POLICY FORUM

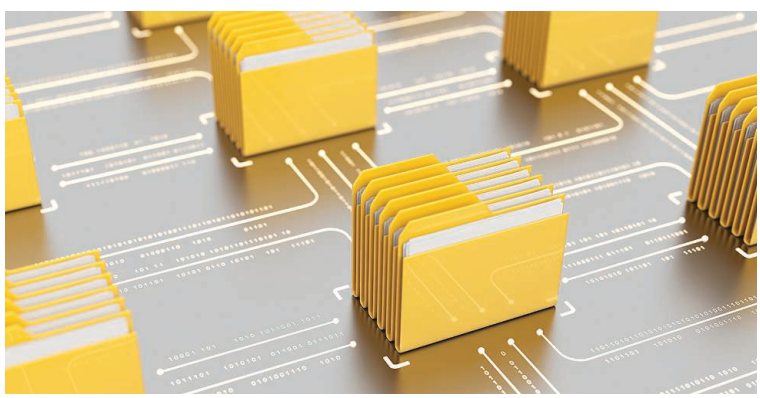
DATA

# Time for NIH to lead on data sharing

## A draft policy is generally supportive but should start mandating data sharing

By **Ida Sim[1,2], Michael Stebbins[3], Barbara E. Bierer[4,5], Atul J. Butte[1], Jeffrey Drazen[6], Victor Dzau[7], Adrian F. Hernandez[8], Harlan M. Krumholz[9], Bernard Lo[10], Bernard Munos[11], Eric Perakslis[8], Frank Rockhold[8], Joseph S. Ross[9], Sharon F. Terry[12], Keith R. Yamamoto[1], Deborah A. Zarin[4], Rebecca Li[2,13]**

The U.S. National Institutes of Health (NIH), the largest global funder of biomedical research, is in the midst of digesting public comments toward finalizing a data sharing policy. Although the draft policy is generally supportive of data sharing (*1*), it needs strengthening if we are to collectively achieve a long-standing vision of open science built on the principles of findable, accessible, interoperable, and reusable (FAIR) (*2*) data sharing. Relying on investigators to voluntarily share data has not, thus far, led to widespread open science practices (*3*); thus, we suggest steps that NIH could take to lead on scientific data sharing, with an initial focus on clinical trial data sharing.



In 2013, the White House directed all U.S. federal research funding agencies with more than $100 million in annual research and development expenditures to develop programs to ensure access to the results of publicly funded research, including peer-reviewed publications and digital data (*4*). The directive was explicit: "[D]igitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze." In 2015, the NIH issued its plans for responding to this directive, asserting its explicit intent "to make public access to digital scientific data the standard for all NIH-funded research" and to "[e]nsure that data management plans include clear plans for sharing research data" (*5*). In November 2019, the NIH assured the U.S. Government Accountability Office that "it is in the process of developing an agency-wide data management and sharing policy, including compliance mechanisms, to fully implement its public access plan" (*6*).

Under the draft policy, NIH would require researchers to submit a plan describing the "rationale for decisions about which scientific data will be preserved and shared." However, the draft policy does not specify a minimum standard or time frame for data sharing and, most importantly, stops short of a definitive mandate for sharing. In the absence of an explicitly stated requirement, we are concerned that researchers will be able to comply with this policy even if their plan was to effectively withhold data from public access (for example, some current plans amount to little more than "email me"). As written, the draft policy may not practically result in data being shared by default for NIH-sponsored research. Moreover, the draft policy requires submission of data sharing plans only after the proposal review process has concluded, which researchers could interpret as meaning that data sharing plans are not a core part of good scientific practice, unlike trial recruitment or statistical analysis plans.

To be sure, there are challenges to implementing FAIR data sharing. For some types of data, sharing may be legitimately delayed or restricted to protect confidential commercial information or for reasons of national or personal security. Privacy considerations are paramount when sharing individual participant-level data from human studies, which legitimize additional protections.

Although it would advance the entire research enterprise, mandatory data sharing would have perhaps its broadest and most immediate impact on clinical trials, where sharing of participant-level data will not only accelerate discovery but would also meet the ethical imperative to honor trial participants' assumption of personal risk by maximizing the potential scientific value of the data. Substantial advances have been made in recent years in the technology, infrastructure, and governance of participant-level clinical trial data sharing. Several repositories have established successful models of sharing and have demonstrated assurance of patient privacy and security and feasibility of use and are experiencing accelerating user uptake (*7*–*9*). Concern over where and how to share clinical trial data is no longer a viable rationale for delay, even as we acknowledge that more needs to be done to ease researcher use of these repositories.

One argument for delaying mandated sharing is the desire to introduce the requirement only when all support infrastructure is in place. Although additional standards, policy, and support infrastructure are needed, NIH should not let the perfect be the enemy of good progress. A phased program beginning with mandatory sharing of clinical trial data, with expansion to other types of data as standards and best practices around data stewardship emerge, seems wise. NIH al-

[1]University of California San Francisco, San Francisco, CA, USA. [2]Vivli, Cambridge, MA, USA. [3]Science Advisors, Washington, DC, USA. [4]Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University, Cambridge, MA, USA. [5]Department of Medicine, Harvard Medical School and Brigham and Women's Hospital, Boston MA, USA. [6]Pulmonary and Communications Divisions, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [7]National Academy of Medicine, Washington, DC, USA. [8]Duke University, Durham, NC, USA. [9]Yale University, New Haven, CT, USA. [10]Greenwall Foundation, New York, NY, USA. [11]FasterCures (Milken Institute), Washington, DC, USA. [12]Genetic Alliance, Washington, DC, USA. [13]Center for Bioethics, Harvard Medical School, Boston, MA, USA. Email: ida.sim@ucsf.edu

ready has the requisite technical (such as ClinicalTrials.gov) and policy infrastructure (*10*) that can be extended to support a mandate for participant-level clinical trial data sharing. Mandates may also drive additional public and private investments to bolster a wider data sharing ecosystem. For clinical trials, for example, substantial private investment in data repositories has helped to expand data sharing capacity for both industry and academic trials (*11*).

Another concern about a sharing requirement is the imposition of rigid standards. Certainly, investigators should have some flexibility—for example, when to share data and which platform to use—but should be held accountable for meeting minimum standards of data stewardship and availability. Flexibility in how to share, not whether to share, should be the policy framework. For clinical trials in particular, refusing to share breaks trust with trial participants' strong desire to share (*12*). Past experience has shown that science can flourish because of, not in spite of, mandated data sharing: NIH's 2014 Genomic Data Sharing Policy implemented explicit data sharing requirements with sanctions for noncompliance. After some early resistance from the genetics investigator community, genomic data sharing is now widely accepted by researchers and research institutions and was a major enabler of the extraordinary scientific and economic value gained from the Human Genome Project.

Specific, practical, and implementable NIH policies can help transform academic culture and practice toward routine data sharing. Building on the current draft plan, we recommend that NIH establish mandated sharing of participant-level data from interventional clinical research, for which the ethical arguments for sharing are most compelling. Below, we recommend new enforceable policies for implementing such a mandate. These recommendations should apply to all prospective human subjects research and could subsequently be adapted for other biomedical research data sharing.

NIH should establish standards for clinical research data sharing repositories, maintain a list of approved repositories, and promote awareness and use of these repositories. NIH should take the lead in facilitating interoperability among approved repositories and fostering close coordination with ClinicalTrials.gov to minimize burdens to investigators and ensure that the data are "findable" and can be understood in the context of the full range of studies on a particular topic.

NIH should require all clinical research proposals to include a data sharing plan that commits to sharing participant-level data in an approved repository. This data sharing plan should be explicitly scored in the grant review process; scores should affect the overall funding decision. Once funded, researchers should be required to post the data sharing plan, selected repository, and anticipated date of data availability to ClinicalTrials.gov before enrollment of the first participant to provide public accountability. Subsequent-year funding (for the duration of the study) should be contingent on meeting these new ClinicalTrials.gov reporting requirements. Applications should include the methods and appropriate budget in the main grant proposal to ensure appropriate data stewardship so that data will be findable and sharable in approved repositories at the conclusion of the study.

In addition to the current requirement that human studies report summary results to ClinicalTrials.gov within the time frames established under the law (generally 1 year, with some exceptions) (*13*), NIH should also require reporting of participant-level data to an approved repository within a reasonable time period. Although there is disagreement about when participant-level data should be reported, the U.S. National Academy of Medicine has deemed 18 months after trial completion to be a reasonable embargo period (*14*).

NIH should establish mechanisms for applicants to report (such as in biosketches) whether and how they executed on data sharing plans from previous grants. Consequences for failing to report and share results should include loss of eligibility for future funding. Conversely, exemplary past data sharing practices should be recognized and rewarded within the grant review process. Because technical and policy issues remain, NIH should continue to support efforts to address challenges to effective FAIR data sharing, such as data management, alignment for reuse, and sharing infrastructures.

Of course, academic institutions must be partners in this effort. Academia must train and support investigators to meet data sharing objectives; partner with approved data sharing platforms; recognize data contributorship in hiring, promotions, and tenure decisions (*15*); and train and reward investigators for reusing data as a valuable complement to generating data through primary studies. As a major source of funding for academic medical centers, the Clinical and Translational Science Award program and other major NIH networks should include institutional data sharing practices in its evaluation and funding criteria. Academia, NIH, scientific societies, and other stakeholders should work together to achieve clear and consistent rewards for and enforcement of all data sharing requirements, including sanctions for noncompliance. Fair and robust oversight is essential to ensure that the fruits of federally funded research are put to maximal use.

We suggest that limited data sharing arises not from culture but from policy. Researcher reluctance to share is a rational response to existing incentive systems that measure and reward individual achievement partly on the basis of the accumulation and use of closely held data sets. NIH has outsized influence to adjust these incentives by mandating and making funding contingent on data sharing, realigning researcher behavior with core values of scholarship. Once data sharing becomes the norm, researchers and the general public will benefit, and in turn, sharing will itself become an incentive. ■

## REFERENCES AND NOTES


1. NIH, "Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance" (notice no. NOT-OD-20-013, NIH, 2019); https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-013.html.
2. M. D. Wilkinson *et al.*, *Sci. Data* **3**, 160018 (2016).
3. A. Bergeris, T. Tse, D. A. Zarin, *JAMA* **319**, 406 (2018).
4. J. P. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research" (Office of Science and Technology Policy, 2013); https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
5. NIH, "Plan for Increasing Access to Scientific Publications" (NIH, 2015); https://grants.nih.gov/grants/nih-public-access-plan.pdf.
6. U.S. Government Accountability Office (GAO), "Federal Research: Additional Actions Needed to Improve Public Access to Research Results" (report no. GAO-20-81, GAO, 2019); https://www.gao.gov/assets/710/702847.pdf.
7. F. Rockhold, P. Nisen, A. Freeman, *N. Engl. J. Med.* **375**, 1115 (2016).
8. R. Li *et al.*, *NAM Perspect.* 10.31478/201811b (2018).
9. J. S. Ross *et al.*, *Sci. Data* **5**, 180268 (2018).
10. NIH, "NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information" (NIH, 2017); https://grants.nih.gov/policy/clinical-trials/reporting/understanding/nih-policy.htm.
11. R. Banzi *et al.*, *Trials* **20**, 169 (2019).
12. M. M. Mello, V. Lieou, S. N. Goodman, *N. Engl. J. Med.* **378**, 2202 (2018).
13. D. A. Zarin, T. Tse, R. J. Williams, S. Carr, *N. Engl. J. Med.* **375**, 1998 (2016).
14. Institute of Medicine, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* (National Academies Press, 2015).
15. H. H. Pierce, A. Dev, E. Statham, B. E. Bierer, *Nature* **570**, 30 (2019).


## ACKNOWLEDGMENTS


I.S, R.L., and B.E.B. are cofounders of Vivli, a nonprofit platform for clinical trial data sharing, and currently serve on the Vivli Board of Directors. R.L. is an employee of Vivli, and I.S. is a paid consultant of Vivli. M.S. serves on the Vivli Board of Directors, and B.L. and J.D. serve on Vivli's External Advisory Committee. F.R. serves as a Vivli Senior Advisor and Principal Investigator of Supporting Open Access for Researchers, Duke Clinical Research Institute's data sharing platform partially funded by Bristol-Myers Squibb. Vivli receives funding support from the Arnold Ventures, Doris Duke Charitable Trust, the Helmsley Charitable Trust, the pharmaceutical industry, and academic institutional partners. J.S.R. and H.M.K. are co-founders of the Yale Open Data Access Project, a Yale University platform for clinical trial sharing, which shares trials for and receives support from Johnson and Johnson. A.J.B. receives research support from the National Institute of Allergy and Infectious Diseases as principal investigator of ImmPort, an open-access data repository, including clinical trials data. Additional author disclosures are provided as supplementary materials.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6484/1308/suppl/DC1


10.1126/science.aba4456

# Time for NIH to lead on data sharing

Ida Sim, Michael Stebbins, Barbara E. Bierer, Atul J. Butte, Jeffrey Drazen, Victor Dzau, Adrian F. Hernandez, Harlan M. Krumholz, Bernard Lo, Bernard Munos, Eric Perakslis, Frank Rockhold, Joseph S. Ross, Sharon F. Terry, Keith R. Yamamoto, Deborah A. Zarin and Rebecca Li

*Science* **367** (6484), 1308-1309.
DOI: 10.1126/science.aba4456