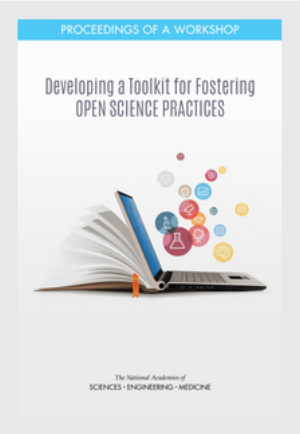

Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop (2021)

DETAILS

120 pages | 6 x 9 | PAPERBACK

ISBN 978-0-309-09361-3 | DOI 10.17226/26308

CONTRIBUTORS

Thomas Arrison, Jennifer Saunders, and Emi Kameyama, Rapporteurs; Committee on Developing a Toolkit for Fostering Open Science Practices: A Workshop; Board on Research Data and Information; Policy and Global Affairs; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine. 2021. *Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26308.

Appendix C

Toolkit Elements

This appendix includes examples of draft elements of a toolkit that have been developed by members of working groups of the National Academies of Sciences, Engineering, and Medicine's Roundtable on Aligning Incentives for Open Science. The following materials were developed to stimulate discussions at the November 5, 2020, workshop on Developing a Toolkit for Fostering Open Science Practices:

I. **Open Science Imperative**. This essay communicates the benefits of open science using approachable language.
II. **Open Science Signaling Language Template and Rubric**. These resources provide specific language that can be adapted and adopted to signal an organization's interest in open science activities at specific points of high leverage (e.g., grant applications, job postings).
III. **Good Practices Primers**. These concise guides offer policy makers a high-level overview of open sharing.
IV. **Open Science by the Numbers Infographic**. This infographic communicates the benefits of open science in a graphic form.
V. **Open Science Success Stories Database**. This database compiles research articles, perspectives, case studies, news stories, and other materials that demonstrate the myriad ways in which open science benefits researchers and society alike.

VI. **Reimagining Outputs Worksheet**. This table enumerates the range of research products stakeholders may choose to consider as they develop open science policies.

The toolkit is primarily intended to assist university leadership, academic department chairs, research funders, learned societies, and government agencies about how such a toolkit might be used, what additional materials are needed, and how such a toolkit should be disseminated for broad adoption. As a result of the workshop, a few sections in the Open Science Imperative and Good Practices Primers have been revised by the working group authors.

III. GOOD PRACTICES PRIMERS[4]

Nicholas Gibson, John Templeton Foundation
Jerry Sheehan, National Institutes of Health
Stuart Buck, Formerly, Arnold Ventures
J. C. Burgelman, Vrije Universiteit Brussel
Anne-Marie Coriat, Wellcome
Anne Koralova, Helmsley Trust
Heather Pierce, Association of American Medical Colleges
Dawid Potgieter, Templeton World Charity Foundation
Greg Tananbaum, Open Research Funders Group

Many organizations, particularly those that perform or fund research, are in the information-gathering stage with respect to their open science policies and practices. These concise primers are intended to provide decision makers with a high-level overview of the *what's* and *how's* of open sharing of various research outputs. Each primer (1–2 pages) addresses a different output type, delving into exemplars, dependencies, resourcing, and a range of other considerations. The following drafts provide a sense of what the primers will encompass. They do not provide a detailed rationale for adopting an open science policy, an analysis of the barriers, or a comprehensive guide to implementation, including the pros and cons of various approaches.

ARTICLES

Relevance to Open Ecosystem

Unrestricted access to, and reuse of, published journal articles benefits the research community by facilitating the dissemination of new information, thus maximizing opportunities for that work to lead to new insights and discoveries.

Considerations

Among the key issues that organizations will wish to address in developing a policy to make articles open are the following:

[4] The views expressed are those of the authors and do not necessarily reflect the official policies or positions of their employing organizations.

- ***Fulfillment.*** Can researchers adhere to the policy by publishing in a fully open access journal, a "hybrid" journal (a subscription-based journal that allows authors to make individual articles open access immediately on payment of an article publication charge), or by posting a copy of a paper in an open, trusted repository? If the latter is permissible, must a certain version (e.g., version of record, approved manuscript) be posted?
- ***Timing.*** Does the policy require that the articles be made openly available immediately, or is some embargo (e.g., 6 months) permissible?
- ***Financial Support.*** Will the policy maker provide funding to defray costs of open access (e.g., article processing charges)? If so, is there a cap on the amount? Must the researcher explicitly account for these expenses at the time of project design? Is there a mechanism for the researcher to have such costs covered after grant close?
- ***Discoverability.*** How will potential readers discover the openly available content? Will it be picked up by major indexing services or be made available in leading disciplinary repositories?
- ***Licensing and Reuse.*** What type of licensing requirements will the policy include to facilitate reuse? Free to read, preferably permanent, is often the primary focus of open access policies, but reuse considerations (including, but not limited to, text and data mining) also merit consideration.

Approaches

The practical implementation of a policy requiring access to published articles can take different forms (see Box 1). Some policies require publication in an open access journal or a hybrid journal. This can introduce a modest restriction on researchers' choice of publication venue, although thousands of journals are open access or offer a hybrid option.

Some policies promote deposit of a copy of the paper (which may not be the final, formatted version, depending on publisher or funder requirements) in a trusted repository. As virtually all journals allow some form of self-archiving, this approach places fewer restrictions on authors (see Box 2). It does require authors to proactively identify and deposit the paper in an appropriate repository. Some journals will, however, deposit articles or final submitted manuscripts in a selected repository on behalf of authors.

SPARC (Scholarly Publishing and Academic Resources Coalition) maintains a succinct resource for tracking, comparing, and understanding

BOX 1
Examples of Open Access Policies Requiring Publication in Open Access Journals

The Bill & Melinda Gates Foundation and the Wellcome Trust require funded researchers to publish their articles in open access journals, with no embargo period.[a] The option to publish in hybrid journals is being phased out by both organizations in 2021.

[a] See https://www.gatesfoundation.org/how-we-work/general-information/open-access-policy and https://wellcome.org/news/wellcome-updates-open-access-policy-align-coalition-s.

BOX 2
Examples of Self-Archiving Open Access Policies

- All U.S. federal science funding agencies require submission of the author's final manuscript or final published article to a designated repository such as PubMed Central, with public access provided no later than 12 months after publication.[a]
- Harvard University is among the many universities that asks faculty to deposit a version of their articles ("the accepted author manuscript") in Harvard's institutional repository.[b]
- The Academic Senate of the University of California adopted a systemwide open access policy in 2013 designed to make research articles authored by faculty available to the public at no charge.[c]

[a] See https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

[b] See https://osc.hul.harvard.edu/policies.

[c] See https://osc.universityofcalifornia.edu/for-authors/open-access-policy.

U.S. federal funder article-sharing policies;[5] ROARMAP (Registry of Open Access Repository Mandates and Policies) provides similar information about funders and universities;[6] and the federal interagency group CENDI posts information about federal agency public-access policies.[7] These sites can be used to compare and contrast different approaches that stakeholders are taking to open access policies.

Resourcing

Once open policies are implemented, organizations can undertake a range of activities to manage them. At the low-touch end of the spectrum, organizations can require researchers to document how they intend to comply. Depending on internal resources, some organizations spot-check these plans, while others simply rely on the honor system. Other organizations take a more engaged approach, requiring proof of compliance from researchers and checking this against internal expectations and guidelines. Additionally, funders are increasingly able to rely on emerging research infrastructure such as author and funder registries to automate aspects of the reporting process. Organizations without open policies may view administration and compliance as daunting tasks. However, each organization can make its own appropriate determination about the resources it is able to devote to these activities. Compliance monitoring can often be embedded within other regular research-reporting processes without adding significant burden on researchers or administrative staff.

Next Steps

The Open Research Funders Group (ORFG) can provide support and insight into best practices and available resources.[8] The ORFG Incentivization Blueprint provides model language that can be adapted and adopted by funders and other organizations.[9] It offers a stepwise approach to deploying a policy that can grow to encompass not only open access articles but also data, code, and other research outputs.

[5] See http://researchsharing.sparcopen.org/articles.
[6] See https://roarmap.eprints.org.
[7] See https://www.cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html.
[8] See http://www.orfg.org.
[9] See http://www.orfg.org/incentivization-blueprint.

DATA

Relevance to Open Ecosystem

The ability to independently confirm results and conclusions is critical for evaluating scientific rigor and informing future research activities. Openly shared data can support reanalysis and confirmation of research findings. They can also shed light on research that is not published, which can occur when tested hypotheses are not confirmed or research is considered unproductive, thereby mitigating publication bias and improving the efficiency of the research process, and can lead to novel lines of inquiry. In particular, shared data can be reused for new analyses, whether independently or in combination with other data.

Considerations

Several issues merit consideration by organizations developing open data policies, including the following:

- ***Scope.*** What data are needed for the independent verification of research results? Which data are most valuable to preserve for reuse? What is the appropriate balance between making available large volumes of raw data versus smaller amounts of more processed data?
- ***Metadata.*** What documentation and descriptive details are necessary to allow others to use the data properly and without confusion? How does the policy ensure that information about the methodology and procedures used to collect the data, details about codes, definitions of variables, variable field locations, frequencies, and the like are properly collected and disseminated? Are there disciplinary-specific metadata schemas that should be used to facilitate discovery and reuse?
- ***Timing***. Starting with the baseline expectation that data underlying reported results will be made available concurrent with the posting of research findings, are there legitimate exceptions? Should researchers be given a period of exclusivity to analyze research data additional to those directly supporting reported findings before sharing them with the community? If data are

not reported in a publication, what is an appropriate time line for sharing the data?

- ***Financial Support.*** Who will provide funding to defray costs of preparing and/or depositing the data? What costs are recoverable? If so, is there a cap on the amount? Must the researcher explicitly account for these expenses at the time of project design?
- ***Licensing.*** What type of licensing requirements will the policy include to facilitate reuse of the data?
- ***Proprietary Software.*** To the extent that the data can only be accessed or analyzed through software that is not open source, what steps can be taken to reduce restrictions on its reuse?
- ***Data Management Plans.*** What support and guidance will the organization provide to help the researcher clearly articulate at the outset of a project what, how, and where data will be shared? What mechanisms are in place to ensure that the researcher adheres to the data management plan?
- ***Data Standards.*** For the study type in question, or for the field in which the work is centered, are there best practices for how the data should be formatted, to enable wider and more efficient reuse and interoperability?
- ***Preservation.*** What constitutes an appropriate deposit location for the data? Is there a repository that is appropriate for the subject matter in question, and/or has emerged within a specific research community as the default resource in that field? Is the repository secure, stable, and open for all to access?
- ***Discoverability.*** How will data be discoverable? Even if it is deposited in a particular repository, how will other possible users know where to look? Will the data be assigned a unique persistent identifier, and will that identifier be promulgated through related publications?
- ***Privacy/Confidentiality.*** Some datasets may contain human subject details that cannot be fully disseminated, due to the Health Insurance Portability and Accountability Act of 1996 (HIPAA) (Public Law 104-191; 104th Congress), the Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 C.F.R. Part 99), the European Union's General Data Protection Regulation 2016/679 (EU GDPR) (O.J. L. 119, 04.05.2016; cor. O.J. L. 127, 23.5.2018), or other privacy restrictions. Such datasets, however, can often be shared after anonymization or deidentifica-

tion techniques (including adding statistical noise, suppression of small cells, etc.), or under protected mechanisms such as a virtual data warehouse accessible only with a confidentiality agreement in place. How will such datasets be handled in a way that maximizes sharing while protecting privacy? Can analytic opportunities be made openly available while the confidential aspects of the data remain restricted?

- ***Compliance monitoring.*** How can compliance with data management and sharing requirements/expectations be easily monitored, for example, by funders, other institutions, or individuals?

Approaches

One common approach to facilitate data sharing is to develop policies requiring data to be findable, accessible, interoperable, and reusable, that is, to meet the Findability, Accessibility, Interoperability, and Reuse of digital assets (FAIR) data principles. While data can be FAIR without necessarily being publicly open, the FAIR principles broadly support open science. Specific definitions and operationalizations of each of these principles, together with practical guidance on how to satisfy each requirement, have been prepared by the GO FAIR Initiative.[10] To render data FAIR, metadata and datasets should be prepared in a standardized, descriptive manner that makes it easier for both humans and machines to find and use.

With respect to data accessibility, a common rule of thumb in the open science community is that data should be shared in a manner that promotes reuse and transparency while recognizing that certain safeguards may be required to protect sensitive information that could compromise subject privacy or other norms and regulations. While the default position needs to shift to "open," legitimate restrictions on access need to be taken into account.

Many U.S. federal science agencies require researchers to submit a data management plan either as part of a grant application or before issuing an award. These plans provide general information about the types of data to be collected in a research study, the repository into which they will be deposited, and the time lines and other conditions of access. For certain types of research studies, federal science agencies have developed more

[10] See https://www.go-fair.org/fair-principles.

specific guidance or requirements (see the National Institutes of Health [NIH] example in Box 3).

Some organizations, such as the National Science Foundation, provide a general set of guidelines on data sharing, articulating to researchers that they are expected to share their data with their peers under reasonable cir-

BOX 3
Examples of Open Data Policies

- The National Institutes of Health (NIH) issued a Data Management and Sharing Policy that applies to all data generated by funded research[a] as well as specific policies that apply to genomic data, clinical trial data, and other specific research programs and data types.[b] NIH has also provided information for selecting a data repository.[c]
- The American Heart Association requires grant applicants to include a data-sharing plan as part of the application process. Any research data that are needed for independent verification of research results must be made freely and publicly available within 12 months of the end of the funding period (and any no-cost extension).
- The European Open Science Cloud (EOSC) has developed a strategic implementation plan for the creation of a data commons housing interoperable, machine-readable data across domains, consistent with FAIR (findable, accessible, interoperable, and reusable) principles.[d]
- The Yale University Open Data Access (YODA) Project facilitates clinical trial data access to promote independent analyses of the data. It also provides a formal vetting of the data to ensure consistency with informed consent and confidentiality requirements.[e]

---

[a] See https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

[b] For genomic data, see https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing; for clinical trial data, see https://grants.nih.gov/policy/clinical-trials/reporting/understanding/nih-policy.htm; and for other specific research programs and data types, see https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html.

[c] See https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html.

[d] See https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en.

[e] See https://yoda.yale.edu.

cumstances.[11] Others, such as the NIH, have overarching data management and sharing policies that apply to all funded research, while also having more focused policies that provide explicit guidance as to the timing, licensing, and dissemination of data of particular types (e.g., genomic data) or associated with particular research programs (e.g., the Cancer Moonshot).[12]

Resourcing

For data specifically, it is important to ensure that appropriate metadata and documentation are provided so that datasets are properly contextualized. Organizations will also benefit from in-house or outsourced expertise to assess the appropriateness of data management plans and informed consents, to ensure these allow data sharing to the extent that the organization desires.

Once open policies are implemented, organizations can undertake a range of activities to manage them. At the low-touch end of the spectrum, organizations can require researchers to document how they intend to comply. Depending on internal resources, some organizations spot-check these plans, while others simply rely on the honor system. Other organizations take a more engaged approach, requiring proof of compliance from researchers and checking this against internal expectations and guidelines. Additionally, funders are increasingly able to rely on emerging research infrastructure, such as author and funder registries, to automate aspects of the reporting process. Organizations without open policies may view administration and compliance as daunting tasks. However, each organization can make its own appropriate determination about the resources it is able to devote to these activities.

Next Steps

There are a range of resources that can contribute to a detailed understanding of policy options and approaches, including the following:

- GO FAIR provides a starter kit with a wealth of information on data management plans, license options, and repositories.[13]

[11] See https://www.nsf.gov/pubs/policydocs/pappg19_1/pappg_11.jsp#XID4.
[12] See https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/funding/public-access-policy#requirement.
[13] See https://www.go-fair.org/resources/rdm-starter-kit.

- The Transparency and Openness Promotion (TOP) Guidelines provide sample language for three levels of open data policies.[14] This wording can be adapted and adopted to suit the specific circumstances of various organizations.
- The Open Research Funders Group Incentivization Blueprint offers sample open data policy language that can be adapted for a range of use cases.[15]
- The American Heart Association's website contains a detailed FAQ page that articulates questions commonly asked by researchers subject to an open data policy.[16]
- The DMPTool site is an excellent resource for both browsing the data policies of hundreds of organizations and generating data management plans to fit a range of requirements and circumstances.[17]
- NIH is developing various resources to assist researchers in complying with its Data Management and Sharing Policy, including clarifications about the contents of a data management and sharing plan, selection of data repositories, and allowable costs.[18]

[14] See https://osf.io/bcj53.

[15] See http://www.orfg.org/incentivization-blueprint.

[16] See https://professional.heart.org/en/research-programs/aha-research-policies-and-awardee-hub/open-science-frequently-asked-questions#:~:text=The%20AHA%20open%20data%20policy%20requires%20any%20data%20needed%20for,but%20the%20most%20exceptional%20circumstances.

[17] See https://dmptool.org.

[18] See https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

PROTOCOLS AND PREREGISTRATION ANALYSIS PLANS

Relevance to Open Ecosystem

Unreported flexibility in data analysis can reduce the credibility of reported results and invalidate common tools of statistical inference. By submitting a detailed study protocol and statistical analysis plan to a public registry prior to conducting the work (i.e., preregistering with an analysis plan), the scientist makes a clearer distinction between planned hypothesis tests (i.e., confirmatory tests) and unplanned discovery research (i.e., screening or exploratory research). Preregistration of laboratory protocols—detailed descriptions of the methods used in the experiment, including equipment and reagents—is becoming more common and facilitates replicability. Preregistration is particularly important for studies that make an inferential claim from a sampled group or population, as well as studies that are reporting and testing hypotheses. After a project is completed, protocols and preregistration analysis plans can be used in conjunction with the final study and analysis by researchers seeking to replicate, reproduce, and build upon findings.

Considerations

- ***Scope.*** Should preregistration address the study protocol (how a study or experiment will be conducted), the laboratory protocol (detailed description of methods), the analysis plan (how the collected data will be organized and evaluated), or all three? Of primary interest in ensuring the integrity of the research outcome is information about the prespecified outcome measures/endpoints. However, decisions made during analysis can also affect the integrity of the reported findings, so many registries encourage preregistration of both.
- ***Documentation.*** Should preregistration include disclosure of the full-study protocol or just summary information about the protocol and statistical analysis plan? Submission of summary information can be more time consuming, but it also allows for structured data entry to facilitate searching and cross-study comparison. If a summary is submitted, then what specific information needs to be provided?
- ***Data Privacy.*** Protocols and analysis plans can contain proprietary or other protected information (e.g., names of study person-

nel). To what extent can information be redacted without undermining the benefits of access? The desire to promote meaningful preregistration must be balanced against the provision of necessary protections/redactions of information.

- ***Deposit Location.*** Where and how should a scientist register their protocol and/or analysis plan? There are a limited number of established public repositories. For clinical trials of health-related interventions, NIH's ClinicalTrials.gov is the default system.[19] Within the social, behavioral, and preclinical sciences, the Open Science Framework is becoming a default registry.[20] Some public repositories tend to be disciplinarily focused.
- ***Timing.*** How long before or after a study begins must it be registered? When should a preregistration be updated? Earlier may be better, but additional information may be needed about its status (e.g., has Institutional Review Board approval been received). The timing of an update is also linked to the degree to which a change has implications on the full preregistration (e.g., challenges in recruiting a full sample may necessitate moving from a single cohort to a multicohort design). Protocols shared at study initiation can more clearly establish a project's aims and plan. Does the registry support time-stamped versioning?
- ***Discoverability.*** Are preregistrations automatically made public after a fixed period of time? Does the registry support public searches for preregistrations?
- ***Scope.*** To date, the majority of registries are for causal impact studies, typically carried out either in a small-scale experiment or a large randomized clinical/field trial. However, there may be a strong rationale to consider preregistering exploratory studies at the time of funding or at the beginning of a study so as to capture strong theory-driven exploratory questions as opposed to post hoc "fishing" analyses.
- ***Results.*** To what extent should a funder require the ultimate posting of a study's results in a way that can be compared to whatever was preregistered? Federal law requires the posting of results at ClinicalTrials.gov for certain clinical trials; should this be a broader expectation?

[19] See https://clinicaltrials.gov.
[20] See https://osf.io.

Approaches

There are a range of different preregistration locations available, primarily driven by discipline. All NIH-funded clinical trials and most clinical trials of Food and Drug Administration (FDA) regulated drugs, biologics, and devices must be preregistered at NIH's ClinicalTrials.gov not later than 21 days after first recruitment. Summary information is provided in a highly structured format. Final protocols for NIH-funded clinical trials and most FDA-regulated clinical trials of drugs, biologics, and devices must be submitted to NIH's ClinicalTrials.gov as part of summary data reporting after a trial has completed. These policies also require that the statistical analysis plan be submitted if it is not considered part of the protocol. (See Box 4 for examples of preregistration and protocols policies.)

BOX 4
Examples of Preregistration and Protocols Policies

- The Chan Zuckerberg Initiative (CZI) requires grantees to make laboratory protocols publicly available and has nurtured dedicated protocol communities of CZI-funded investigators.[a]
- The American Economic Association (AEA) encourages researchers to register their randomized controlled trials (including research designs and analysis plans) in the AEA Randomized Controlled Trials (RCT) Registry.[b]
- CHDI Foundation has established an Independent Statistical Standing Committee (ISSC) to provide unbiased evaluation and expert advice on developing protocols and statistical analysis plans, and evaluation of prepared study protocols.[c]
- Arnold Ventures requires all funded empirical studies that involve statistical inference to be preregistered before the start of intervention or data collection on the Open Science Framework.[d]

[a] See https://cziscience.medium.com/power-to-the-protocols-388fe92001be and https://www.protocols.io/workspaces/neurodegeneration-method-development-community1.

[b] See https://www.socialscienceregistry.org.

[c] See https://chdifoundation.org/independent-statistical-standing-committee.

[d] See https://www.arnoldventures.org/guidelines-for-investments-in-research and https://osf.io.

Other disciplines have their own community-promoted repositories. Researchers carrying out causal studies in education have the opportunity to preregister their work in the Registry of Efficacy and Effectiveness Studies.[21] Researchers in the social, behavioral, and cognitive sciences often use the Open Science Framework platform.[22] The Registry for International Development Impact Evaluations hosts impact evaluations related to development in low- and middle-income countries.[23]

Resourcing

Organizations considering preregistration will need to consider whether resources are needed to support a preregistration repository for collecting preregistration reports and protocols. It is also important that there is a transparent link among any disseminated findings (preprints, articles, etc.), data, and preregistrations to determine whether there are significant deviations from the intended analysis.

Organizations and publishers will also need to ascertain how to indicate where preregistration records and protocol information exist for a published article. Multiple publishers and other organizations offer modalities for publishing study protocols, laboratory protocols, and registered reports. To be most effective, preregistrations and protocols should be closely linked to associated publications and other study information so that they can be easily discovered and accessed by those examining the study results.

Next Steps

The TOP Guidelines provide sample language for three levels of policies for study preregistration and analysis plan preregistration.[24] This wording can be adapted and adopted to suit the specific circumstances of a range of organizations. The TOP recommendations include (1) disclosing whether work was preregistered or not, (2) verifying that any preregistered work adheres to the prespecified plans, and (3) requiring preregistration for relevant research studies (typically inferential and hypothesis-testing work).

[21] See https://sreereg.icpsr.umich.edu/sreereg.
[22] See https://osf.io/prereg.
[23] See https://ridie.3ieimpact.org.
[24] See https://osf.io/bcj53.

The Center for Open Science provides multiple resources on how to preregister studies and analytic plans, including templates.[25] NIH provides a number of resources to facilitate the development of protocols, including the National Institutes of Health e-Protocol Writing Tool and protocol templates for clinical trials and behavioral/social science research.[26]

[25] See https://www.cos.io/initiatives/prereg and https://osf.io/zab38/wiki/home/?view.
[26] See https://e-protocol.od.nih.gov/#/home and https://grants.nih.gov/policy/clinical-trials/protocol-template.htm.

REGISTERED REPORTS

Relevance to Open Ecosystem

Peer review of study protocols with analysis plans, along with dissemination of findings regardless of outcome, addresses publication bias against null results. It also provides the benefits of preregistration by making a clearer distinction between hypothesis tests and discovery research. By submitting funded studies to journals as a registered report, the scientist improves study planning, increases study rigor, and improves scientific credibility. Funders who support this process anticipate that peer-review feedback could change study processes that result in budget changes and are prepared to consider such amendments in response to journal reviewer feedback. Funders can also partner with journals to coordinate review for funding and publishing decisions.

Considerations

- ***Scope.*** Registered reports are most appropriate for specific experiments or studies, not for grants that fund a research program over several years. Such grants could still include one or more registered reports, but it would likely not cover the entire program.
- ***Research Scope.*** Registered reports are best for studies that test hypotheses and in disciplines that could suffer from publication bias (typically against null results). Registered reports are not appropriate for purely exploratory or discovery science, until those studies are ready to use traditional hypothesis tests.
- ***Timing.*** By design, registered reports include additional time at the beginning of a project. Project plans should account for this. Additional time devoted to peer review in the early stages of the project is also required to ensure that the study methods are as rigorous as possible and that results will be disseminated regardless of outcome.

Approaches

There are a number of ways in which an organization can promote registered reports. On the low end of engagement, a funder or agency can ask grantees to specifically state whether all or part of the work would be

BOX 5
Examples of Funders Encouraging/ Requiring Registered Reports

- The Flu Lab is partnering with PLOS and the Center for Open Science to promote replications and registered reports of influenza research.
- Cancer Research UK is collaborating with the journal *Nicotine & Tobacco Research* on an integrated review process for grant proposals and preregistered reports.[a]

[a] See https://academic.oup.com/ntr/article/19/7/773/3106460.

appropriate for a registered report. This will remind grantees that registered reports are a valued addition to a proposed study. Principal investigators can be encouraged to notify their communities—via social media, their websites, CVs, and other appropriate channels—when their precollection hypotheses and data analysis plans have been reviewed and registered. Organizations may also wish to educate researchers on the benefits of registered reports, particularly researchers in domains where the practice is not currently widespread.

For specific grants, programs, or initiatives where projects are appropriate for the format, agencies and funders may elect to make registered report submissions to a journal before data collection a requirement. If a study does not receive an in-principle acceptance offer from a journal, the plan can still be preregistered by the authors on a platform like the Open Science Framework and submitted for publication after the study is completed.

Some funders are partnering directly with discipline-appropriate journals to integrate the registered reports model in the grant application process. One example is the Children's Tumor Foundation,[27] which is partnering with the journal *PLOS ONE* to concurrently evaluate grant proposals and the ethics and rigor of the experimental design. Accepted proposals will simultaneously receive both funding and a commitment to publication of the study results in *PLOS ONE*. (See Box 5.)

[27] See https://grants.nih.gov/policy/clinical-trials/protocol-template.htm.

Resourcing

Given the relative novelty of registered reports, organizations may need to educate grantees about the merits and mechanics of this approach. Organizations that seek to integrate grant proposals and registered reports will need to establish a review process that allows for independent evaluation of the latter along a timescale and workflow that supports the former. This may also require negotiation of a direct partnership with a journal or publisher.

Absent this type of embedded relationship, researchers may require guidance to evaluate the growing number of journals that accept and publish registered reports. The Comparison of Registered Reports site provides an interactive tool to assist in this process.[28] Policies that require registered reports will also require some form of monitoring, ranging from spot-checking to soliciting proof of compliance.

Next Steps

The Center for Open Science provides a comprehensive registered reports resource,[29] including FAQs, workflow suggestions, and other foundational materials. The Center for Open Science also provides a simple Q&A tutorial to assist authors in the drafting of registered reports.[30] The Open Science Framework provides a searchable database of registered reports across a range of disciplines.[31] These may offer useful guidance to better understand the core elements of a well-constructed registered report.

[28] See https://katiedrax.shinyapps.io/cos_registered_reports.

[29] See https://www.cos.io/initiatives/registered-reports.

[30] See https://osf.io/93znh/?_ga=2.100491997.298846709.1580837996-1159488863.1580234077.

[31] See https://osf.io/registries/discover?provider=OSF&type=Registered%20Report%20Protocol%20Preregistration.

SOFTWARE AND CODE

Relevance to Open Ecosystem

Research projects may generate code that is used as a means to run, analyze, or interpret research data. The ability to independently confirm results and conclusions is critical for evaluating scientific rigor and informing future research activities. To extract maximum value from research findings and available data, any code deployed to process these data must therefore be widely and freely available. Research findings are not fully open unless the tools necessary to understand and test them are also made available. Research projects may also generate software that is the product of the project rather than the byproduct, a specified deliverable designed to perform a specific task. Making the underlying code for this type of research output open source can encourage collaboration, further development, community engagement, and enhanced return on funders' investment.

Considerations

As organizations develop open science policies pertaining to code and software, among the issues they must consider are the following:

- ***Software/Code Maintenance.*** What are the expectations for the duration and extent to which code should be kept up to date? Should the version used to produce the reported findings be maintained?
- ***Proprietary Software.*** To the extent that some or all of the code base upon which an experiment relies is not open source, what steps can be taken to reduce restrictions on its reuse?
- ***Timing.*** Does the policy require that the code or software be made openly available immediately upon the posting of research findings (e.g., publication of an article, deposit of a dataset), or is some embargo (e.g., 6 months) permissible? If research findings are not published or posted, should code and software be made publicly available no later than grant close?
- ***Financial Support.*** Will the policy maker provide funding to defray costs of preparing and/or depositing the code or software? If so, is there a cap on the amount? Must the researcher explicitly account for these expenses at the time of project design? If code

or software is made publicly available after the conclusion of the grant, does the grantee have a mechanism to request additional financial support?

- ***Licensing.*** What type of licensing requirements will the policy include to facilitate reuse? Do the grantee and/or the funder retain any stake in the intellectual property?
- ***Metadata.*** What documentation and descriptive details are needed to understand and execute the code or run the software program? How will the computational environment in which software or code was originally executed be described and archived? Should researchers establish virtual environments (e.g., Docker)?
- ***Preservation.*** What constitutes an appropriate deposit location for the code or software? Is there a repository that is appropriate for the subject matter in question and/or has emerged within a specific research community as the default resource in that field? Is the repository secure, stable, and open for all to access? Does the repository assign persistent digital identifiers to code?

Approaches

The TOP Guidelines advise that researchers should "provide program code, scripts for statistical packages, and other documentation sufficient to allow an informed researcher to precisely reproduce all published results ... through a trusted digital repository."[32] More funder-specific TOP guidance may be found at https://www.cos.io/initiatives/top-funders.

Some agencies within the U.S government use open source code as a matter of policy. For example, the Consumer Financial Protection Bureau unequivocally states, "When we build our own software or contract with a third party to build it for us, we will share the code with the public at no charge."[33] Other agencies, such as the Department of Education, make the source code for their prominent public-facing initiatives (in ED's case, the College Scorecard)[34] openly available. Both of these organizations deposit these research outputs (software as a product, not a byproduct, of the grant) on GitHub. When code is developed to interpret or analyze research findings (code as a secondary output of the grant), organizations such as

[32] See https://osf.io/bcj53.

[33] See https://www.consumerfinance.gov/about-us/blog/the-cfpbs-source-code-policy-open-and-shared.

[34] See https://collegescorecard.ed.gov.

BOX 6
Examples of Open-Code and Software Policies

- NASA's Earth Science Data Systems (ESDS) Program requires that all software developed through research and technology awards be made available to the public as open source.[a] All funding proposals must include software development plans that are vetted as part of the application process.
- The U.S. government's Federal Source Code Policy includes a pilot program that "requires agencies, when commissioning new custom software, to release at least 20 percent of new custom-developed code as Open Source Software for three years."[b]
- Several learned societies that publish flagship disciplinary journals, including the American Geophysical Union and the American Astronomical Society, require or strongly encourage authors to make openly available any code used to generate results or analyses reported in their papers.[c]

[a] See https://earthdata.nasa.gov/collaborate/open-data-services-and-software/esds-open-source-policy.
[b] See https://www.cio.gov/2016/08/11/peoples-code.html.
[c] See https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Policies/Data-policy and https://journals.aas.org/news/policy-statement-on-software.

the Wellcome Trust typically require the code to be shared at the time the primary research is published.[35] (See Box 6 for examples of open-code and software policies.)

Resourcing

For code specifically, some technical expertise may be required to ensure that the code and software are operable and can be accessed and used by the wider community.

Once open policies are implemented, organizations can undertake a range of activities to manage them. At the low-touch end of the spectrum, organizations can require researchers to document how they intend to

[35] See https://wellcome.org/news/our-new-policy-sharing-research-data-what-it-means-you.

comply. Depending on internal resources, some organizations spot-check these plans, while others simply rely on the honor system. Other organizations take a more engaged approach, requiring proof of compliance from researchers and checking this against internal expectations and guidelines.

Next Steps

The TOP Guidelines provide sample language for three levels of open-code policies.[36] This wording can be adapted and adopted to suit the specific circumstances of a range of organizations. For a deeper dive into policy formulation, interested parties can download the National Academies of Sciences, Engineering, and Medicine's report *Open Source Software Policy Options for NASA Earth and Space Sciences*.[37] This comprehensive document provides a deep dive into the established approaches, best practices, and practical considerations that can help effectively shape an open code policy.

[36] See https://osf.io/bcj53.

[37] See https://www.nap.edu/catalog/25217.