can make a fork. Even if they do not have access to push to the original repository, they can make a fork and edit it independently. Forks are linked to the original GitHub repository and “upstream” changes (<i>i.e.</i> , those in the original repository) can be merged to keep the fork up to date with the original project. Changes made in the fork can be integrated into the original project via pull requests .
clone: Cloning a repository is a way of making a local copy (<i>i.e.</i> , on your computer) of a GitHub repository . If you have access to push to a repository , this can be a first step to contributing to a project.
branch: Git workflow timelines or repositories are analogous to trees, with a main working project and diverging branches that are pointers to changes during the development process. A git branch is an alternative line of development for a project (repository). Branches allow users to add new features or modifications to the project without affecting the main part of the project. Development branches can be created at any point in time and work on each branch can continue independently. Branching is useful for testing out new ideas (both code and text) which may or may not eventually get integrated into the main branch of the project. Branches can also be used to isolate contributions of multiple contributors. Each person working on their own branch eliminates problems that may arise if conflicting edits are pushed to the same remote branch. Changes in a development branch can be merged into the main branch via pull requests . Branches can only be made by those who are given access to the project repository .
commit: Commits are snapshots of the development of a project. In Git, versions of files and directories are uniquely identified as “commits”, allowing one to identify and track modifications line-by-line. Commits can include changes in multiple files and must include a brief commit message describing the changes made. A typical workflow is to make some related changes in files, add a commit message (e.g., “generate and include fig1 in results”), and after several commits push those commits to the remote (<i>i.e.</i> , cloud-based) GitHub repository .
push/pull: When commits are made in a project locally, they must be synced with the remote GitHub repository by pushing them. Changes on a GitHub repository can then be pulled to keep your local version of the project up-to-date.
pull request: A pull request is a request that the owner(s) of a GitHub repository integrate changes you’ve made on either a branch in the repository or in your own fork . When you initiate a pull request, you must provide a description of what changes were made. Some automated tests may be run and review may be required before integrating your changes into the main branch .
merge: Combining commits from two different branches together into one branch
release: At any point a release can be made on GitHub to mark a significant milestone in the progression of a repository . While this GitHub feature is designed with releases of new versions of code in mind (e.g., v1.0.0), it can also be used to create a snapshot of a repository at significant stages like submission, revision, and acceptance of an associated manuscript.
community: A forum where GitHub users can ask for advice, offer solutions to questions, and share ideas (https://github.community/).

Twelve practical ways GitHub can accelerate research in ecology and evolution

Storing and sharing research compendia

An EEB research compendium includes all computational materials related to research production, including data, code for analyses and protocols. Having copies of these files safely stored is essential to protect against accidental modifications or deletions. Many researchers begin using GitHub as a means to store (or backup) their research compendium¹⁷ to a centralized, readily-available remote server (see [Box 1](#)). A centralized research compendium stored in a version-controlled repository has the advantages of facilitating collaboration, integrating data and code archiving services, allowing file versions to be accessed and restored, and contributes to open science (see sections below).

GitHub limits committed file sizes to 100 Mb (megabytes¹⁸), which can make it challenging for centralizing research compendia containing larger file sizes. Users may still version large files using Git Large File Storage¹⁹ text pointers, but may have to rely on external file storage alternatives (such as local or cloud-hosting).

Virtual laboratory notebooks

Laboratory notebooks help researchers track their research notes, methods, policies and protocols²⁰. Virtual laboratory notebooks can be stored in GitHub repositories and provide the benefits of simultaneous, centralized and selective access, and allows for the easy update of policies and experiment protocols²¹. Researchers have been increasingly using GitHub to maintain versions and share digital laboratory notebooks³ (e.g., <https://scheuerell-lab.github.io/lab-book>; https://github.com/HuckleyLab/how_we_work).

At least for aspects of a research project that involve writing code, a GitHub repository is a form of a laboratory notebook; when changes are made to files in a version-controlled repository, the author of those changes makes a commit ([Box 1](#)) accompanied by a description of changes. Later, the entire history of commits and their commit messages are viewable and can be audited similar to a physical laboratory notebook⁸. GitHub issues (see project management section below) can be used to prioritize laboratory objectives and goals, as well as track any status updates.

Project management

Modern research in ecology and evolution is highly collaborative, bringing together multidisciplinary teams from various institutions. On GitHub, collaborators can share feedback, brainstorm ideas, and troubleshoot problems (Figure 1). Project management can happen via three GitHub repository features: "Issues", "Discussion" and "Projects". Github Issues allow for discrete tasks and sub-tasks to be identified, assigned to team members, and categorized with custom labels. Github Discussions serve as a message board for conversation. Finally, GitHub Projects integrate issues and pull requests on automated spreadsheets and project boards, providing users with real-time tracking of project priorities and status²². Scripts, commit messages, and pull requests can be linked directly to issues, discussions, and projects providing a clear record of project workflow.

Using GitHub for all project-related conversation and planning, rather than e-mail or messaging tools, makes it easier to keep track of progress throughout the lifespan of a project. Unlike emails and messages which can get lost as more new tasks arise, GitHub issues are intentionally closed by repository administrators hiding the issue from view (closed issues remain accessible but not immediately visible). Project management in GitHub can also be integrated with third-party applications, such as Zenhub (<https://www.zenhub.com>) or Slack (<https://slack.github.com>). ZenHub allows for the enhanced visualization and organization of repositories and their issues, while the GitHub for Slack integration allows notifications from GitHub events to be sent directly to users or group channels²³.

Educational materials

GitHub supports a broad set of mechanisms for hosting educational materials. The entire process of running a course, workshop, or lecture, can all be done openly on GitHub including material development, web hosting, and delivery, and even submission and grading of assignments. While there are other purpose-built platforms for this, GitHub provides a free, open-source alternative.

Making presentations, syllabi and other course materials can be done through most major high-level programming languages such as R, with `RMarkdown` ²⁴, Python, with `python-ppt` (<https://python-pptx.readthedocs.io>), and Julia, with `Remark.jl` (<https://juliapackages.com/p/remark>), and be version-controlled and stored in GitHub. Once content is made, hosting a course website can be done through GitHub Pages²⁵ (e.g., <https://github.com/topics/course-website>). This way, course content can be available to enrolled students, as well as a global pool of learners and teachers interested in the material. Content can then be delivered via the course website, and/or a GitHub Organization with, for example, template repositories for assignments. Student submissions are perhaps the most challenging component, but the new GitHub Classroom tool (<https://classroom.github.com>) allows instructors to host private assignments to be submitted collaboratively or individually as code or PDF files, and even build autograding tests. Although time-consuming to establish, using these features can integrate learning version control and GitHub with the learning course content, and thus boost students' feelings of self-efficacy and confidence²⁶.

Hosting a website

Personal or laboratory websites can improve the sharing of research findings, build online presence, and increase coordination of research efforts²⁷. Despite many researchers in ecology and evolution having little experience in building or hosting webpages, many tools have been developed to help this process. Static websites can now be easily built using independent software and languages, such as Jekyll (<https://jekyllrb.com>), Hugo (<https://gohugo.io>), Quarto (<https://quarto.org>), and wowchemy (<https://wowchemy.com>), or with the help of dependencies in the programming languages scientists commonly use, such as the `distill` (<https://github.com/rstudio/distill>) and `blogdown` (<https://bookdown.org/yihui/blogdown>) R packages. The resulting files from static websites can then be hosted in repositories, from which one can activate GitHub's Pages (<https://pages.github.com>) feature, allowing for the direct live hosting of HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript files within a free github.io domain.

Website templates are readily available on GitHub (<https://github.com/topics/website-template>) or in user-specific repositories, which can be forked (Box 1) and customized²⁸. Aside from free hosting services, GitHub Pages also allows websites to be autogenerated and instantly updated following content modifications⁷. Creating and hosting websites on GitHub pages can be time-consuming, as this process usually requires more knowledge in web design than out-of-the-box platforms (e.g., Wix, Weebly, Google Sites). However, free hosting, widely available template customization, and versioning are strong advantages over the alternatives.

Archiving citable code and data

Government, funding agencies, and publishers exercise rigorous open access data policies and mandates^{29,30}. However, code and data sharing may be met by individual reluctance, temporary embargoes, or partially prevented by privacy and confidentiality reasons³¹⁻³³. Still, data deposition and ensuring its availability can amplify the outreach of published studies³⁴, increase citation rates³⁵, and among many other reasons, enable the reproducibility and robustness of scientific advances³⁶⁻³⁸. While public repositories on GitHub make it easy to store and share data files, they are not considered long-term repositories for research materials. This is because GitHub, a for-profit company, does not have long-term data availability guarantees, allowing users to delete or make repositories private after publication. Also, GitHub does not issue Digital Object Identifiers (DOI) for

content uploaded to their servers. DOIs are persistent and citable unique alpha-numeric identifiers assigned to digitally stored research materials. Because of this, scientists sharing code and data through GitHub are strongly encouraged to independently submit their research materials to long-term data archives (e.g., Zenodo, Figshare, Dryad, OSF^{39,39}; [Table 1](#)). Some of these options (Zenodo, Figshare and OSF) integrate with GitHub, allowing project, code, and data releases ([Box 1](#)) to be archived with versioned, citable DOIs. Linking GitHub repositories with a DOI helps research become findable, properly cited, and can ensure long-term stability⁴⁰.

An important aspect of making code and data citable and reusable is to add an appropriate license to protect intellectual property. Code published without a license is under exclusive copyright (by default), protecting it from copy, distribution, and modifications. One may grant specific rights to their code for reuse by adding licensing files and specifications within GitHub repositories⁴¹. The Choose a License (<https://choosealicense.com/non-software/>) website offers further guidance on the licenses available for research and creative products. For example, Creative Commons (CC; <https://creativecommons.org/licenses/>) licenses can specify that shared code is intended for a specific analysis. A CC BY 4.0 license specifies that any code (or other creative products) must be appropriately credited to its original author when distributed, adapted or reused.

Collaborative and asynchronous code editing

Researchers can asynchronously communicate and individually or collaboratively work in GitHub. By forking, branching, and cloning repositories ([Box 1](#)), researchers can simultaneously work on different issues of the same research project to then later merge changes to the main branch project with pull requests. Explicit project organization and increased communication in GitHub Issues, Discussions, or within pull requests can help with project development and with potential merge conflicts due to users simultaneously working on the same sections⁴². Moreover, version control⁴³ and commit reversal features allow researchers to track and progress without worrying about irreparably modifying someone else's work. By enabling more comprehensive remote collaboration, GitHub encourages the exchange of ideas among researchers at different institutions and in different countries, which can serve to improve the quality of the research itself by providing open access to data and code. In academic settings, GitHub can also facilitate interactions between research advisors and advisees, providing a platform for students or other trainees to share in-progress code, and flag specific challenges or questions for their supervisors or mentors ([Table 2](#)). Periodic code review⁴⁴ can also help advisors identify errors early in the process, and inform further training and mentorship to fill gaps in skills.

Writing a manuscript

Beyond supporting collaborative code development, GitHub can be used for writing manuscripts. Writing a manuscript and storing its associated data and code in GitHub increases scientific reproducibility because text, code, and data can be found in one place. Writing a manuscript on GitHub may take more time when compared to using conventional text processors⁸. Nevertheless, GitHub has many features that can allow for a powerful collaborative workflow when writing manuscripts. Text documents stored and versioned in GitHub can be instantly displayed when written in Markdown, a lightweight markup language increasingly popular among scientists. Co-authors can contribute changes or suggest revisions to a manuscript written in GitHub through pull requests ([Box 1](#)). Pull requests provide line-by-line views of proposed changes, which can be commented, modified, or approved by designated reviewers and collaborators ([Table 2](#)). Relevant literature or issues can be made using the Discussions and Issues features. Moreover, real-time collaboration on text documents stored in GitHub repositories can be achieved with the help of other platforms (e.g., HackMD (<https://hackmd.io>) for Markdown documents).

We wrote this manuscript using Manubot⁴⁵, a workflow implemented in GitHub to automatically render manuscripts and automate bibliographical tasks. Manubot uses GitHub's automation workflow, GitHub Actions, to combine and convert individual Markdown files into a single LaTeX document, which can then be converted to a Word or PDF document, and displayed as a webpage. Citations and bibliographic references are automatically managed with citable persistent identifiers (e.g., DOIs, PubMed IDs, ISBNs, URLs). The resulting manuscript can be rendered with document templates and citation style language formatting to meet journal formatting requirements. Every change made to the manuscript triggers its rendering, so that updates are readily displayed and made publicly available. Additional GitHub Actions can be integrated with Manubot, such as ones creating figures or generating tables (e.g., <https://github.com/SORTEE-Github-Hackathon/manuscript/tree/main/.github/workflows>). For more examples of manuscripts built via Manubot, see⁴⁵.

Peer review

Peer review is the standard process for assessing and judging whether research done in ecology and evolution should be published in a scientific journal. GitHub provides an open and transparent platform that can be used for either directly providing feedback on research products or addressing changes recommended by reviewers. GitHub Issues can be used to organize and discuss reviewer suggestions and to assign them to co-authors (e.g., <https://github.com/SORTEE-Github-Hackathon/manuscript/issues?q=label%3A%22Reviewer+Comment%22+>). When reviewer comments are posted as separate issues, authors can comment on the issues to discuss possible changes and assign themselves to indicate which comments they intend to handle. Co-authors can then integrate their edits and responses to reviewers using pull requests.

GitHub can also assist reviewers during the peer review process. If the code associated with a manuscript is made available at the time of submission (e.g., as a link to a GitHub repository within the Data Availability Statement), peer reviewers may be able to offer more comprehensive suggestions on the code and written materials, potentially recognizing errors before publication. Certain journals or software development communities require submitted work or research code to be hosted on GitHub and their review processes make use of GitHub Issues (e.g., rOpenSci (<https://ropensci.org/software-review/>), Journal of Open Source Software (<https://joss.readthedocs.io/en/latest/submitting.html>)). rOpenSci's efforts have resulted in many well-used R packages for ecology research including `rfishbase`⁴⁶ and `taxize`⁴⁷.

Open science discussion

Scientific publications often omit part of their intellectual and computational workflows, including the treatment of raw data and analytical steps (e.g., model assumption testing). Publishing data and reproducible workflows along with manuscripts can provide readers with all details about analytical steps and enable reproducing research experiments and results⁴⁸. In addition to storing data and code, GitHub repositories can provide a time-stamped (version controlled) preregistration of research plans and hypotheses.

Conventional research practices typically separate tasks among collaborators (*i.e.*, data entry, analysis, writing). It is common that coauthors discuss, but do not actively verify, edit, or execute research tasks that are not their main responsibility. GitHub can serve as a tool for open and tractable research development. Collaborators can directly interact with code and data, inspect for errors, and potentially identify scientific misconduct prior to manuscript submission (e.g.,⁴⁹). Collaborators and readers are better positioned to discover erroneous or questionable findings if they have complete and transparent access to projects.

This transparency can be extended beyond coauthors to the entire scientific community and to the public. Supplying code for (novel) methods that are proposed or used reduces barriers to knowledge, improving the ability of others to build on existing work. This results in greater proliferation and accessibility for a broader audience. Projects can make use of GitHub Discussions (<https://docs.github.com/en/discussions>) to communicate among repository members (collaborators) and to engage with other scientists and the general public. Moreover, researchers can also use the GitHub Community (<https://github.community/>) forum to share expertise or request help from others on their analyses and ideas ([Table 2](#)).

Project continuity

Projects in ecology and evolution often involve graduate students, research assistants, and post-doctoral fellows, who hold limited-term positions⁵⁰. Without clear plans on project continuity, the research code and data management upkeep tends to fall off as researchers move on to new projects or to other institutions. Additionally, code and data can be difficult to access when kept only on personal devices⁵¹.

GitHub can facilitate project continuity among research code by making code handover easier^{8,50}. Through version control, the history of code and data from projects in ecology and evolution can be tracked and accessible to future laboratory members and collaborators¹². Repositories and organizations can have designated data and code owners (or more appropriate, stewards; see^{40,52}), who can also change through time allowing for the transition of code from one cohort of researchers to the next. Other project collaborators can offer many contributions to repository design and development, and their active involvement can both aid authors ability to act as guarantors of the project, and the clarity and reproducibility of the project for future users. In ([Figure 1](#)), we highlight several elements of good repository structure, and the various ways that contributors may interact with them.

GitHub also allows users to describe and store information about software and code dependencies, ensuring that code can run using the same version of software as when it was initially developed. This can be achieved through the access of GitHub repositories containing the necessary releases of such software, or through the usage of actions with specific software versions in its automation tool, GitHub Actions.

GitHub Organizations

GitHub Organizations are shared virtual spaces that allow teams to work in different repositories, while remaining tied together under a larger group, such as a laboratory, department, or project involving several teams. Organizations are well-suited to ensure larger projects with many steps or moving parts are constrained to one virtual space, where outputs and sub-projects can be easily accessed and located without relying on any one individual. Because the repositories are grouped, members can reference and contribute to each other's work without necessarily being part of the same repository, broadening the accessibility and longevity of code and writing contributions.

Contributors can be assembled into teams within an organization, which allows administrators to assign roles, tasks, and repository modification permissions to organization members. Whereas access to repositories is usually assigned to individual contributors, Organizations facilitate the management of access permissions by allowing teams to be granted access to specific repositories. This ensures repositories with sensitive information remain as restricted as needed, while others stay open and accessible to selected member groups. The organization structure also allows for easy issue tracking and discussions related to research content and progress.

As an example, GitHub Organizations are particularly well-suited to house documents and projects within a laboratory, such as research compendia, codes of conduct, protocols, training documents, and other such documents that evolve collaboratively over time and are relevant to many colleagues. In this way, teams can have full ownership of repositories within an organization, while ensuring that these materials stay accessible to the laboratory after people have moved on or lost their copies of project data. This application extends to research centres, which may include several distinct projects that remain linked to institutions, such as the German Centre for Integrative Biodiversity Research (iDiv, <https://github.com/idiv-biodiversity>). The team organizing the hackathon which inspired this paper used a GitHub Organization (SORTEE-Github-Hackathon, <https://github.com/SORTEE-Github-Hackathon>) to keeping track of all materials as ideas developed and took shape in one virtual space. Organizations are also convenient for hosting learning materials, including lectures or workshops, such as the Québec Centre for Biodiversity Science R Workshop Series (QCBSRworkshops, <https://github.com/QCBSRworkshops>) or the University of Edinburgh's Coding Club (Coding Club, <https://github.com/ourcodingclub>), which may be updated by an ever-evolving group of contributors over time.

Discussion

The promise of GitHub for EEB researchers

There have been many calls for researchers outside of the software development community to adopt GitHub to improve their collaborative research^{3,39,53}. This call comes in light of both the continuous shift towards open-science and the increasing computational and data requirements in EEB. Until now, resources and practical guidance on using GitHub within the EEB community have been dispersed in blog posts and video tutorials ([Box 2](#)). These resources have been extremely useful for learning to use GitHub in our own work. We also feel that a collation of the main uses of GitHub in EEB into one medium, while adding on our personal perspectives, will be useful to the EEB community.

In this paper, we described twelve ways that EEB researchers can leverage GitHub to enable more transparent and collaborative research (Figure [2](#)). Many of the examples are specifically meant for first-time GitHub users and can likely be adopted with just a little practice (storing data, creating virtual notebooks, making code citable). For example, storing code and data and making it citable involve creating a repository on GitHub, pushing code to the repository, and then going through the necessary steps (*e.g.*, connecting a GitHub repository with Zenodo; see below) for creating a DOI. These actions are often covered in GitHub tutorials and require little effort to implement. On the other hand, some examples we described here, including course material development, web hosting, and delivery, may require a greater time commitment, but have the potential to make EEB research more open, accessible, and collaborative. Managing full research projects or laboratories on GitHub requires careful thought as to how to delegate tasks such as reviewing pull requests or creating issues. For example, collaboratively writing a paper using GitHub, as we have done here, may involve a learning curve for co-authors less familiar with the intricacies of GitHub, and requires overhead to set up the repository using GitHub Actions. Despite the many potential applications of GitHub to EEB research, we acknowledge that there will still be many times when researchers might look to other platforms for research collaboration.

Other platforms for collaboration

Despite its strong collaborative potential, we describe two use cases where GitHub's features fall short of highly collaborative work emblematic of EEB research.

First, real-time document editing is still best performed on other platforms (*e.g.*, cloud-stored documents from Microsoft Word, Google Docs, hackMD (<https://hackmd.io/>)). Second, operations that

are dependent on other software might not be easily accomplished in GitHub, such as manipulating figures or tables. Although creating tables and figures can be achieved through code, users may choose other software to collaboratively brainstorm figures and tables (e.g. Google Slides, Google Sheets).

Why aren't more EEB researchers using GitHub?

Although GitHub has been available as a platform for more than a decade, its uptake among EEB researchers, especially as a tool for collaboration, has been slow. Here, we discuss four potential barriers to GitHub use in EEB.

First, there may be hesitation to independently adopting and learning a new tool. Institutional encouragement and instructional resources focused on researchers in ecology and evolution may be limited. When GitHub is taught in an EEB context, it is usually accompanying coursework in topics such as statistical programming. It can be challenging to learn Git alongside scripting languages, statistical theory, and file system navigation, especially when many may be inexperienced with programming. Instructors likewise may confuse the expected digital literacy of students with computational fluency, even when modern technology increasingly abstracts concepts through search optimization and preponderant integrated development environments (IDE), or 'point-and-click' user interfaces.

Second, while many EEB researchers take advantage of GitHub for individual use, collaborative use may lag due to how researchers traditionally divide labor within projects. Despite broad utility, GitHub remains a tool predominantly used by computer scientists and software developers. EEB researchers may take the view that GitHub is a platform that only needs to be used by individuals writing code, and may silo those aspects of projects to a single individual. Those assumptions may obscure the utility of GitHub for tasks other than traditional data analysis and code development. However, we emphasize that there are opportunities for collaboration using GitHub by researchers of all skill levels or time constraints (Table 2); for example, project stakeholders can provide a list of use-cases or highlight important conceptual components of a project using GitHub's issues or discussions features.

A third barrier may come from general reluctance to share data and code publicly, or technical and logistical issues¹⁶. GitHub is, by default, a public and open platform. This openness may add additional pressure to students and scientists learning to use the platform. We also note that large file storage is discouraged (and limited) on GitHub but add-ons do exist (e.g., <https://git-lfs.github.com/>) that permit data storage, and increasing integration between platforms (e.g.,⁵⁴) allows data to be stored away from GitHub and linked dynamically. We suspect a major additional barrier to EEB researchers is a distinct lack of GitHub help documents for non-English researchers in ecology and evolution, meaning that EEB researchers potentially miss the opportunity to fully understand the importance of version control, reproducibility, and other benefits of GitHub.

Lastly, GitHub has both free and paid plans. When projects get highly collaborative, they may have to pay for additional GitHub support. The acquisition of GitHub by Microsoft has led to some concerns over the future of free plans and several biodiversity data managers have begun to switch to Open-Source Git services (e.g., Bitbucket and GitLab). At this point however, there is little practical difference for EEB researchers between the paid and free GitHub plans.

Box 2

Ten tips for getting started in GitHub

Ten tips for getting started in GitHub

1. **Check for an existing solution to your problem.** The GitHub Help webpage (<https://docs.github.com/en>) contains extensive and detailed documents with helpful screenshots. It is a good starting point for handling an issue, and has troubleshooting tips for specific problems. Alternatively, consider Tweeting your issue. There is a large community of GitHub users around the world who have likely faced analogous problems and may be able to provide quick solutions. Third, try to follow blogs (e.g., <https://github.blog>), Twitter accounts or YouTube channels that regularly post practical solutions about the most widely-used web platform for common GitHub issues.
2. **Consider taking free courses.** such as those from Software Carpentry⁵⁵ and sharing these courses with your lab members or colleagues.
3. **Take advantage of GitHub as an asynchronous working tool for team-based projects.** See the repository for this paper (<https://github.com/SORTEE-Github-Hackathon/manuscript>) as an example of a collaborative manuscript that includes discussions, issues, and a website via GitHub.
4. **The GitHub Learning Lab** (<https://lab.github.com>). allows you to learn GitHub basics through short projects and tasks, and allows you to get feedback from their Learning Lab bot.
5. **Check out the following markdown cheatsheet** (<http://markdownguide.org/basic-syntax>). so that you can write clear metadata README files for your repositories.
6. **The Jenny Bryan universe of GitHub material.** provides a thorough and accessible introduction for a multitude of research-related uses for GitHub, and includes a book⁵⁶, statistics course⁵⁷ and academic article⁷.
7. **Don't be afraid of trial-and-error.** One of the best ways to learn GitHub is the trial-and-error method. Learning from your own mistakes can be the better way to master your GitHub abilities. In any case, GitHub has the advantage of making it easy to go back to any steps that you desire via version controlling if you make mistakes.
8. **If you are an educator, include lectures on reproducibility and tools for creating reproducible workflows in the curricula.** Some graduate programs include coursework on course Rmarkdown and GitHub. Getting students started with these tools earlier will prevent the resistance that comes from working with a less reproducible workflow for a longer period of time. (see example https://github.com/rmcelreath/stat_rethinking_2022)
9. **Try to begin committing with GUI (Graphical user interface) tools.** like GitHub Desktop (<https://desktop.github.com>), git-gui (<https://git-scm.com/docs/git-gui>), RStudio (<https://www.rstudio.com>), Visual Studio Code (<https://code.visualstudio.com>), Atom (<https://atom.io>), GitKraken (<https://www.gitkraken.com>) tools instead of CLI (Command line interface) tools such as Terminal or Console for more advanced features.
10. **Get help deciphering GitHub Notifications.** Try using tools like Octobox (<https://octobox.io>) to disentangle and manage multiple notifications from distinct GitHub projects.

Conclusion

We provide twelve practical ways that ecologists and evolutionary biologists can adopt GitHub to improve their research workflow. We provide definitions (Box 1) and types of users (Figure 1) to help researchers identify and prioritize the skills and tools to learn first. We highlight tools providing high collaborative potential (e.g., open science discussion, collaborative code editing) to more individual focused (e.g., storing code and data, building a website). We argue that the tools readily available in GitHub have the potential to make ecology and evolution more open, reproducible and transparent. With this comprehensive review of how EEB researchers can use GitHub, we encourage researchers at any career stage to adopt GitHub as a platform for sharing and collaboration.

Author Contributions

We indicate author contributions using the [CRedit Taxonomy](#).

- Conceptualization: RCO, BPME, KH, EJH, LLSR, PHPB
- Investigation: RCO, BPME, KH, EJH, LLSR, ERS, MJG, VF, ADB, CBB, KMG, SSHS, AG, FH, HW, DGEG, PHPB

- Methodology: RCO, BPME, KH, EJH, LLSR, PHPB
- Project administration: RCO
- Software: RCO, BPME, KH, EJH, LLSR, ERS, MJG, VF, ADB, CBB, KMG, SSHS, AG, FH, HW, DGEG, PHPB
- Visualization: RCO, BPME, KH, EJH, LLSR, ERS, MJG, VF, ADB, CBB, KMG, SSHS, AG, FH, HW, DGEG, PHPB
- Writing – original draft: RCO, BPME, KH, EJH, LLSR, ERS, MJG, VF, ADB, CBB, KMG, SSHS, AG, FH, HW, DGEG, PHPB
- Writing – review and editing: RCO, BPME, KH, EJH, LLSR, ERS, MJG, VF, ADB, CBB, KMG, SSHS, AG, FH, HW, DGEG, PHPB
- Resources: PHPB

Acknowledgements

This manuscript arose from a hackathon at the Society for Open, Reliable, and Transparent Ecology and Evolution (SORTEE) virtual meeting in July 2021. We thank Ciera Martinez for being a co-organizer of the SORTEE hackathon that started our discussion on GitHub in EEB and for creating our hackathon website on GitHub.

RCO was funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Earth and Environmental Sciences Division, Data Management program under contract number DE-AC02-05CH11231.

Code and data availability

The source code and data for this manuscript are available at <https://github.com/SORTEE-Github-Hackathon/manuscript>. The source code will be archived in the ESS-DIVE data repository prior to publication.

Figures

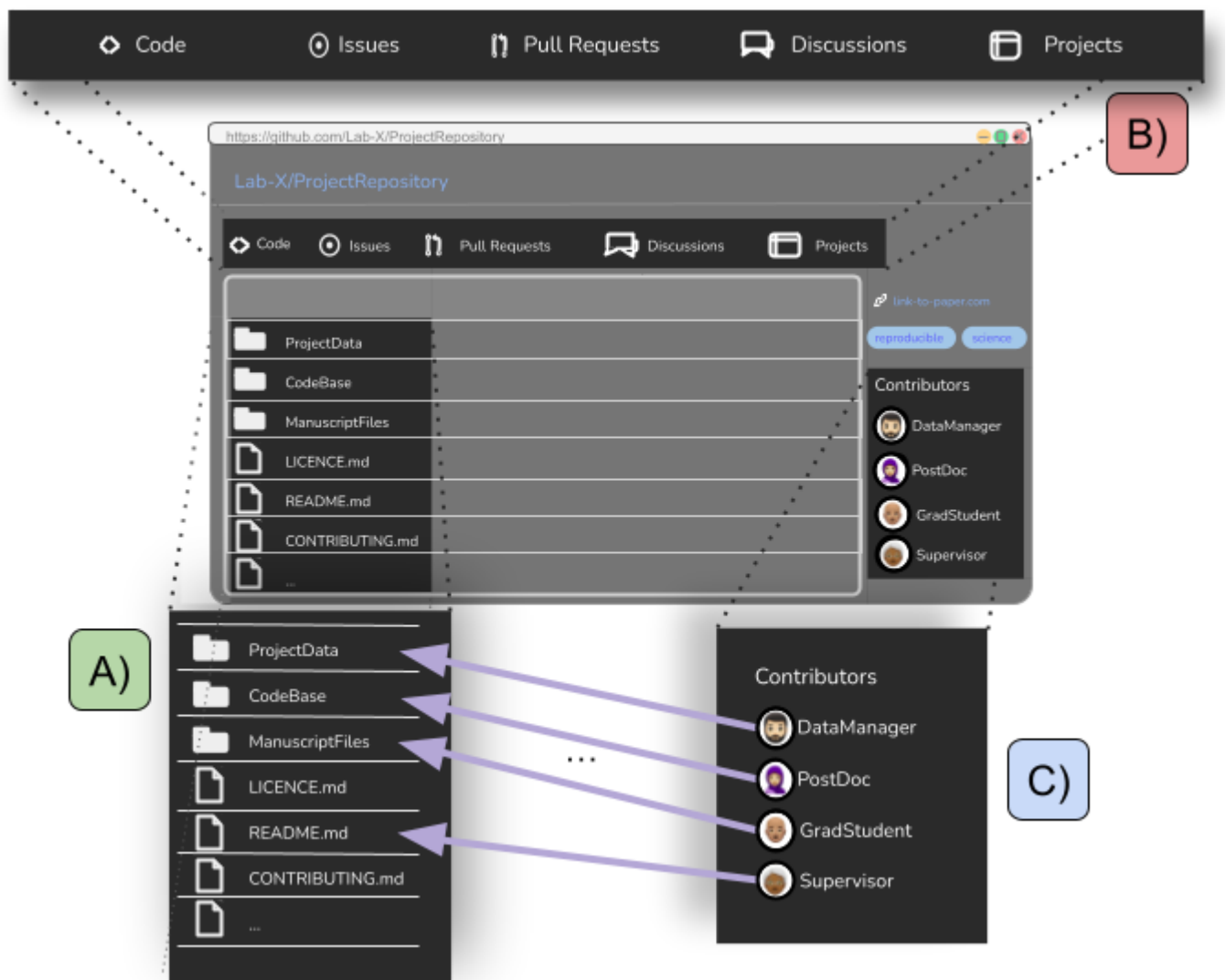


Figure 1: An overview of Git's core features. A) Multi-faceted components allow for code writing, small data storage, manuscript writing, and project management to all be done in one place. `CONTRIBUTING.md`, `LICENCE.md`, and `README.md` files allow new team members, or others wanting to use materials, to understand the project components and learn how they can engage with the project and existing team members. B) Issues, Pull Requests, Discussions, and Projects allow for team members to ask for feedback, suggest fixes, discuss related ideas, and keep track of all the moving parts of a project. C) All collaborators on a project can be a part of a single repository, with varying push privileges and responsibilities

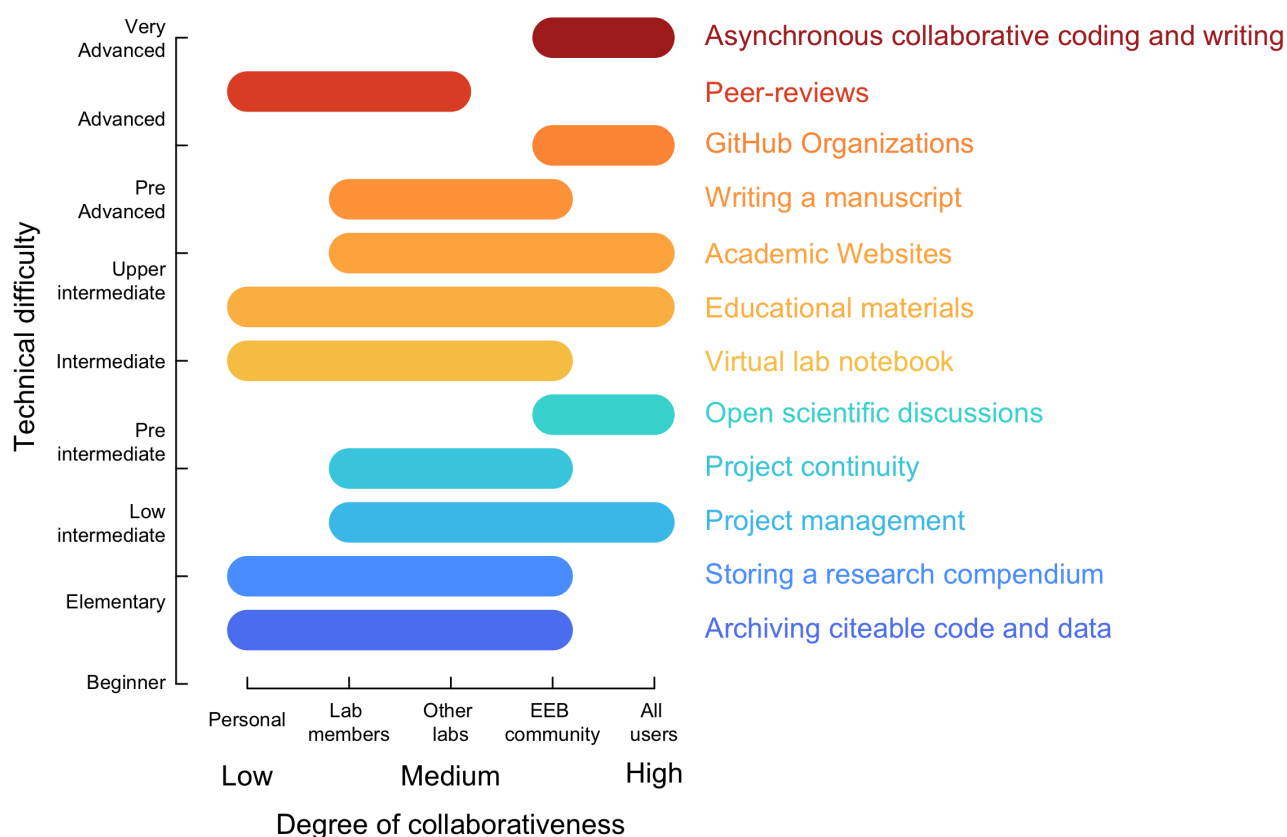


Figure 2: A summary of ways GitHub can be used showing technical difficulty and degree of collaboration for each. Activities higher on the vertical axis require usage knowledge of more GitHub features than activities lower on the axis. On the horizontal axis, each activity spans a region representing who is potentially involved with or benefits from each activity. For example, storing data and code mainly benefits individual researchers or members of a lab group while making data and code citable and reproducible benefit other labs and the larger community as well. Independently of a users knowledge level of GitHub features, there are ways to use GitHub that allow tapping unto one of the most salient benefits of the platform: facilitating and enhancing collaboration.

Tables

Table 1: a comparison of technologies commonly used for collaborating on research in Ecology and Evolutionary Biology. In the first column, we group platforms for collaboration into broad guilds. The second column lists the platform for collaboration. The remaining columns indicate whether the platform for collaboration includes certain features.

Guid	Software	Versions	Basic (controls)	Passive - time collaboration	Active real - collaboration	Free \$	Permanent (DOI)	Storage limits	GitHub Integration
Multi-tool	Git Hub	yes	yes	yes	NA	Broadly limited free version. Advanced features are provided for free to students and education professionals.	A DOI can only be obtained when integrating to other services that can mint DOI (e.g. Zenodo, OSF).	100 MB per file, 500 MB per private repository (2 GB for paid accounts). 100 GB for public repositories. Larger files (up to 2 GB) can be attached to releases	NA
Multi-tool	OSF	yes	yes	yes	yes	yes	yes	25 GB for private projects, up to 5GB per file, plus partner add-ons, 50GB for public projects	yes
Long-term (public) data repositories	PAN GAEA	yes	yes	yes	NA	yes	yes	10 GB free	NA
Long-term (public) data repositories	Zenodo	after publication	after publication	NA	NA	yes	yes	50 GB per dataset	yes

Long - term (public) data repositories	Dryad	after published	after published	NA	NA	some journals cover cost	yes	300 GB per publication	Can link to individual files (not entire repository); not really integrated
Long - term (public) data repositories	Figs hare	yes	yes	yes	NA	yes	yes	20 GB free, up to 5 TB	yes
Temporary (personal) drive storage	Google Drive	yes	yes	yes	yes	limited free version & paid	NA	15 GB free, up to 100 GB with Google One	yes
Temporary (personal) drive storage	Box	limited	yes	?	?	NA	NA	Unlimited total size for subscription	yes
Temporary (personal) drive storage	DropBox	limited	yes	yes	yes	limited free version & paid	NA	2 GB free	yes
Temporary (personal) drive storage	One Drive and the Office Suite	yes	yes	yes	yes	limited free version & paid	NA	5 GB free, up to 1TB paid	yes

Collaborative code/text editors	Overleaf (online LaTeX editor)	yes	yes	yes	NA	NA	NA	1MB for individual .tex, 50 MB for individual files, unlimited project size	yes
Collaborative code/text editors	Jupyter Notebook	yes	?	yes	with Colab	yes	NA	via Binder: no hard limit, but suggests no files >100 MB, can also store on GitHub or Google Colab	yes
Collaborative code/text editors	HackMD	yes	yes	yes	yes	yes	NA	3 documents free, private invitee limits	yes

Table 2: A non-exhaustive collection of ideas for how various GitHub features could be utilized for a research project. Here we have categorized contributors/collaborators into five roles. A Project Manager owns the GitHub repository for a project, and leads the academic project (e.g., lead author of a manuscript). A co-author contributes to writing and other aspects of research, but may have limited or no experience with programming, git, and/or GitHub. A code contributor writes or edits analysis code for the project. A code reviewer could be a project collaborator or a peer reviewer who reviews project code. They are familiar with coding, but not necessarily with git or GitHub (but they are willing to learn). Finally, community members could be other researchers or non-researchers interested in reproducing results, re-using code or data, or communicating with researchers involved in the project. These roles are not mutually exclusive—a co-author could also be a code contributor and code reviewer, for example. For definitions of the GitHub features, see Box 1.

Role	GitHub repository	README	Issue	Discussion	Pull Request	Fork	GitHub Pages
Project manager	Set contributor permissions, share code of conduct	Project description, citation, DOIs	Assign tasks to collaborators	Discuss project directions and goals	Approve and incorporate edits to code and/or writing		Share up-to-date reports, figures, or draft manuscript
Co-author	Edit Markdown text or add files		Propose changes involving code (e.g. analyses, figures)	Discuss proposed changes to manuscript			
Code contributor			Suggest code changes		Contribute changes to code, initiate code review		Contribute to project website

Role	GitHub repository	README	Issue	Discussion	Pull Request	Fork	GitHub Pages
Code reviewer	Find all code related to a project		Highlight specific lines of code and make suggestions		Review or recommended changes in code		
Community			Suggest additional features and report bugs	Ask questions about data and code		Create a linked, editable copy of the repository	View project website

References

1. Hannay, J. E. *et al.* How do scientists develop and use scientific software? in *2009 ICSE Workshop on Software Engineering for Computational Science and Engineering* (IEEE, 2009). doi:[10.1109/secse.2009.5069155](https://doi.org/10.1109/secse.2009.5069155).
2. Perkel, J. M. [Challenge to scientists: does your ten-year-old code still run?](#) *Nature* **584**, 656–658 (2020).
3. Perkel, J. [Democratic databases: science on GitHub](#). *Nature* **538**, 127–128 (2016).
4. Prabhu, P. *et al.* A survey of the practice of computational science. in *State of the Practice Reports on - SC '11* (ACM Press, 2011). doi:[10.1145/2063348.2063374](https://doi.org/10.1145/2063348.2063374).
5. Wilson, G. *et al.* [Best Practices for Scientific Computing](#). *PLoS Biol* **12**, e1001745 (2014).
6. Build software better, together. *GitHub* <https://github.com>.
7. Bryan, J. [Excuse Me, Do You Have a Moment to Talk About Version Control?](#) *The American Statistician* **72**, 20–27 (2018).
8. Ram, K. [Git can facilitate greater reproducibility and increased transparency in science](#). *Source Code Biol Med* **8**, (2013).
9. Perez-Riverol, Y. *et al.* [Ten Simple Rules for Taking Advantage of Git and GitHub](#). *PLoS Comput Biol* **12**, e1004947 (2016).
10. Hester, J. B., the STAT 545 TAs, Jim. [Let's Git started / Happy Git and GitHub for the user](#).
11. Coding Club: A Positive Peer-Learning Community. <https://ourcodingclub.github.io/>.
12. Lowndes, J. S. S. *et al.* [Our path to better science in less time using open data science tools](#). *Nat Ecol Evol* **1**, (2017).
13. Leibzon, W. Social network of software development at GitHub. in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, 2016). doi:[10.1109/asonam.2016.7752419](https://doi.org/10.1109/asonam.2016.7752419).
14. Briney, K., Coates, H. & Goben, A. [Foundational Practices of Research Data Management](#). *RIO* **6**, (2020).
15. Alston, J. M. & Rick, J. A. [A Beginner's Guide to Conducting Reproducible Research](#). *Bull. Ecol. Soc. Am.* **102**, (2021).
16. Gomes, D. G. E. *et al.* Why don't we share data and code? Perceived barriers and benefits to public archiving practices. (2022) doi:[10.31222/osf.io/gaj43](https://doi.org/10.31222/osf.io/gaj43).
17. Marwick, B., Boettiger, C. & Mullen, L. [Packaging Data Analytical Work Reproducibly Using R \(and Friends\)](#). *The American Statistician* **72**, 80–88 (2018).
18. About large files on GitHub. *GitHub Docs* <https://ghdocs-prod.azurewebsites.net/en/repositories/working-with-files/managing-large-files/about-large-files-on-github>.
19. Git Large File Storage. *Git Large File Storage* <https://git-lfs.github.com/>.

20. Kanza, S. *et al.* [Electronic lab notebooks: can they replace paper?](#) *J Cheminform* **9**, (2017).
21. Schnell, S. [Ten Simple Rules for a Computational Biologist's Laboratory Notebook](#). *PLoS Comput Biol* **11**, e1004385 (2015).
22. About projects (beta). *GitHub Docs* <https://ghdocs-prod.azurewebsites.net/en/issues/trying-out-the-new-projects-experience/about-projects>.
23. Slack. GitHub for Slack. *Slack Help Center* <https://slack.com/help/articles/232289568-GitHub-for-Slack>.
24. Xie, Y., Allaire, J. J. & Grolemond, G. *R Markdown: the definitive guide*. (CRC Press, Taylor and Francis Group, 2019).
25. Quickstart for GitHub Pages. *GitHub Docs* <https://ghdocs-prod.azurewebsites.net/en/pages/quickstart>.
26. Trujillo, G. & Tanner, K. D. [Considering the Role of Affect in Learning: Monitoring Students' Self-Efficacy, Sense of Belonging, and Science Identity](#). *LSE* **13**, 6–15 (2014).
27. Smaglik, P. [Creating better lab websites gives potential collaborators and recruiters a clearer window into your world](#). *Nature* **447**, 347–347 (2007).
28. Dawson, C. *Building tools with GitHub: customize your workflow*. (O'Reilly, 2016).
29. Tenopir, C. *et al.* [Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide](#). *PLoS ONE* **15**, e0229003 (2020).
30. Nugroho, R. P., Zuiderwijk, A., Janssen, M. & de Jong, M. [A comparison of national open data policies: lessons learned](#). *Transforming Government: People, Process and Policy* **9**, 286–308 (2015).
31. Wicherts, J. M., Bakker, M. & Molenaar, D. [Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results](#). *PLoS ONE* **6**, e26828 (2011).
32. Figueiredo, A. S. [Data Sharing: Convert Challenges into Opportunities](#). *Front. Public Health* **5**, (2017).
33. Tenopir, C. *et al.* [Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide](#). *PLoS ONE* **10**, e0134826 (2015).
34. Pronk, T. E., Wiersma, P. H., van Weerden, A. & Schieving, F. [A game theoretic analysis of research data sharing](#). *PeerJ* **3**, e1242 (2015).
35. Piwowar, H. A., Day, R. S. & Fridsma, D. B. [Sharing Detailed Research Data Is Associated with Increased Citation Rate](#). *PLoS ONE* **2**, e308 (2007).
36. [On data availability, reproducibility and reuse](#). *Nat Cell Biol* **19**, 259–259 (2017).
37. Mislán, K. A. S., Heer, J. M. & White, E. P. [Elevating The Status of Code in Ecology](#). *Trends in Ecology & Evolution* **31**, 4–7 (2016).
38. Baker, M. [1,500 scientists lift the lid on reproducibility](#). *Nature* **533**, 452–454 (2016).
39. Crystal-Ornelas, R. *et al.* [A Guide to Using GitHub for Developing and Versioning Data Standards and Reporting Formats](#). *Earth Space Sci* **8**, (2021).

40. Hampton, S. E. *et al.* [The Tao of open science for ecology](#). *Ecosphere* **6**, art120 (2015).
41. Adding a license to a repository. *GitHub Docs* <https://ghdocs-prod.azurewebsites.net/en/communities/setting-up-your-project-for-healthy-contributions/adding-a-license-to-a-repository>.
42. Vale, G., Schmid, A., Santos, A. R., de Almeida, E. S. & Apel, S. [On the relation between Github communication activity and merge conflicts](#). *Empir Software Eng* **25**, 402–433 (2019).
43. Crystal-Ornelas, R. *et al.* *Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution*. <https://SORTEE-Github-Hackathon.github.io/manuscript/> (2022).
44. Song, X., Goldstein, S. C. & Sakr, M. Using Peer Code Review as an Educational Tool. in *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (ACM, 2020). doi:[10.1145/3341525.3387370](https://doi.org/10.1145/3341525.3387370).
45. Himmelstein, D. S. *et al.* [Open collaborative writing with Manubot](#). *PLoS Comput Biol* **15**, e1007128 (2019).
46. Boettiger, C., Lang, D. T. & Wainwright, P. C. [rfishbase: exploring, manipulating and visualizing FishBase data from R](#). *Journal of Fish Biology* **81**, 2030–2039 (2012).
47. Chamberlain, S. A. & Szöcs, E. [taxize: taxonomic search and retrieval in R](#). *F1000Res* **2**, 191 (2013).
48. Culina, A., van den Berg, I., Evans, S. & Sánchez-Tójar, A. [Low availability of code in ecology: A call for urgent action](#). *PLoS Biol* **18**, e3000763 (2020).
49. January 18, P. L. B. K. & Pm, 2022. 2:51. #PruittData and the Ethics of Data in Science. *Ecology for the Masses* <https://ecologyforthemasses.com/2020/02/04/pruittdata-and-the-ethics-of-data-in-science/> (2020).
50. Fehr, J., Himpe, C., Rave, S. & Saak, J. [Sustainable Research Software Hand-Over](#). *JORS* **9**, 5 (2021).
51. Vines, Timothy H. *et al.* [The Availability of Research Data Declines Rapidly with Article Age](#). *Current Biology* **24**, 94–97 (2014).
52. About code owners. *GitHub Docs* <https://ghdocs-prod.azurewebsites.net/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-code-owners>.
53. Anbaroğlu, B. [A collaborative GIS programming course using GitHub Classroom](#). *Transactions in GIS* **25**, 3132–3158 (2021).
54. Connect GitHub to a Project - OSF Support. <https://help.osf.io/article/211-connect-github-to-a-project>.
55. Madicken Munk *et al.* swcarpentry/git-novice: Software Carpentry: Version Control with Git, June 2019. at <https://doi.org/10.5281/zenodo.3264950> (2019).
56. Hester, J. B., the STAT 545 TAs, Jim. [Let's Git started / Happy Git and GitHub for the useR](#).
57. Bryan, J. & TAs, T. S. 545. [STAT 545](#).