Open and Reproducible Research on Open Science Framework

Ian Sullivan,[1,2] Alexander DeHaven,[1] and David Mellor[1]

[1]Center for Open Science, Charlottesville, Virginia
[2]Corresponding author: *ian@cos.io*

By implementing more transparent research practices, authors have the opportunity to stand out and showcase work that is more reproducible, easier to build upon, and more credible. Scientists gain by making work easier to share and maintain within their own laboratories, and the scientific community gains by making underlying data or research materials more available for confirmation or making new discoveries. The following protocol gives authors step-by-step instructions for using the free and open source Open Science Framework (OSF) to create a data management plan, preregister their study, use version control, share data and other research materials, or post a preprint for quick and easy dissemination. © 2019 by John Wiley & Sons, Inc.

Keywords: data management • open data • open science • preregistration • reproducible research

How to cite this article:
Sullivan, I., DeHaven, A., & Mellor, D. (2019). Open and reproducible research on open science framework. *Current Protocols Essential Laboratory Techniques*, *18,* e32. doi: 10.1002/cpet.32

INTRODUCTION

The movement to increase transparency in scientific research is gaining widespread support from the scientific community (Baker, 2016). Transparency is the inclusion of additional details and contexts surrounding the research published and the conclusions drawn within the publication. These can comprise inclusion of additional information around the decisions made during the research process, opening up the data used to interrogate the hypotheses, opening up the materials used to conduct the research, or disclosing the plan for the study and analyses. This movement to increase transparency is likely in response to several developments: the challenge of the "Replication Crisis" (Ioannidis, 2005; OSC, 2015), a desire to increase trust through rigor and transparency (Chambers & Munafò, 2019), and perhaps simply as an adaptation to the potential offered by online digital storage and communication technologies. Difficulties in replicating the results of previously published research often result from two general factors: (1) uncertainty about the precise methodology because of unclear or undocumented protocols in the original publication, and (2) presenting tenuous findings using the tools of statistical inference that are only appropriate for rigorous confirmatory findings (Nosek, Ebersole, DeHaven, & Mellor, 2018; Simmons, Nelson, & Simonsohn, 2011). The Open Science movement seeks to address these factors using documentation, materials sharing, and other practices that make the details of research design and execution more transparent. The primary tool developed by the Center for Open Science (COS)—the Open Science Framework (OSF)—is designed to aid researchers in adopting these practices.

As a term, "Open Science" encompasses multiple practices relating to research planning, software selection, data management, and dissemination and publication of research outputs. Researchers may adopt each of the different Open Science practices and tools, individually or collectively, and each practice can be implemented to varying degrees. For instance, researchers might choose just to make more of their research outputs (data, materials, code, etc.) publicly available at the end of a research project but make no changes to how that

CURRENT PROTOCOLS
A Wiley Brand
Current Protocols Essential Laboratory Techniques e32, Volume 18
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cpet.32

research is conducted. Similarly, researchers might adopt the preregistation of their research questions and analysis plans without making any changes to their tools or practices around data management. Here, we outline a protocol for adopting core Open Science practices over the whole course of a research lifecycle, which we believe maximizes the benefits of the individual practices.

BASIC PROTOCOL

The purpose of this protocol is to serve as a guide to creating a more open and reproducible research workflow. This protocol is intended as an overview of how to get started implementing core Open Science practices, not as an exhaustive examination of how to implement those practices in a particular research environment. Where better resources exist externally to this protocol, we will attempt to direct you to them. The following sections cover many of the aspects of a typical research project with steps to use the OSF web tool for implementing the core Open Science practices described in each section. By relating the specific steps for use of OSF to the more general practices, we hope that this document will serve as a useful resource even as the details of software interfaces and features shift over time. OSF is a free, open source research management and collaboration tool designed to help researchers maintain provenance, document projects, and store materials. It is built and maintained by the nonprofit Center for Open Science. Current documentation for all OSF capabilities should always be available from *http://help.osf.io*.

Data Management Planning

"You can't have any sort of reproducibility without good data and project management"
Vicky Steeves, New York University
Librarian for Research Data Management and Reproducibility

[Above quote from Introduction to Research Data Management, *https://doi.org/10.5281/zenodo.2530961*]

Research Data Management includes a wide range of elements in a research project: from how data will be collected and stored to how access to it will be managed during the research period and how that data will be preserved and made reusable after the study is concluded. During project design, these elements of a study may be formally documented in a Data Management Plan (DMP) submitted as part of a grant application. Even when a formal document is not required, we strongly encourage the creation of a detailed plan for managing research data over the full lifecycle of a project, as this planning process can bring multiple benefits.

Selecting metadata standards for your data at the beginning of a project can help to reveal what information about that data needs to be recorded at the time of collection or processing, when that information can be captured most easily and completely. These best practices help improve the accuracy and completeness of the data you are collecting while making that data easier to work with and reuse. Similarly, implementing anonymization and minimization strategies early in the data management workflow can reduce the potential for accidental disclosure of legally or ethically protected data while increasing the portion of your workflow that can be publicly documented. Additional examples of best practices in data management as they relate to a wide variety of different research contexts are available from the DataONE project's best practices database (DataONE Best Practices database, *https://www.dataone.org/best-practices*).

Once the key steps in a data and research management workflow for a study have been identified, it becomes easier to select appropriate tools to enable and support those steps. Recognizing your organizational system and software needs early enables you to avoid the uncertainty and inaccuracies that can creep in when research data needs to be moved around in unplanned ways. Routine requirements such as providing data access to collaborators at different institutions, transitioning ownership of data from a colleague departing the research team, and providing anonymized access to materials during peer review might all lead to an unexpected need to copy or migrate data to new systems, if not considered in advance. When data needs to be moved or processed in ways not addressed in your data management planning, confusion is likely to arise as versioning conflicts or lost metadata (e.g., code books that describe data labels), and may lead to errors or an increase in documentation work when you go to share your research outputs publicly.

Considerations:

- Ensure that all necessary stages of the research are covered by appropriate technology.
- Plan for flexibility to handle unexpected tasks so that procedures are in place for migrating the data easily between different tools and stages of research.

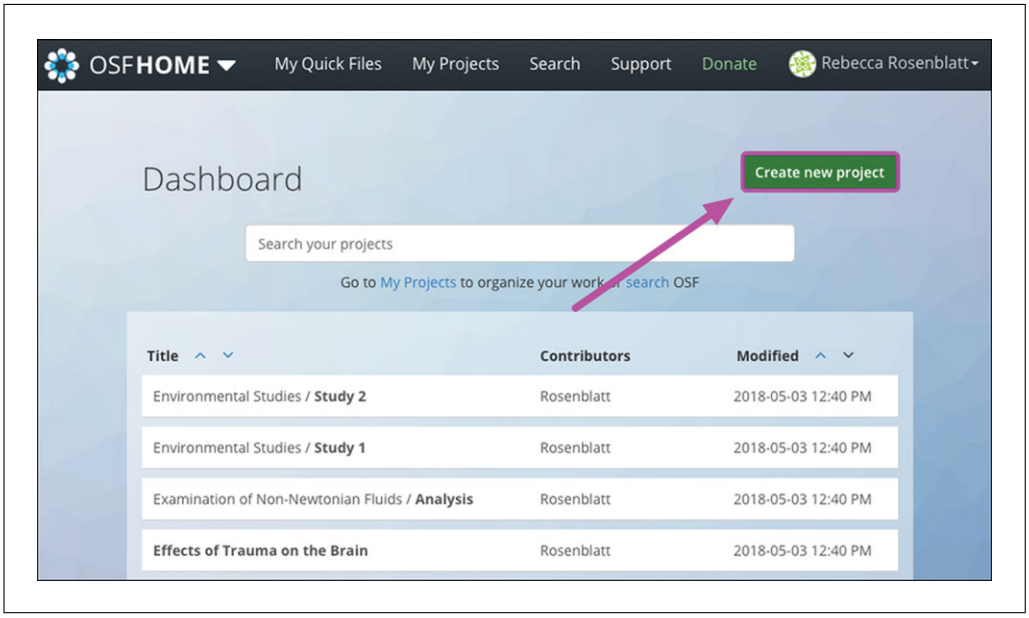

Figure 1 "Create a project" button on an OSF Dashboard.

- Use version control or other provenance-management tools.
- Know where the data, analytical code, and other digital materials are likely to be stored at the end of the project.
- Consider who you wish to use these research outputs and what information they would need in order to locate and make effective use of them.

Implementation steps using OSF

OSF provides controlled-access sharing and persistent online storage, has built-in tools for version control, allows for distinct, structured workspaces, and enables each section of a project to be made public or private independently. These core OSF features enable researchers to build a workspace that is the embodiment of their data management plan. Each step below is supplemented by relevant sections of the OSF help documentation where additional information may be found.

1. Create an account.

Visit *https://osf.io/register/*, where you can log in through your existing ORCiD account or through an affiliated OSF Institution, or provide an e-mail address and password. Accounts are free.

2. Create a project.

When you log into OSF, you will be brought to your Project Dashboard (Fig. 1) where you can create a new project. Create a project and give it a descriptive title to reflect the research you will be pursuing. Once your project is created, go to it by clicking the "Go to project" button that appears or by clicking on the project's name on your Dashboard (OSF Help Guide: Create a Project, *http://help.osf.io/m/projects/l/481539-create-a-project*).

3. Use components to build your project into a structured workspace that fits the data management needs identified in your DMP.

Components are the fundamental unit of organization in OSF. Each component is given a unique URL, called a Globally Unique Identifier (GUID; OSF Help FAQ: What is a GUID, *http://help.osf.io/m/faqs/l/726460-faqs#what-s-a-globally-unique-identifier-guid-what-metadata-is-maintained-about-them*), each can be made public or private independently of any other portion of the project, and each may have a different set of user permissions. Create as many components as appropriate for your planned data management workflows. At a minimum, we suggest creating components for each class of research output your project will generate (Data, Materials, Analysis Scripts, etc.). The "Components" section of your project page contains the "Add Component" button (OSF Help Guide: Create Components, *http://help.osf.io/m/projects/l/481998-create-components*) and displays all existing components at the top level of your project structure (Fig. 2).

You can also create components inside other components (Fig. 3). This enables the creation of multi-level compartmentalized structures for organizing your project and managing access to each section.

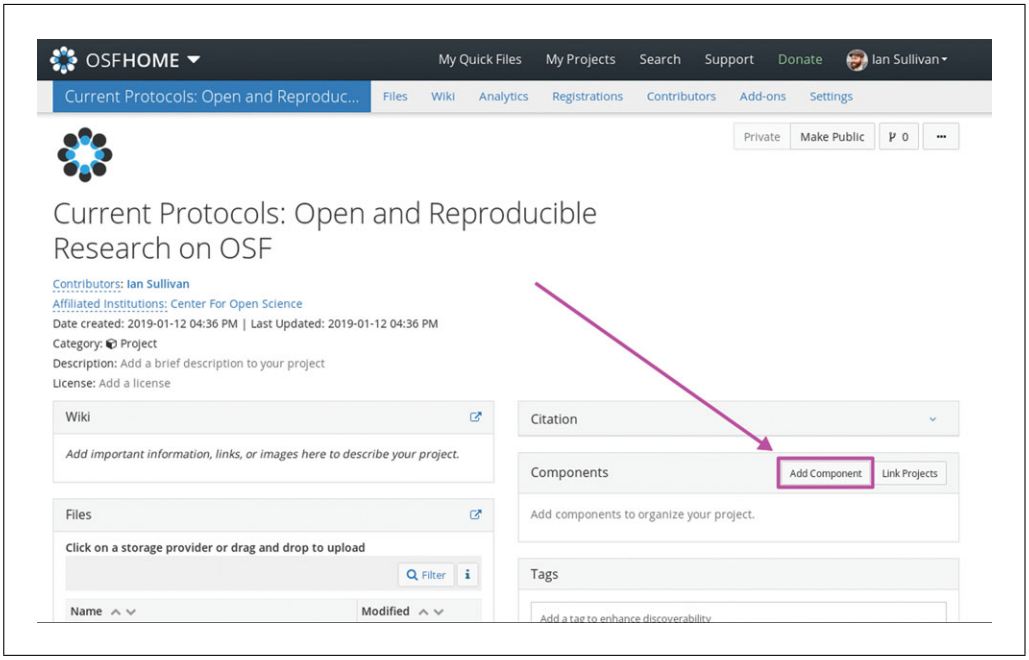

Figure 2 "Add Component" button on an OSF project page.

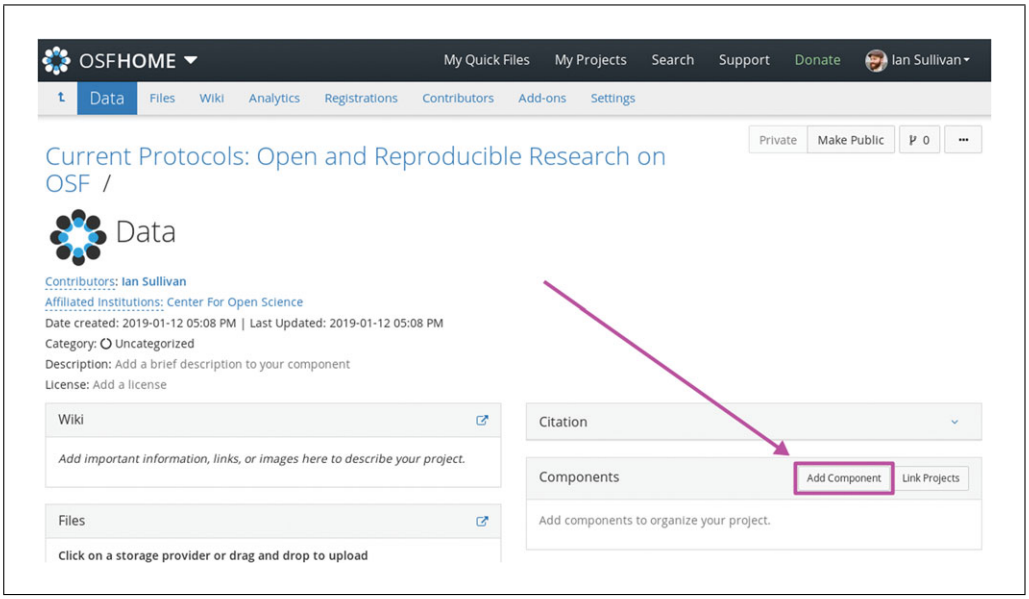

Figure 3 "Add Component" button from inside an existing component of an OSF project.

4. Incorporate documentation into your project structure.

OSF provides multiple ways to add documentation to each area of your project. Each project and each component within it has a title, a description, and a wiki page that can be freely edited by project contributors. All of these can be appropriate locations to include information about your planned processes for key elements identified in the project's DMP such as metadata collection, documentation updating, and the eventual publication of materials (Fig. 4).

Adding these process reminders and implementation details will help ensure that they remain prominently visible and readily accessible to the research team.

5. Configure access permissions for the research team.

All OSF projects and components are ***private by default***, so if you wish to work collaboratively you will need to add the other members of your research team to your project by searching for their OSF user accounts using the "Contributors" tool (OSF Help Guide: Add Contributors to Projects and Components,

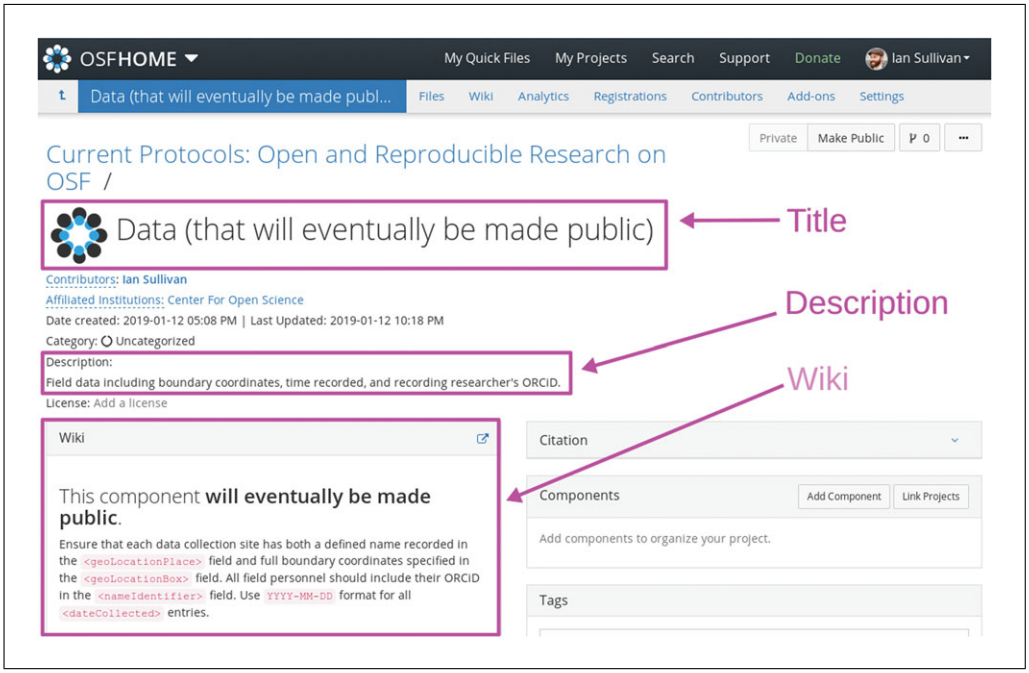

Figure 4 Incorporating references to and reminders about important data management steps in an OSF component.

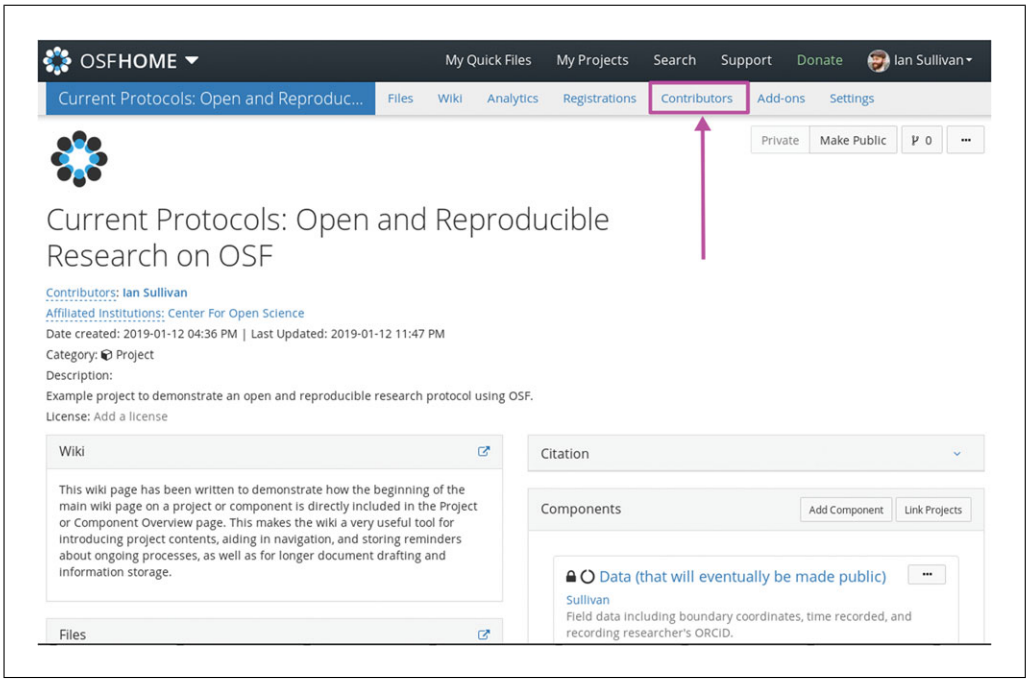

Figure 5 Contributor tool for managing collaborative access to a project or component.

http://help.osf.io/m/projects/l/482001-add-contributors-to-projects-and-components; Fig. 5).

OSF users can be searched for using their names or any other information they may have optionally added to their user profiles (OSF Help Guide: Editing Your Profile, *http://help.osf.io/m/settings/l/524113-edit-your-profile*). OSF does not disclose the e-mail addresses associated with user accounts, so you cannot use that information when searching for users to add to your project.

To quickly add a contributor to multiple components in your project, click on the Contributors tool from the main project overview page. After searching for the contributor's account and selecting the desired permissions, you will then be presented with a menu listing all the components underneath the main project (Fig. 6). Simply check all components

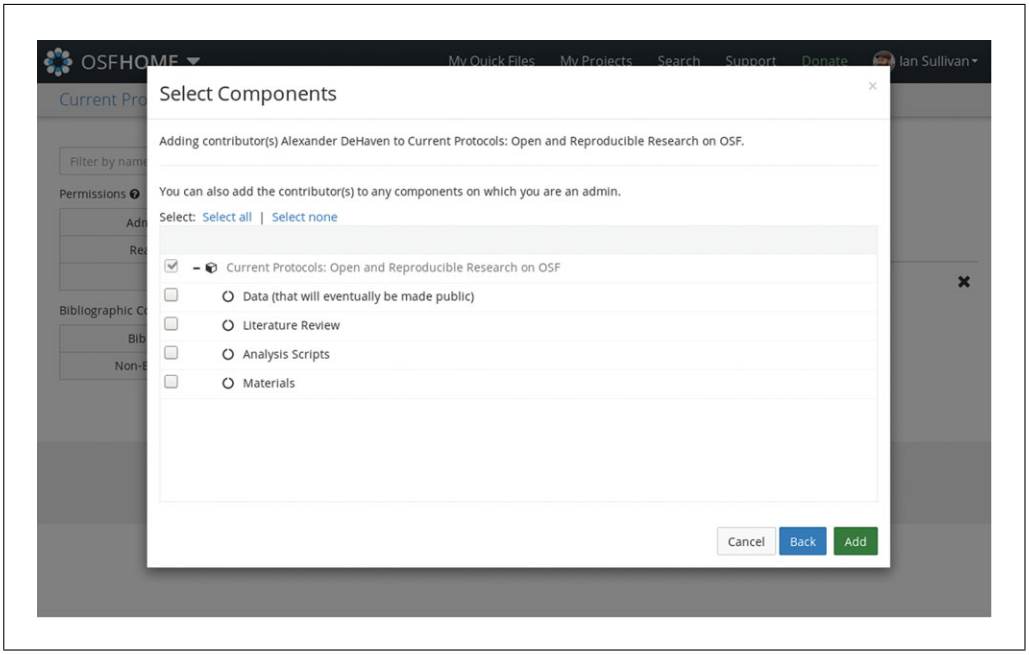

Figure 6 Adding a contributor to multiple components. This option appears when you add a contributor from the main project overview page as opposed to from within an individual component.

that you wish to add a contributor to. Should you need to expand on your project structure at a later date by adding new components, you can save time by clicking the "Add contributors from <parent project|component name>" option during component creation (Fig. 7), which will import the contributors from whichever section of the project you are currently in.

OSF gives you substantial control when setting up user permissions. Each contributor can be given one of three levels of permission (read, read+write, or administrator; OSF Help Guide: Understand Contributor Permissions, *http://help.osf.io/m/projects/l/814917-understand-contributor-permissions*), each component can have a separate list of users, and those users can be granted different levels of permissions between different components or between components and the top-level project. This makes it simple to give a team member permissions on the "Data" and "Materials" components but not the "Manuscript" or "Grant Management" ones. We recommend following the principle of least permission (Saltzer, 1974) by assigning collaborators the minimum permissions needed to accomplish their research tasks in each section of the project. In particular, you should give careful consideration to who is given Administrator permission for any portion of the project.

- Things to be aware of:
  - Where possible, ensure that there are two Administrator users for each component in the project to ensure continuity should one leave the research team.

The second administrative user could be a departmental administrator or other colleague uninvolved with the research project, assuming that data protection and blinding considerations allow it. In such a case, the user not contributing to the research should likely be added as a "Non-bibliographic" (OSF Help Guide: Understand Contributor Permissions, *http://help.osf.io/m/projects/l/814917-understand-contributor-permissions*) contributor so that the contributor's name does not appear publicly should you make some or all of the project public.

- Document the procedure that should be followed when members of the research team depart, especially if you intend to make portions of the project public later.

To help ensure that all contributors receive credit for work made public on the platform, OSF generates an automatic citation for each project and component based on the list of contributors to that section. Removing someone from the list of contributors therefore has the dual effect of ending their access to any materials while also removing his or her name from the citations for each of those project areas. If members of the research team leave a project, due to shifting responsibilities, jobs, etc., and their access to the research materials needs to be completely removed for the remainder of the project, you may wish to

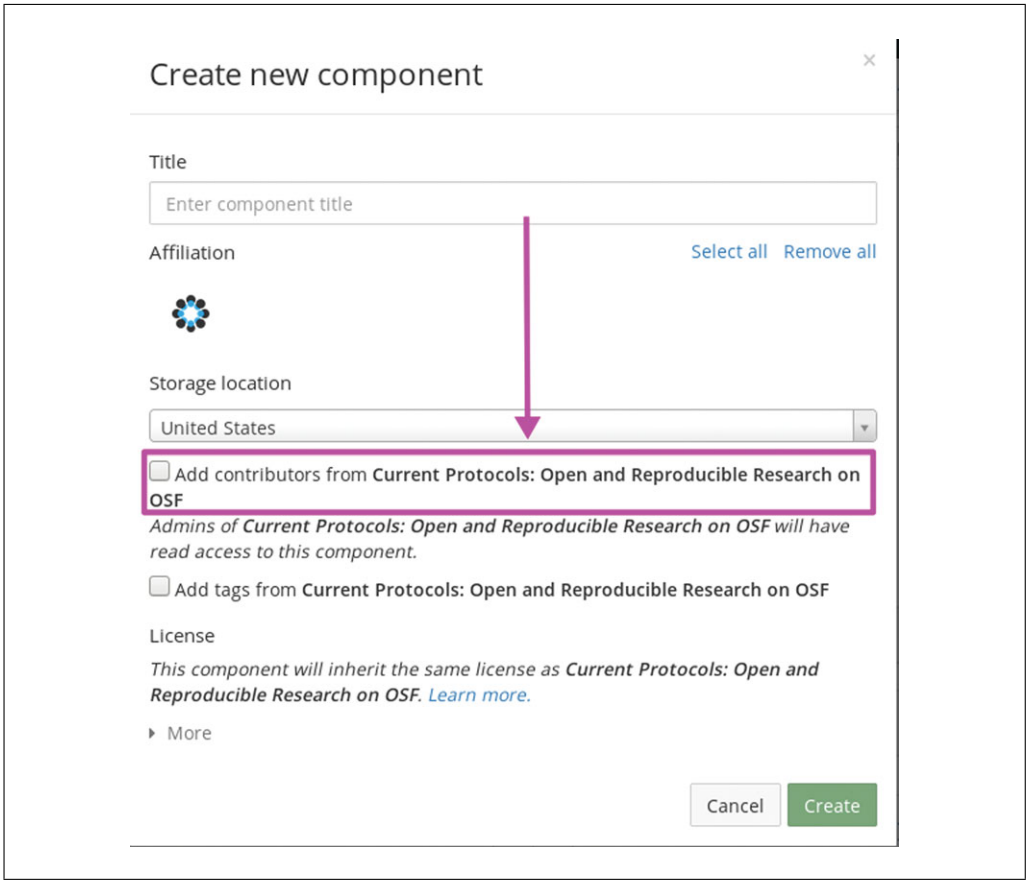

Figure 7 Option during the creation of a new component to have it import the current contributor settings of its parent component/project.

consider either manually editing the citation for those project areas (OSF Help Guide: Creating Custom Citations, *http://help.osf.io/m/sharing/l/985015-create-custom-citations*) to preserve their name or making a note to re-add the departing members as "read-only" contributors to those sections when you make them public. The project or component's activity log (OSF Help Guide: View Recent Activity, *http://help.osf.io/m/projects/l/488030-view-recent-activity*) is an invaluable resource for reviewing the contributions of each research team member.

Preregistration

Preregistration is the practice of documenting your research plan at the beginning of your study and storing that plan in a read-only public repository such as *clinicaltrials.gov* (U.S. National Library of Medicine Clinical Trials Registry, *https://clinicaltrials.gov/*) or osf.io/registries (OSF Registry site, *https://osf.io/registries*). The benefits of preregistration depend directly on what kind of information, and how much of it, is included inside that preregistration. A very sparse outline of a study plan may be sufficient to increase the discoverability of the research and thus help to address the file drawer effect (Franco, Malhotra, & Simonovits, 2014; Rosenthal, 1979), yet insufficient to assist in evaluating claims resulting from that research. Including a detailed analysis plan in the preregistration may additionally help reduce unintentional false positive inflation of results (Forstmeier, Wagenmakers, & Parker, 2017) and better enable readers to distinguish exploratory from hypothesis-testing elements in a study (Nosek et al., 2018). Both modes of research are essential for science to advance, but presenting the results of data-dependent, exploratory discoveries using the tools of statistical inference designed for confirmatory studies makes the results appear more surprising and publishable, which comes at the expense of their credibility (Nosek, Spies, & Motyl, 2012). While we recommend preregistering all types of research, the most benefits accrue when performing hypothesis testing and confirmatory research, and these benefits are of particular importance to addressing issues of reproducibility in the published literature (Munafò et al., 2017). Our strongest recommendation is therefore to preregister confirmatory research

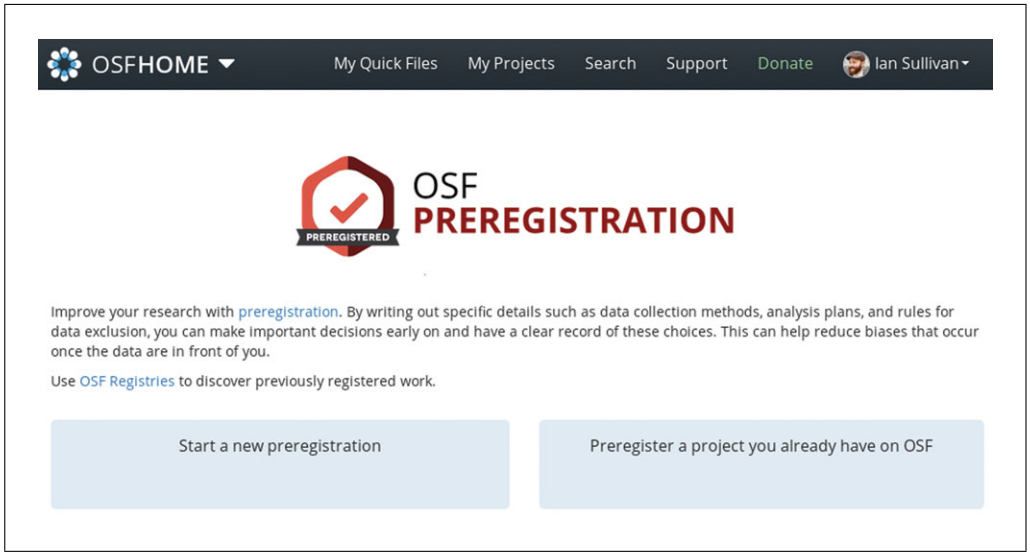

Figure 8 OSF Preregistration site. Preregistrations can be associated with an existing OSF project or started from scratch.

with a detailed analysis plan. If setting out on purely exploratory research or pilot studies, preregistration can still help you remember that intention at the end of that project.

When reporting the results of preregistered research, there are four basic rules to follow:

1. Include a link to the original preregistration. This will allow readers to compare the planned and reported study.

2. Report the results of all pre-specified work. If you set out to run 20 tests, report all of their results. This is especially urgent if only a few of those tests were statistically significant. Since some will be significant by chance alone, readers deserve the full context of the results in order to interpret the overall findings.

3. Clearly label any unplanned analyses. Plans change. Diverging from your preregistered study plan does not invalidate that preregistration (*https://cos.io/blog/preregistration-plan-not-prison/*). If changes from the original plan are made, they should simply be transparently documented in the resulting research. This transparency enables readers to better evaluate any impact that changes from an original plan may have on the results (Simmons et al., 2011).

4. Finally, include a "Transparent Changes" document for any deviations that occurred from your preregistered plan. This may be unnecessary if your exploratory analysis section covers all differences, but in many studies there will be additional changes. Perhaps participant recruitment was tougher than imagined. Or perhaps some measures were impossible to collect. Whatever the changes are, document them and include them in a footnote or separate section.

For additional materials discussing what information should be included in a preregistration, how preregistration can be effectively implemented, and examples of selected preregistrations from multiple disciplines, refer to cos.io/prereg (Center for Open Science Preregistration information page, *https://cos.io/prereg/*).

OSF Example

OSF welcomes preregistrations from any field or discipline, and we support researchers using OSF as a preregistration platform whether or not they will also be using its data storage, collaboration management, or publishing capabilities. To begin a preregistration on OSF, visit *https://osf.io/prereg/* (Fig. 8), where you will be given the opportunity to begin a new registration from scratch or to connect the preregistration to an existing project (OSF Help Guide: Create a Preregistration, *http://help.osf.io/m/registrations/l/546603-create-a-preregistration*). Note that only Contributors with administrator permissions can register any portion of an existing OSF project.

To assist researchers in drafting their preregistrations, OSF offers a number of different templates that researchers can select from (Fig. 9). The default, "OSF Preregistration," presents a comprehensive list of questions to address. This form is a useful place to start if you are new to preregistration, is our general recommendation for researchers unsure what template to use, and is the form used by

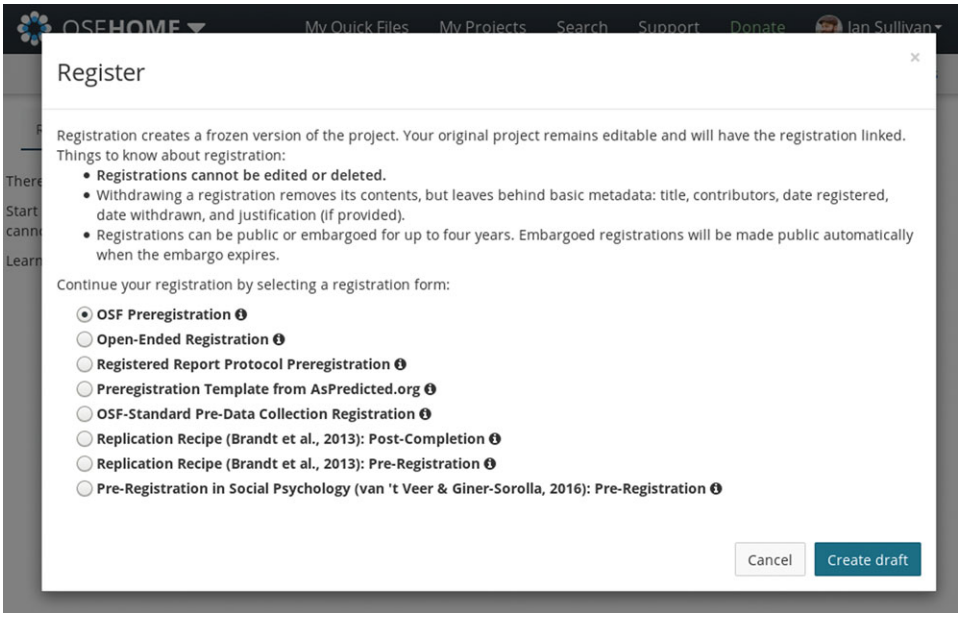

Figure 9 List of templates available when making a registration on OSF. The default 'OSF Preregistration' option is recommended for preregistrations.

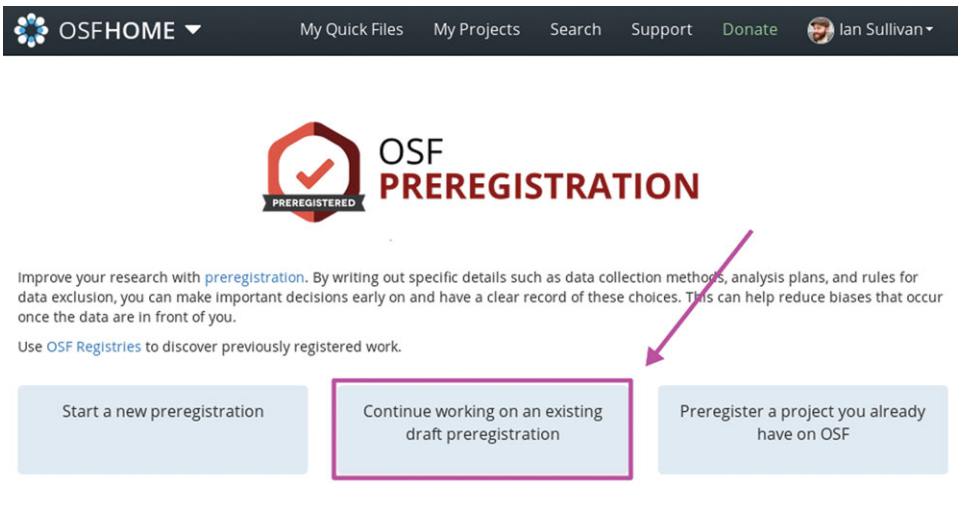

Figure 10 OSF Preregistration site showing the option to return to editing an existing draft preregistration.

researchers in all of the collected exemplar preregistrations available on *https://cos.io/prereg*. If you begin a preregistration but cannot enter all of the information during a single session, you can return to the draft later by visiting *https://osf.io/prereg/*, where a new option to "Continue working on an existing draft preregistration" will be available to you (Fig. 10).

As the final stage of submitting a preregistration you have the option to make that preregistration public immediately or to enter it into an embargo of up to four years (OSF Help Guide: Create a Preregistration, *http://help.osf.io/m/registrations/l/546603-create-a-preregistration*). On an OSF project, administrators can end an embargo early (OSF Help Guide: End an Embargo Early, *http://help.osf.io/m/registrations/l/524207-end-an-embargo-early*), but embargo dates cannot be **extended** once set. Once a registration has been submitted, all contributors to that project will be notified of that fact by an e-mail from OSF. Everyone will have 48 hr to review the contents of that registration, and all project

administrators will be able to cancel the registration during that time. After 48 hr, administrators on a registration can still remove it from OSF through a withdrawal process (OSF Help Guide: Withdraw a Registration, *http://help.osf.io/m/registrations/l/524209-withdraw-a-registration*), but a record is left in the public OSF registry detailing when the registration was made, when it was withdrawn, and any reason for the withdrawal provided by the registration's team as they withdrew it. We strongly encourage using the initial 48-hr period after a registration is submitted to review the contents of that registration and verify that all materials are present and correct, and that only materials you are legally and ethically allowed to make public are included.

Version control

Managing and tracking research materials can prove a significant challenge during the course of research, especially in collaborative projects. Improving the ability to consistently track and retrieve each version of a file can lead to more efficient collaboration and increased accuracy of research results. This is most effectively accomplished through the use of version control systems that automate various portions of the storage and record keeping. Version control systems vary widely in capabilities and complexities, from automatic tools that sync sequential versions of your files with various cloud storage offerings, to tools like git (see also gitlab, github) that allow for active management of multiple branching and merging versions of the same file. When selecting a version control system for your research, the primary consideration should be what can be most consistently implemented in your current environment. If your chosen tools are too challenging to use or are not available on all platforms where research materials need to be managed, individual researchers may begin use of their own tools and systems, resulting in the same kind of confusion that you hoped to avoid. Existing familiarity with a particular tool, availability of tools, and ease of configuring those tools on new devices are all important considerations when selecting a version control system.

Once you have identified what systems can be implemented consistently in your research environment, some additional considerations are worth including in your evaluation. In addition to general software selection considerations [price, size of the community using and supporting the tool, amount of community control, Free and Open Source (FOSS) licensing terms, etc.], some considerations specific to version control tools include:

- How much complexity is functionally required by your particular task and workflows? For instance collaborative software development may benefit from more complex tools than are needed to track sequential versions of a data file during analysis.
- How many versions do you need to keep and for how long? Some tools may only allow you to keep a certain number of versions or to only keep them for a certain period of time, or may limit whether you can migrate those versions to other tools. Ensure that any limitations are in line with your research and archival needs or select tools without such limits.
- How much of the metadata and provenance tracking data specified in your Data Management Plan will this tool record as it is implemented in your workflow? If your collaboration dynamic delegates all interactions with the version management system to a small number of team members, you may end up with less complete or less accurate provenance tracking than that tool would collect if all team members were interacting with it directly. Both the theoretical capabilities of the tool and how those capabilities are to be utilized in practice should be considered.

OSF Example

Managing files on OSF begins the first time that you upload a file to the system. Any kind of file may be stored on OSF as long as each individual file is 5 Gigabyte (GB) or less in size. Hundreds of file types (list of file types that will be rendered directly in the browser by OSF, *https://github.com/CenterForOpenScience/modular-file-renderer/blob/develop/supportedextensions.md*), including a variety of audio, video, and common office document types, along with less common formats like .stl 3D scan files, will be rendered directly in the browser by OSF. You add files to your OSF project using the Files tool (Fig. 11), which will list your main project level and all of the components underneath it (Fig. 12). Each of these will have an "OSF Storage" item listed under the project or component name. Advanced users who make

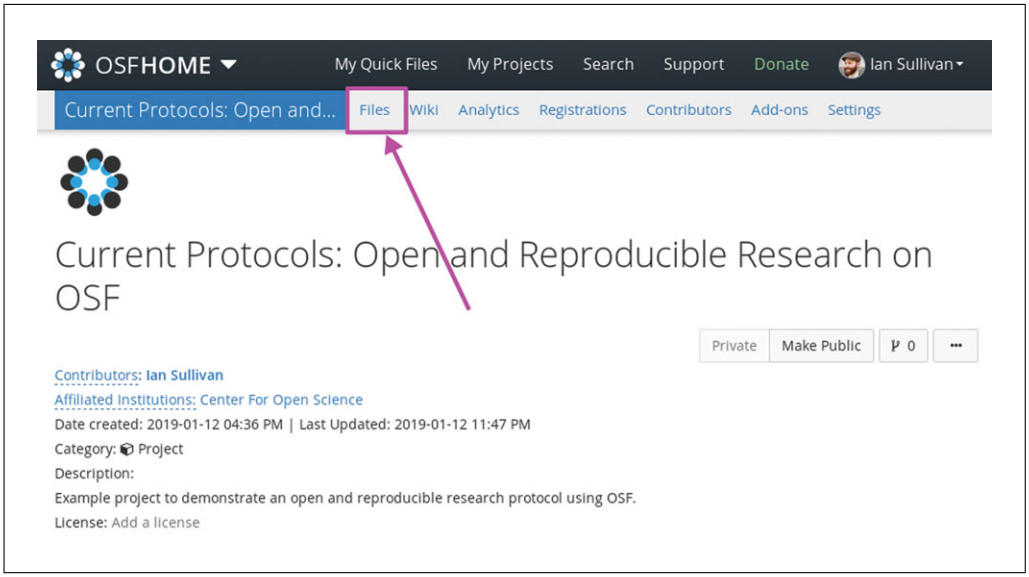

Figure 11 Location of the "Files" tool on an OSF project or component's overview page.

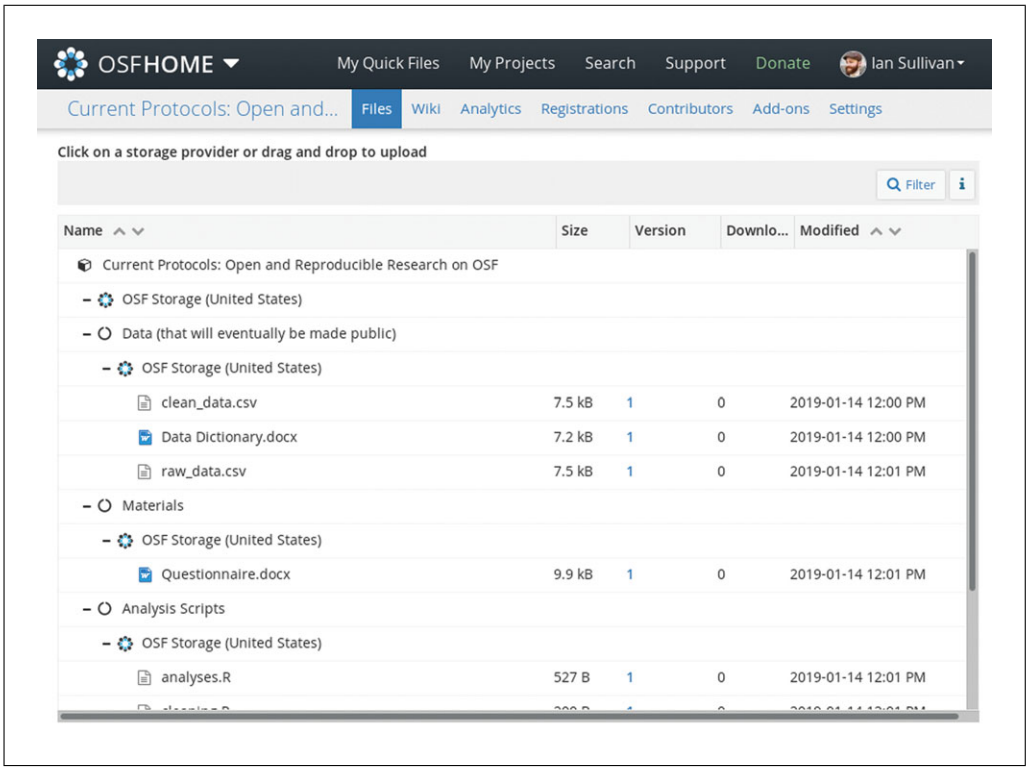

Figure 12 The Files page on an OSF project.

use of OSF's Add-on system (OSF Help: Add-ons, *http://help.osf.io/m/addons*) to integrate other tools and services for file storage will also see logos for whichever services they have enabled in a component next to the "OSF Storage" logo for that component. To add a file to a component, simply drag files from your computer to the "OSF Storage" icon underneath the component or click on the "OSF Storage" item and use the "Upload" button that will appear at the top of the file page (Fig. 13). Files can also be uploaded using the same process through the miniature files section of the main project and component overview pages.

Once a file is uploaded to OSF, you can click on its name to view that file (Fig. 14) and gain access to all file-related OSF actions. For plain text files like the R script in Figure 13, you are able to edit them directly through OSF (Fig. 15). This capability is accessed by clicking on the "Edit" button on the file view

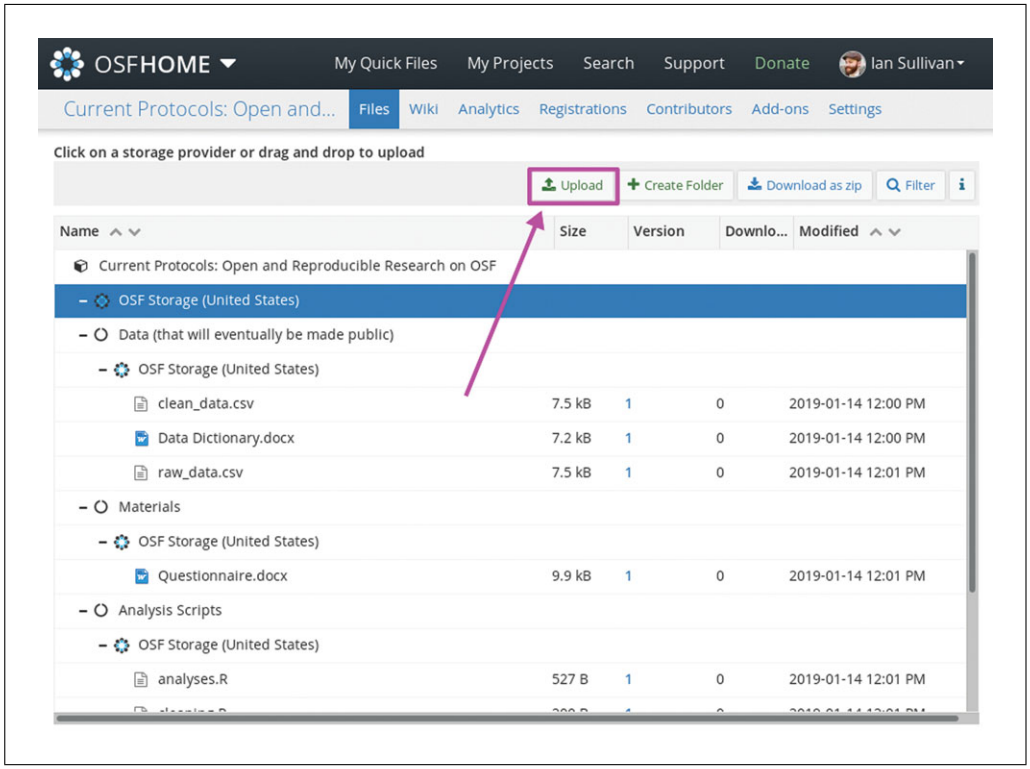

Figure 13 The upload button that appears when one of the "OSF Storage" items is clicked on. Note the highlighted blue row indicating which OSF Storage the button relates to.

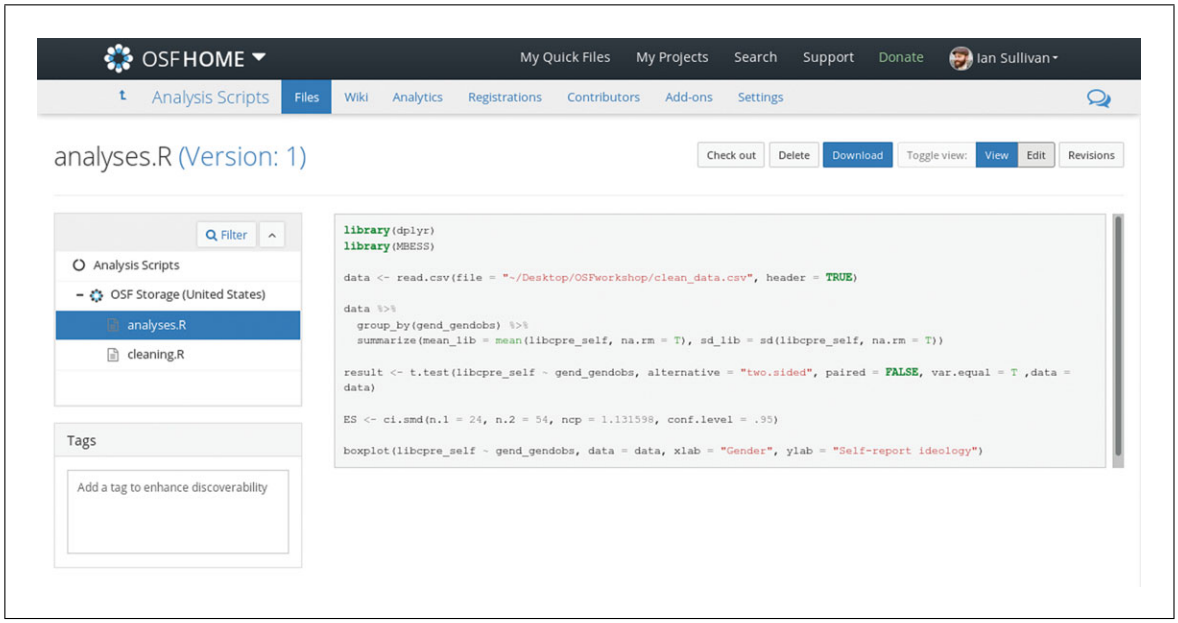

Figure 14 View of an individual file, an R script, on OSF.

page (Fig. 16). Once you have edited a plain text file, hitting "Save" creates a new version of the file. Plain text files are the only ones that can be edited directly through OSF. For all other file types, including audio, video, word processor, etc., files should be edited locally and then re-uploaded to OSF following the same process by which they were originally uploaded. Simply upload the new version to the same location within your OSF project or component and use the same file name. This will tell OSF that the file is actually a new version of the earlier file.

All versions of a file stored on OSF are saved and made available through the file's revisions view. This view is accessible by clicking on the "Revisions" button or the file's version number (Fig. 17). The revisions

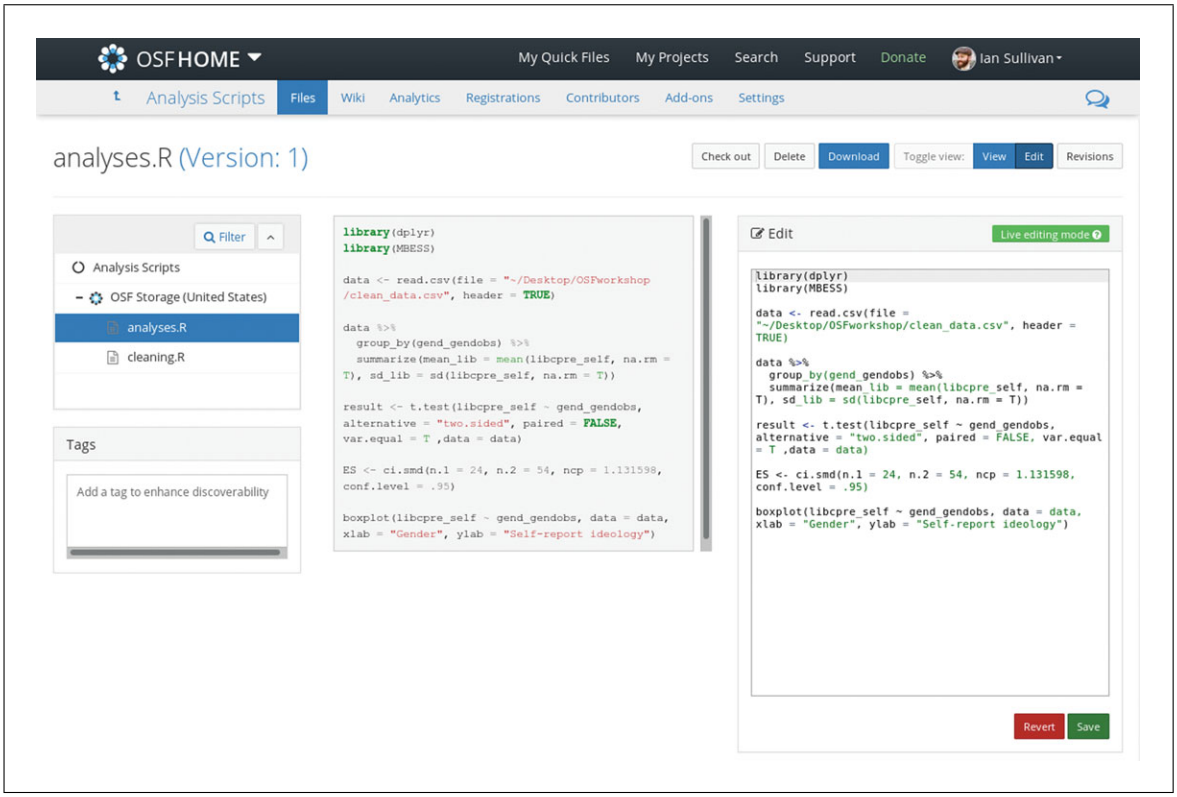

Figure 15 View of a plain text file being edited on OSF. The edit window is on the right and a live preview is in the middle of the page.

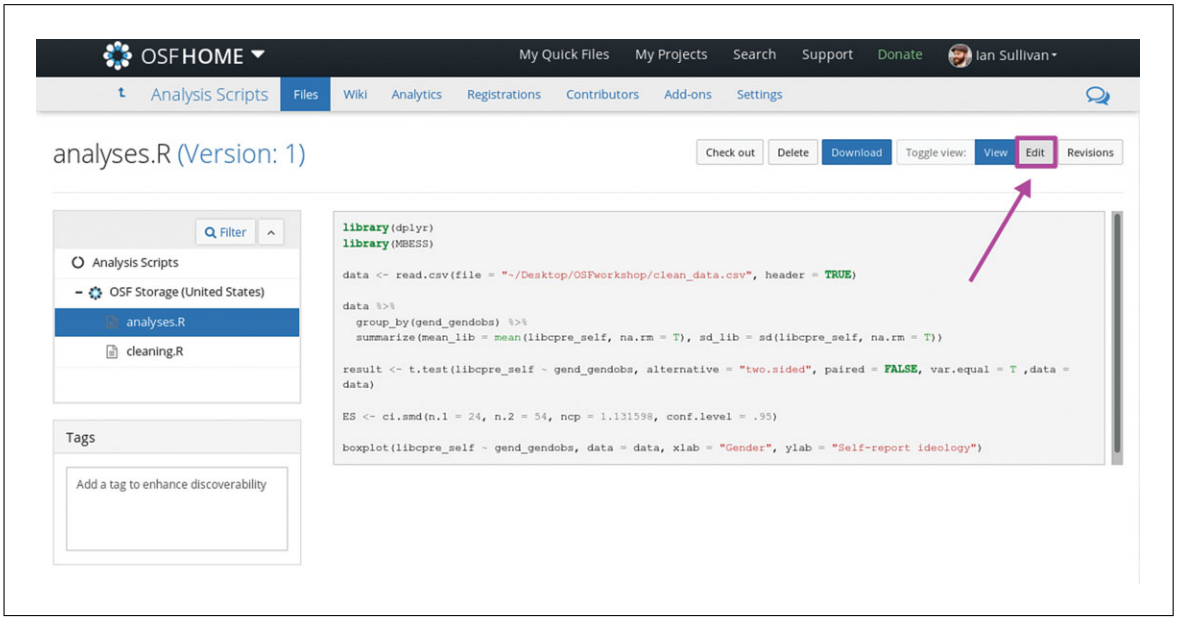

Figure 16 The "Edit" button that enables plain text files on OSF to be edited directly in the browser.

view (Fig. 18) includes information about each version of the file, who made it, and when that version was added to OSF. Each version has an independent "Download" button, and old versions can be rendered in the browser by clicking on their version number. Files in public components or project areas also show a count of how many times they have been download by non-project team visitors.

Reproducible software environment

As the role of software tools in the production of research findings continues to grow, so too do the challenges that they pose for the researcher interested in a transparent and

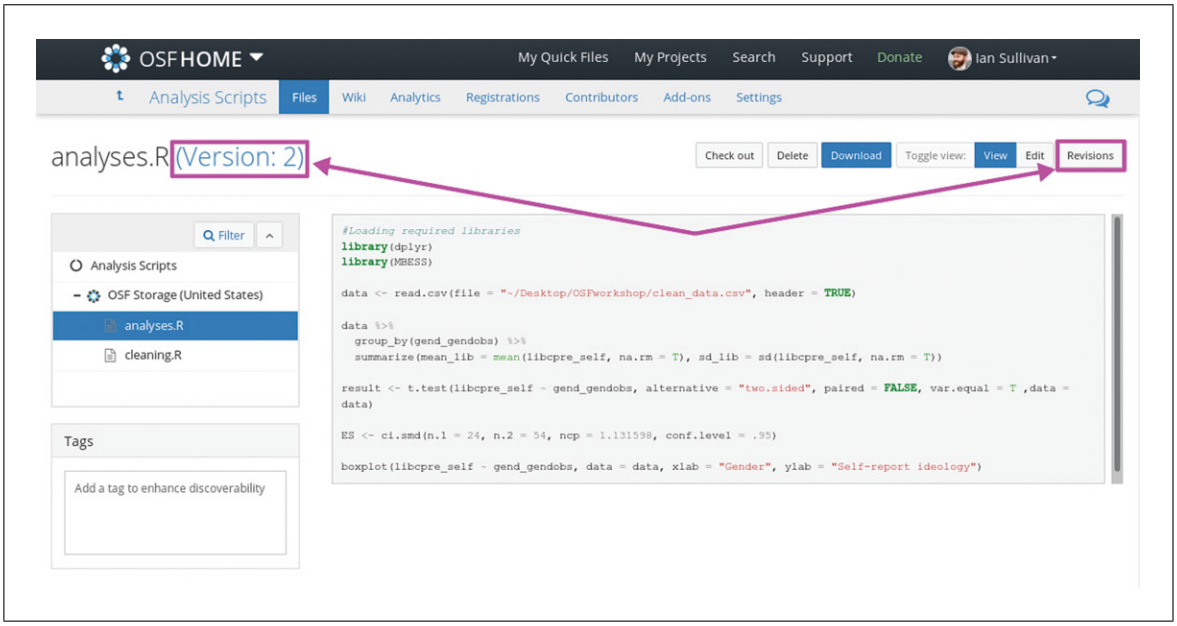

Figure 17 The two ways to view earlier versions of a file stored on OSF: clicking the "Revisions" view button or clicking on the version number next to the file name.

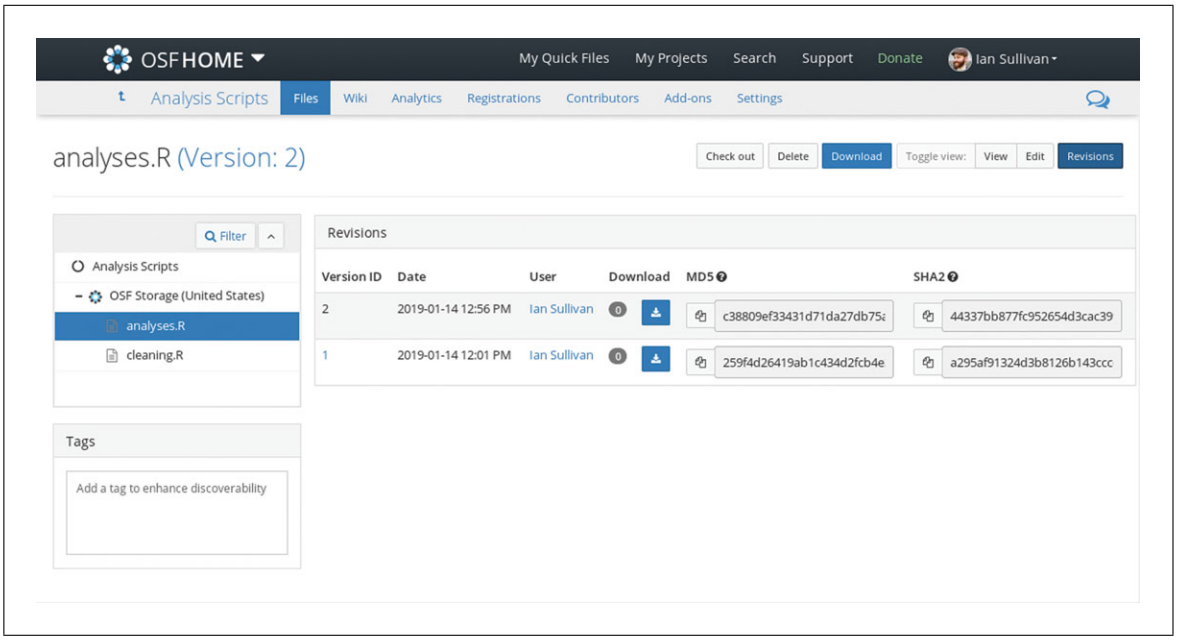

Figure 18 The Revisions view of a file on OSF providing information about and access to earlier versions of the file that were uploaded to the same project location.

reproducible research process. In terms of transparency, both the procedure followed when using a software tool and the tool itself are valid topics for investigation and concern. When documenting workflows, tools that use scripts stored in text files make repeating an exact series of steps significantly easier than those with a graphical user interface where steps must be documented by describing the various menus to click and values to be entered in different windows. Since the software executing these instructions is itself an evolving piece of digital machinery, transparency about how those instructions are carried out and how that process may change over time are important to consider when seeking to understand or expand on a particular research finding. Software available under a free and open source software (FOSS; GNU Project list of FOSS licenses, *https://www.gnu.org/licenses/license-list.html*) license enables this level of transparency, while also allowing the relevant software source code to be preserved for analysis

and use in future. For these reasons, FOSS software is to be preferred wherever possible. OSF is one such FOSS tool, and OSF-related portions of your research workflow (adding files, uploading new versions, etc.) can be scripted through use of the OSF API (OSF API Documentation, *https://developer.osf.io*), which can be interacted with directly or through the use of third-party libraries including osfR (OSF R package, *https://github.com/CenterForOpenScience/osfr*), and osfclient (osfclient python library and command line tool, *https://github.com/osfclient/osfclient*).

The quickly shifting nature of software development poses an additional challenge even when using repeatable scripts with FOSS software. Rapidly changing software versions make it a near certainty that any software running today will no longer run within a few years or, if it does run, that it may behave differently than expected. This is true for software specifically designed for research or analysis, down through layers of shared libraries, to general computer operating systems and drivers for specialized hardware. The effort to create a fully archivable and reproducible computing environment will depend significantly on the software you are using, and is beyond the scope of this protocol. Substantial efforts are being directed at this challenge by community efforts like Refroze (Reprozip project page, *https://www.reprozip.org/*), being developed by the ViDA Group (NYU Visualizing, Imaging, and Data Analysis Center, *https://vida.engineering.nyu.edu/*) out of New York University (NYU), mybinder.org (MyBinder reproducible computing project, *https://mybinder.org*) from the Project Jupyter (Project Jupyter, *https://jupyter.org/*) team, and others. For those interested in taking a first step toward this kind of reproducible software environment, we recommend two steps. First, include a list of the software names and versions used during each research project in the methods section of papers that result from that research. Second, during your research, re-run any key analyses on a second machine. Ideally, this will be a machine free of previous installs of the software necessary for your data processing and analysis steps. Having to set this software up on a second machine will help ensure that all relevant steps are appropriately documented and that your final results are not dependent on hidden settings or outdated references on your primary machine. Finally, where possible, archive the key software and any installation instructions from that second machine and publicly share that archived software alongside the other materials from your research.

Sharing non-paper research outputs

The sharing of all research outputs, including data, code, materials, and other types of information beyond the traditional research paper, has the potential to aid the advancement of scientific progress generally and benefit individual researchers by adding transparency to their research process as well as potentially increasing citations to their work (Piwowar & Vision, 2013). In order to maximize these effects, we encourage sharing of all data, materials, and software code whenever possible, a recommendation most recently echoed by the ERAC Standing Working Group on Open Science and Innovation (European Research Area and Innovation Committee, 2018). The likelihood that sharing any particular research output will contribute to these benefits depends on many factors, including how complete it is, how completely it is documented, where it is stored, and how it is shared.

As mentioned in the section on Data Management Planning at the beginning of this protocol, planning for the publication of research outputs from the beginning of the research process can have a significant impact on the quantity of material available for sharing as well as the quality of documentation and metadata for those materials. Because results are most commonly shared months or years after the point when data is collected and processed, producing the documentation, metadata, and other elements important to sharing research outputs may require more work when done at the end of a research process than it would have required if planned for from the beginning. This potential reduction in the work required to share research outputs makes Data Management Planning an important supporting practice for the sharing of research outputs. Likewise, setting the intention from the beginning of a project that research outputs will eventually be made public may support and assist teams during the process of data management. The assumption that work will be made public may provide additional motivation for researchers to maintain metadata documentation and version management processes agreed upon at the outset of research, or to add additional descriptions and quality control tests to custom software and analysis scripts.

When sharing research outputs, a variety of factors influence how readily those materials

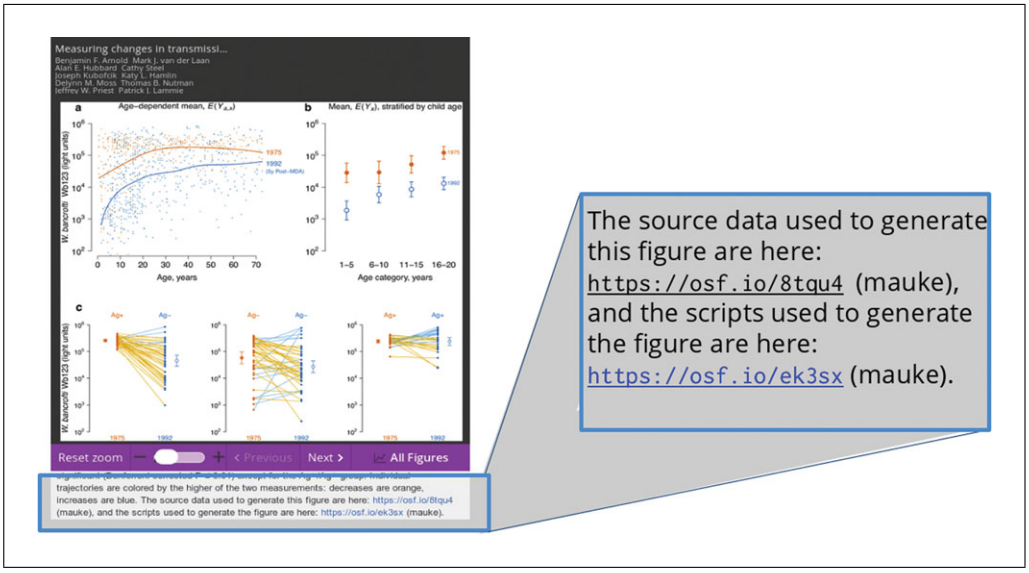

Figure 19 Example of linking to source data and analysis scripts from within a figure caption from Figure 1 in Arnold et al. (2017).

can be found, understood, and used. In addition to the previously mentioned issues of metadata collection, data anonymization, workflow documentation, and reproducible software environment, issues related to choosing an appropriate copyright license (*https://osf.io/qpxv2/*; see Internet Resources), selecting interoperable formats, and others are important considerations. For a more comprehensive look at the range of issues involved and how they may relate to your research outputs, refer to the "FAIR Guiding Principles for scientific data management and stewardship" (Wilkinson et al., 2016). Depending on the type of research outputs you produce, the particular considerations for making them FAIR will vary. Regardless of the type of output, we suggest favoring publishing tools that enable unique citations for each type of research output and as many of the individual elements inside that type as you might wish to refer to directly.

Having a clear citation available for each of your data, materials, code, or other class of research outputs makes it easier to track what elements of your research are being cited and may help reinforce the norm that each of these different types of outputs should be cited if used in new research. Unique and persistent identifiers at the file level additionally enable you to tie your research outputs directly to relevant portions of your research papers. This can take the form of links inside figure captions to the source data and analysis scripts that created the figure (Fig. 19), links from methods sections to more complete protocol information, links from representative image selections to the complete collection of images, or other similar connections. These direct connections add information to your research findings and help your research paper serve as a map to the additional research outputs you make public. In one study evaluating open data (Roche, Kruuk, Lanfear, & Binning, 2015), this behavior of direct linking was a notable feature of the datasets that simultaneously scored highly for both completeness and reusability. OSF generates citations for each component in a project and assigns every project, component, and file a short URL that is globally unique and whose persistence is guaranteed by a data preservation fund currently sufficient to provide 50+ years of public access (OSF Data Preservation Fund, *http://help.osf.io/m/faqs/l/726460#what-if-you-run-out-of-funding-what-happens-to-my-data*).

OSF Example

By default, OSF projects are private and only accessible to those users listed as project contributors. Project administrators are also able to make individual components or the project as a whole public. Making materials public on OSF does **not** change who is able to edit or otherwise change those materials. The only ones able to make any modifications to an OSF project contents are contributors to that project with "Read+Write" or "Administrator" permissions.

Because components can be made public or private independently of each other, each project and component overview page has a

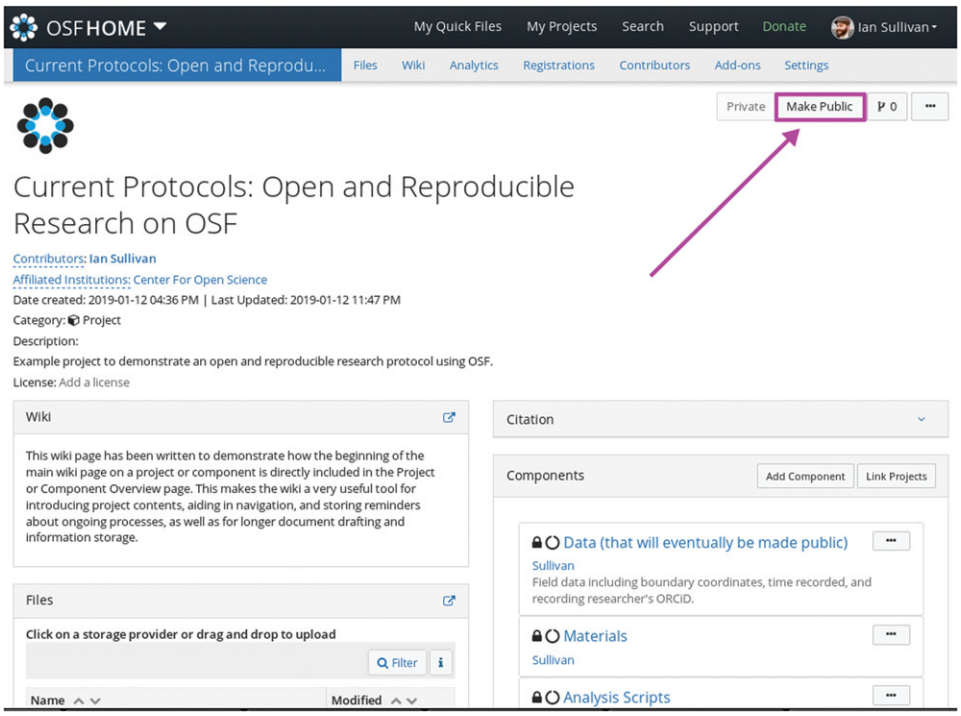

Figure 20 The "Make Public" button on an OSF project overview page.

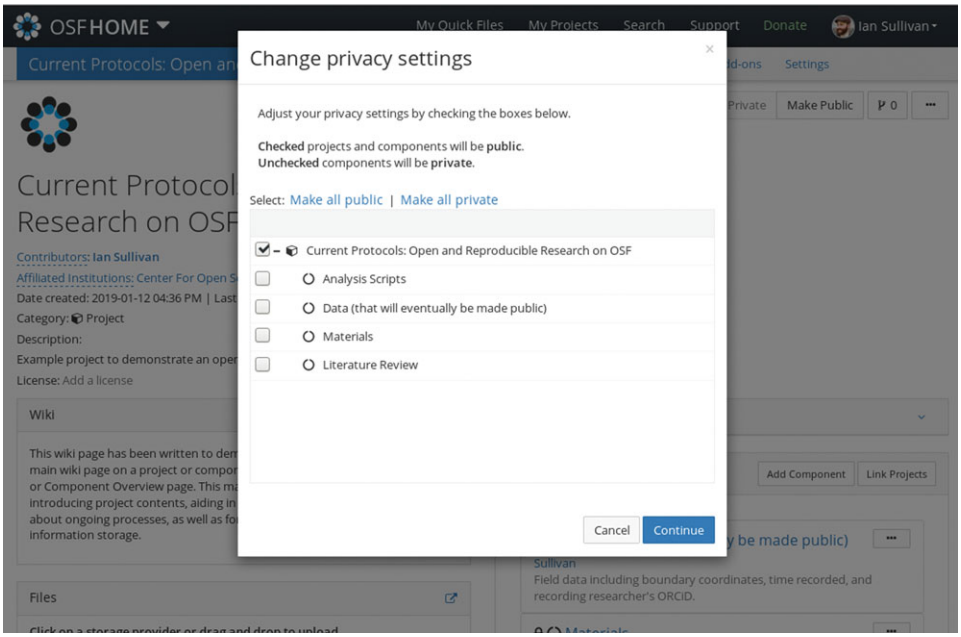

Figure 21 Menu of project components to make public. This menu appears when you select the "Make Public" button on the top-level project overview page rather than from one of the individual component overview pages.

"Make Public" button (Fig. 20) that can be used to change the setting for that individual project area. If you wish to make an entire project or multiple components from within it public at the same time, clicking "Make Public" on the main project page will provide you a menu where you can select any number of project areas to make public (Fig. 21). Once you have made your selections, you will be asked to confirm them a final time (Fig. 22)

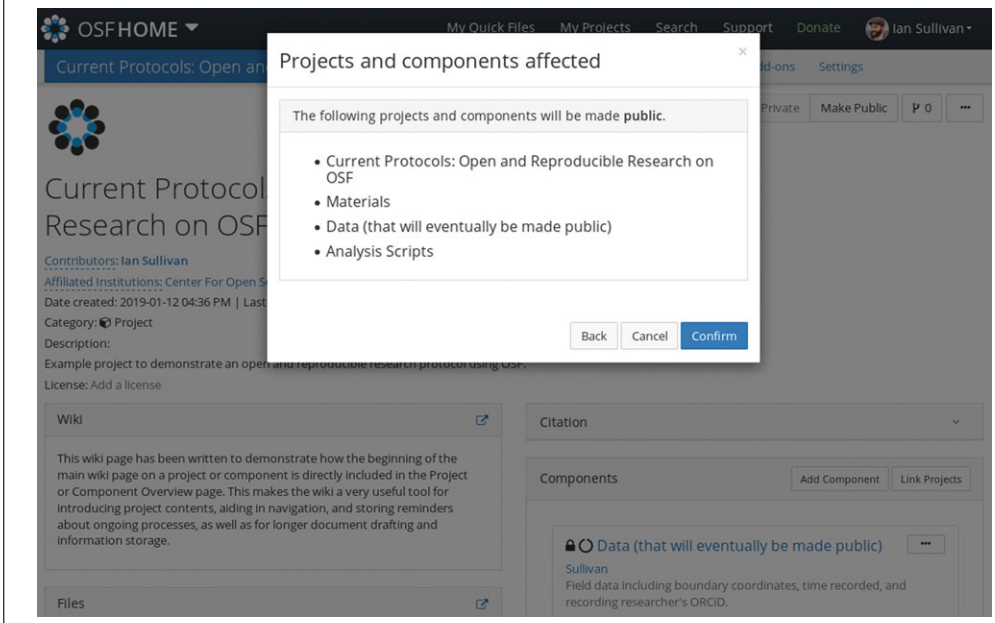

Figure 22 Final confirmation of which components will be made public. This confirmation only appears when multiple areas of a project are being made public at the same time.

before they are made public. Note that this final confirmation window will only appear if you are making multiple project areas public at once, not when changing the privacy setting on a single component.

Public materials on OSF can be read by anyone and are indexed by major search engines as well as available through OSF's internal search page: *https://osf.io/search/*. Each project, component, and file on OSF receives a short unique URL that acts as a persistent identifier for that resource. Projects and components additionally have a "Citation" tool on their overview pages (Fig. 23) that presents the list of contributors and the resource's permanent address in multiple common citation formats. Once a project or component has been made public, OSF can also create a free Digital Object Identifier (DOI) for it by using the "Create DOI" button on the Project or Component overview page (Fig. 24). Each component in a project, as well as the project itself, can be issued a separate DOI if desired.

Preprints

Once research is complete and a manuscript is ready for peer review and publication, the article is also ready for posting online to a preprint repository. Posting these preprints can help to advance the speed of research in a field (Johansson, Reich, Meyers, & Lipsitch, 2018), solicit feedback months or years earlier than the final publication (Desjardins-Proulx et al., 2013), address issues of open access equity (Tennant et al., 2016), and increase the citation count of articles when they are published (Davis & Fromerth, 2007 Serghiou & Ionannidis, 2018), among other potential advantages. The posting of some stage of an article's manuscript was, as of 2015, formally supported by 78% (Why Open Research?, *http://whyopenresearch.org/archiving*) of the peer-reviewed journals tracked by the SHERPA/RoMEO (SHERPA/RoMEO Database, *http://www.sherpa.ac.uk/romeo/index.php*) database of publisher open access policies. Support for preprints has grown since 2015, including via formal procedures for issuing Digital Object Identifiers (DOI) to preprints in 2016 (Crossref Blog: 'Members will soon be able to assign Crossref DOIs to preprints', *https://www.crossref.org/blog/members-will-soon-be-able-to-assign-crossref-dois-to-preprints/*) and the resulting standardization of the "Version of Record" metadata (Crossref Support: Posted content (includes preprints), *https://support.crossref.org/hc/en-us/articles/213126346-Posted-content-includes-preprints*) that enables authors to link together the DOIs for an article's preprint and the version later published in a journal.

Where you can and should post preprints of your work varies based on your field of research, your academic institution, and the particular journals in which you seek to publish.

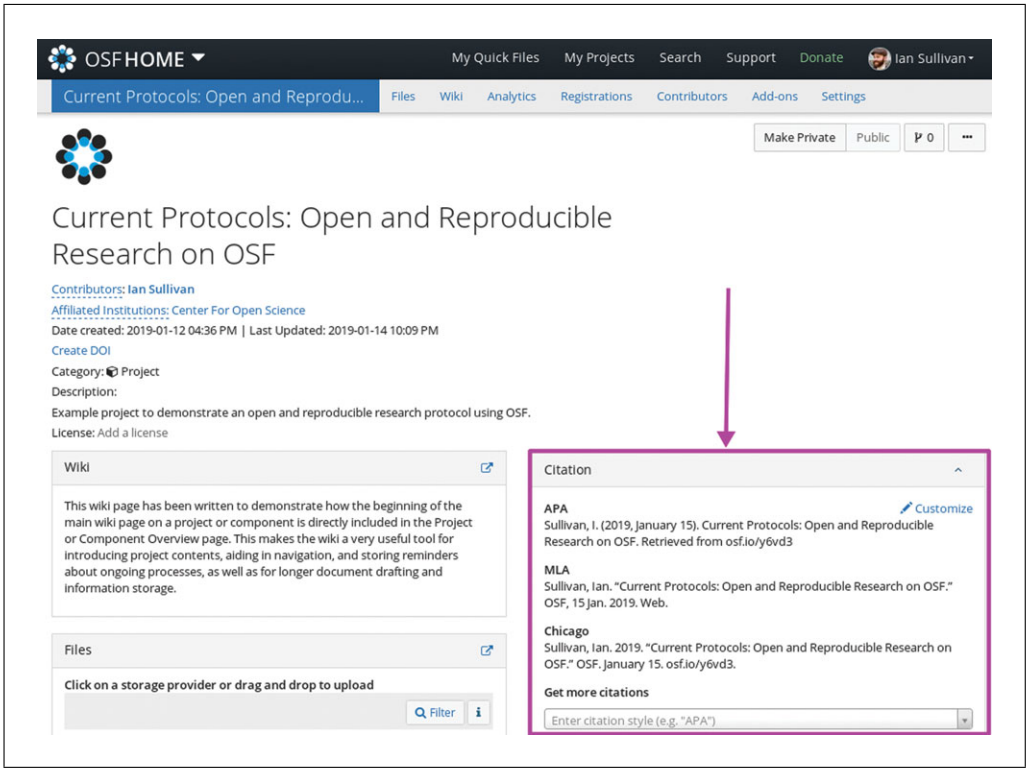

Figure 23 The Citation tool on an OSF project overview page.

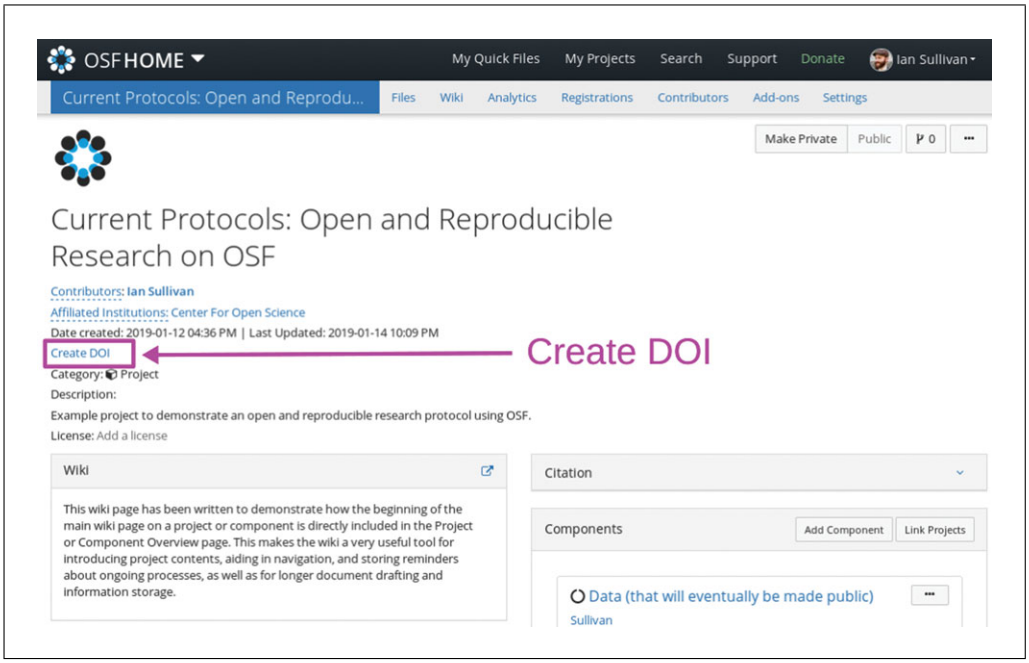

Figure 24 The "Create DOI" button on a public OSF project overview page.

Though services like SHARE (SHARE multi-site preprint index, *https://share.osf.io/discover?type[]=preprint*) exist to enable easy searching across multiple preprint services, if one particular preprint server is most commonly used by members of your research field, you may wish to post your own work there in order to make it easier to discover by the rest of your community. Most commonly preprints are posted:

- on a dedicated preprint server, either one favored by your field, such as *arXiv.org* (arXiv® e-print service, *https://arxiv.org/*), or one open to any field, such as OSF Preprints (OSF Preprints, *https://osf.io/preprints/*);

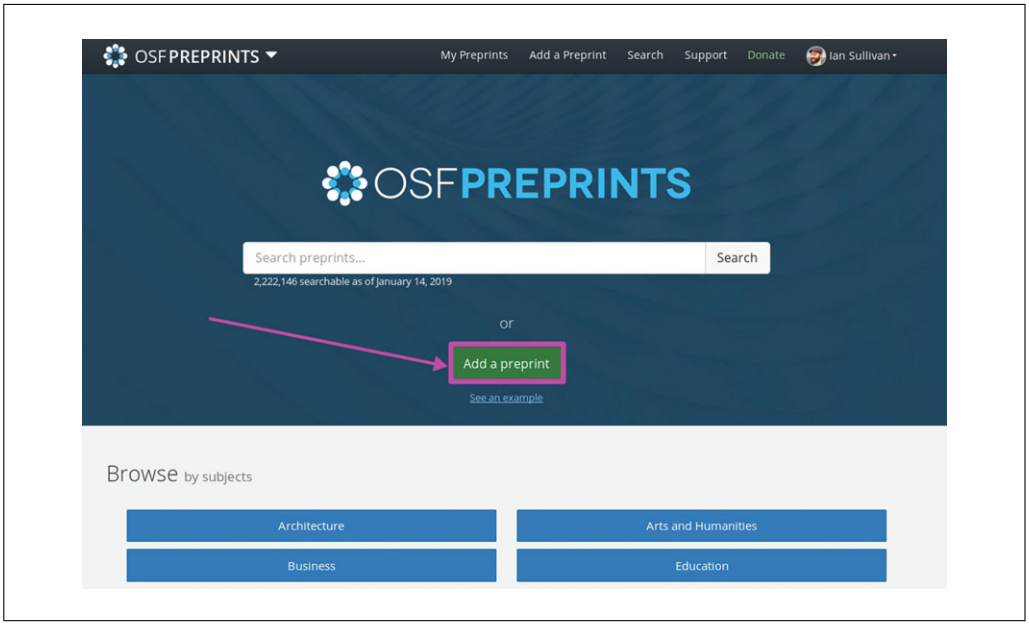

Figure 25 The "Add a Preprint" button on *https://osf.io/preprint*.

- in the local repository operated by your academic library; or
- on the researcher's personal website.

Posting a preprint on your personal site brings with it challenges in metadata, discoverability, and long-term preservation, so we encourage researchers to post to institutional repositories or dedicated preprint servers as the preferred location for their preprints and to include a link to such locations alongside any additional copies posted on a personal site.

OSF example

To post your work as a preprint on OSF, visit *https://osf.io/preprints/* and select the "Add a Preprint" button (Fig. 25; OSF Help Guide: Upload a Preprint, *http://help.osf.io/m/preprints/l/627729-upload-a-preprint*). Your first step is to choose whether to post to one of the OSF community preprint services or to the discipline-agnostic "OSF Preprints." The community preprint services each focus on a particular academic field or community of researchers. We encourage you to investigate these community services to see if one might be a fitting home for your research. The general-purpose OSF Preprints service is available if you are unsure of where to post your research, and is selected by default. At this time, individual preprints may only be posted to one of the OSF services, but all of the services are indexed and searchable through the shared OSF Preprints Search page: *https://osf.io/preprints/discover*.

After you have selected a service, you can either upload your preprint file directly or point to an existing OSF project and select the file from there. To ensure consistent rendering and long-term access, we encourage you to upload a PDF rather than a word processor file as your preprint. Once the file is uploaded, you will be asked for some basic data about the preprint and then have the opportunity to add any co-authors to the preprint. You search for your co-authors just as you would when adding them to an OSF project, by using their name or any other information they have added to their OSf user profile. If your co-authors do not all have OSF accounts, you can add them by their e-mail addresses. These addresses are used to notify all authors when the preprint is submitted and give them the ability to edit the preprint and accompanying information.

The last step in preprint submission is adding any other research outputs you would like to publish alongside your preprint. These are configured in the "Supplemental Materials" section of the form (Fig. 26), where you can either connect the preprint to an existing OSF project or create an empty project linked to the preprint so that you can upload files to it after submitting the preprint. As discussed earlier under "Sharing non-paper research outputs," we strongly encourage sharing all legally and ethically appropriate research outputs resulting from your work. If you are creating an OSF project from scratch at this step, refer to that earlier section for some

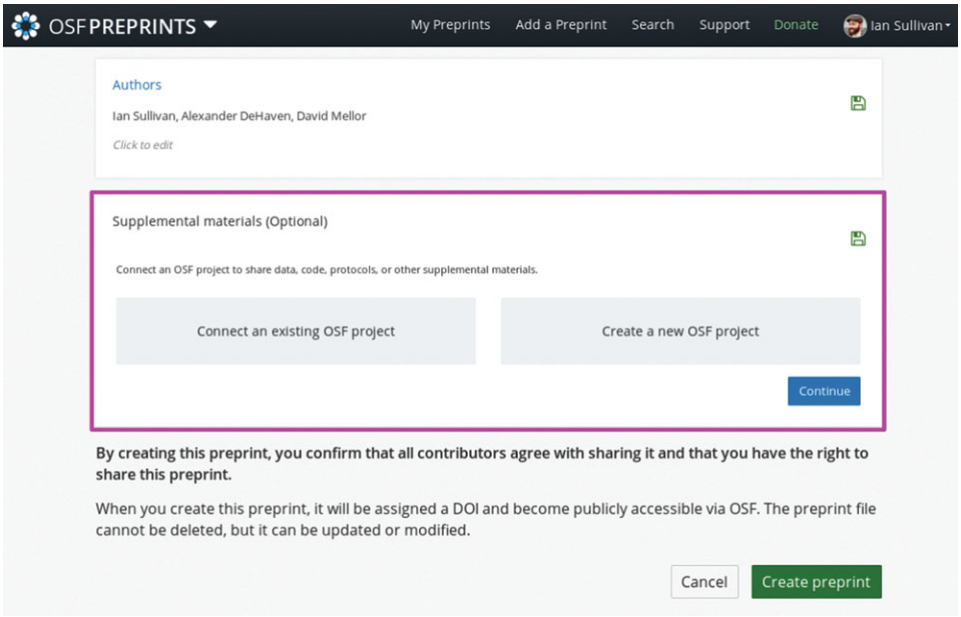

Figure 26 The "Supplemental Materials" section of the OSF Preprint submission form.

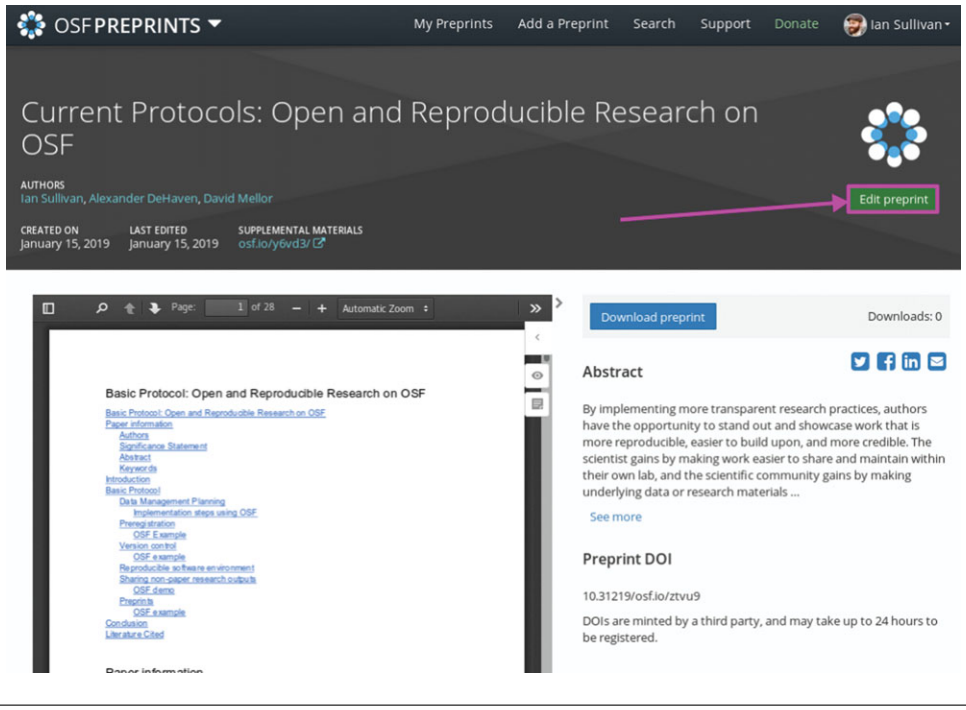

Figure 27 The "Edit" button on an OSF Preprint. This is only visible and available to authors of the preprint.

considerations about ensuring that each type of research output has unique permanent identifiers and how to incorporate such references into your manuscript.

Once your preprint has been submitted to OSF Preprints it will immediately be made public and a DOI will be created for it. Some of the community preprint services have a moderated submission process (OSF Help Guide: Submitting to a Moderated Preprint Service, *http://help.osf.io/m/preprints/l/806116-submitting-to-a-moderated-preprint-service*) that will involve a review period and possible communication with someone on the service's

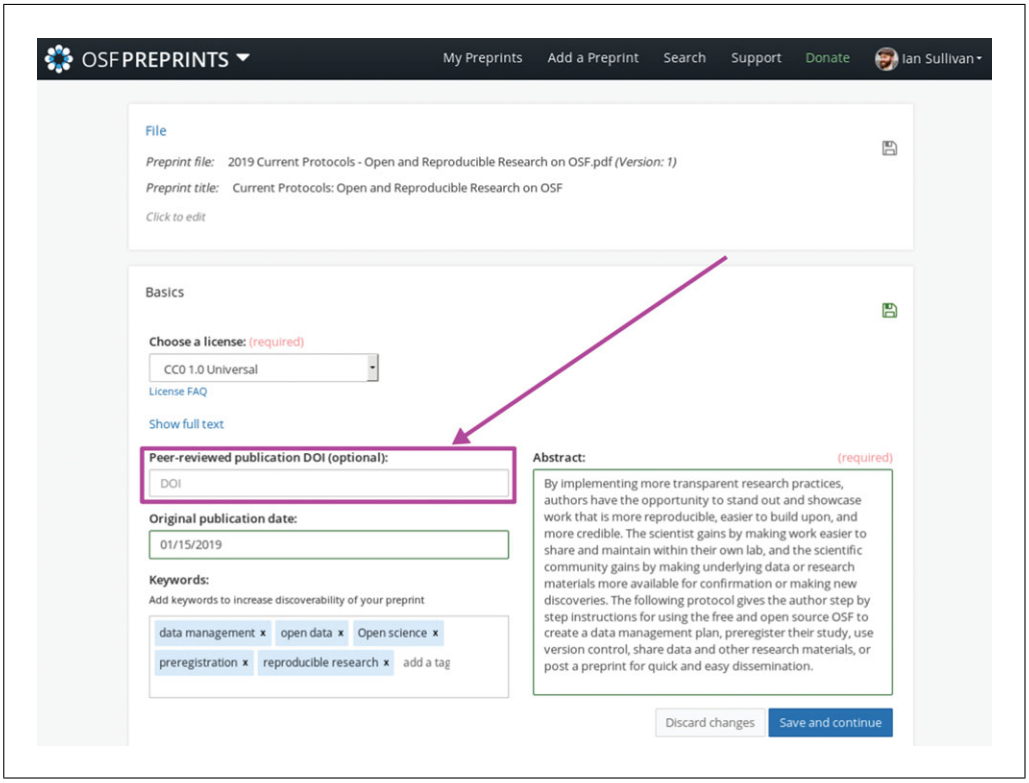

Figure 28 The "Peer-reviewed publication DOI" field in an OSF Preprint's basic metadata section.

review team before your preprint is made public.

If needed, preprints can be edited by their authors using the "Edit Preprint" button (Fig. 27). This enables updating both the content of a preprint and the metadata around it. Just as with files on OSF, all versions of a preprint file are stored, and viewers of the preprint page can access old versions (OSF Help Guide: Downloading Previous Versions of Preprints, *http://help.osf.io/m/preprints/l/983227-download-previous-versions-of-preprints*). Editing metadata is particularly important if your preprint goes on to be published in a journal or other peer-reviewed publication and your work is issued a second DOI. In those cases, one of the preprint authors should return to the preprint and edit the "Peer-reviewed publication DOI" field in the "Basics" metadata section (Fig. 28) to include the new DOI. This helps avoid confusion in the scholarly record and ensures that any visitors to your preprint page know that a peer-reviewed version exists.

CONCLUSION

The traditional scientific article represents a summary of the substantial work that occurred leading up to the final results. Both the desire to tell the most concise and exciting story possible and the biases toward significant findings over null results result in an incomplete reporting of scientific knowledge. Open science practices offer a variety of benefits to science and to scientists. Scientists benefit through improved documentation, a defense against unintentional false positive inflation in research results, and an opportunity for increased citations. Science benefits by allowing greater collaboration, replication, and scrutiny.

A common first reaction to seeing the spectrum of open science practices is to be overwhelmed. It may seem daunting to implement all of the above, which it is. However, *any* of the above-mentioned tools will provide the scientist with the means to stand out through better and more transparent workflows. Most authors will implement one or two of the practices when starting on their next project after first hearing about open science. An easy first step might be to attempt to share an analysis script or a questionnaire that would normally not be part of the article. Some authors are keen to share data or to even preregister their work. Whatever your particular interests are, starting with even one additional open science practice will allow you to better report your work.

LITERATURE CITED

Arnold, B. F., van der Laan, M. J., Hubbard, A. E., Steel, C., Kubofcik, J., Hamlin, K. L., . . . Lammie, P. J. (2017). Measuring changes

in transmission of neglected tropical diseases, malaria, and enteric pathogens from quantitative antibody levels. *PLOS Neglected Tropical Diseases*, *11*(5), e0005616. doi: 10.1371/journal.pntd.0005616.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. doi: 10.1038/533452a.

Chambers, C., Munafò, M., & more than 80 signitories. (2019). Trust in science would be improved by study pre-registration. The Guardian, Available at https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, *71*(2), 203–215. doi: 10.1007/s11192-007-1661-8.

Desjardins-Proulx, P., White, E. P., Adamson, J. J., Ram, K., Poisot, T., & Gravel, D. (2013). The case for open preprints in biology. *PLoS Biology*, *11*(5), e1001563. doi: 10.1371/journal.pbio.1001563.

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings—a practical guide. *Biological Reviews*, *92*(4), 1941–1968. doi: 10.1111/brv.12315.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. doi: 10.1126/science.1255484.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi: 10.1371/journal.pmed.0020124.

Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLOS Biology*, *13*(11), e1002295. doi: 10.1371/journal.pbio.1002295.

Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLOS Medicine*, *15*(4), e1002549. doi: 10.1371/journal.pmed.1002549.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. doi: 10.1038/s41562-016-0021.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. doi: 10.1177/1745691612459058.

Nosek, Brian A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606. doi: 10.1073/pnas.1708274114.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. doi: 10.1126/science.aac4716.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *Peer Journal*, *1*, e175. doi: 10.7717/peerj.175.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi: 10.1037/0033-2909.86.3.638.

Saltzer, J. H. (1974). Protection and the control of information sharing in multics. *Communications of the ACM*, *17*(7), 388–402. doi: 10.1145/361011.361067.

Serghiou, S., & Ioannidis, J. P. A. (2018). Altmetric scores, citations, and publication of studies posted as preprints. *JAMA*, *319*(4), 402–404. doi: 10.1001/jama.2017.21168.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi: 10.1177/0956797611417632.

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. J. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review. *F1000Research*, *5*, 632. doi: 10.12688/f1000research.8460.3.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. doi: 10.1038/sdata.2016.18.

INTERNET RESOURCES

https://www.dataone.org/best-practices

Best Practices | DataONE. (2019 January 11). Retrieved January 11, 2019.

https://osf.io/qpxv2/

Butler, B. (2017 February). Licensing your Research. Video.

https://cos.io/blog/preregistration-plan-not-prison/

DeHaven, A. (2017 May 23). Preregistration: A Plan, Not a Prison. Retrieved January 12, 2019.

https://data.consilium.europa.eu/doc/document/ST-1216-2018-INIT/en/pdf

European Research Area and Innovation Committee. (2018. December 18). Recommendations by the ERAC Standing Working Group on Open Science and Innovation (SWG OSI) on open science and innovation.

http://help.osf.io/m/faqs/l/726460#Backup-Preservation-Policy

FAQs | FAQS | OSF Guides. (2019 January 12). Retrieved January 12, 2019.

https://doi.org/10.5281/zenodo.2530961

Steeves, V. (2017). Introduction to research data management.

https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Strasser. C., Cook, R., Michener, W., & Budden A. (2012). Primer on data management: What you always wanted to know. A DataONE publication, available via the California Digital Library.