



**PUBLIC RESPONSIBILITY IN  
MEDICINE AND RESEARCH**

**Chair**

*Heather H. Pierce, JD, MPH*

**Vice Chair**

*Christine Grady, MSN, PhD*

**Secretary**

*David H. Strauss, MD*

**Treasurer**

*Christian E. Newcomer,  
VMD, MS, DACLAM*

**Board of Directors**

*Albert J. Allen, MD, PhD*

*A. Cornelius Baker*

*Barbara E. Bierer, MD*

*Elizabeth A. Buchanan, PhD*

*Alexander Capron, LLB*

*Owen Garrick, MD, MBA*

*Bruce Gordon, MD*

*Mary L. Gray, PhD*

*F. Claire Hankenson,  
DVM, MS, DACLAM*

*Karen M. Hansen*

*Martha Jones, MA, CIP*

*Natalie L. Mays, BA, LATG, CPIA*

*Robert Nobles, DrPH, MPH, CIP*

*Sally Okun, RN, MMHS*

*Suzanne Rivera, PhD, MSW*

*Stephen Rosenfeld, MD*

*Walter L. Strauss, MD, MPH*

**Ex Officio**

*Elisa A. Hurley, PhD  
Executive Director*

December 3, 2018

Submitted electronically at <https://osp.od.nih.gov/provisions-data-management-sharing/>

Francis S. Collins, MD, PhD  
National Institutes of Health  
6705 Rockledge Drive, Suite 750  
Bethesda, MD 20892-7985

*RE: Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research*

Dear Dr. Collins:

Public Responsibility in Medicine and Research (PRIM&R) appreciates the opportunity to comment on the National Institutes of Health (NIH)'s Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research, published October 10, 2018.

PRIM&R is a nonprofit organization dedicated to advancing the highest ethical standards in the conduct of research. Since 1974, PRIM&R has served as a professional home and trusted thought leader for the research protections community, including members and staff of human research protection programs and institutional review boards (IRBs), investigators, and their institutions. Through educational programming, professional development opportunities, and public policy initiatives, PRIM&R seeks to ensure that all stakeholders in the research enterprise understand the central importance of ethics to the advancement of science.

PRIM&R fully supports initiatives that seek to promote broad data sharing. We agree with the NIH that data sharing can optimize the use of scarce research resources and has the potential to accelerate science and its application to human health. Furthermore, the sharing

of scientific data from clinical trials and other research involving humans honors those subjects' contributions by maximizing the value of their involvement. Relevant to the NIH's point that data sharing can inform "future research pathways," data sharing may also lead to better designed, and safer, future research, serving the ethical imperative to minimize risks to research subjects.

While data sharing has significant benefits, it also involves inherent risks, most notably privacy and confidentiality risks particularly when it comes to certain types of genetic data. These risks are magnified by the fact that research often involves accessing and aggregating multiple primary data sets. Though each of these data sets may include only "de-identified" personal data, their aggregation increases the chances that individuals will be inadvertently identified and their privacy breached. Indeed, advances in technology and the proliferation of data sources mean that no data can be considered permanently de-identified.

These risks, most agree, are, unavoidable, but there are strategies to mitigate them. **We believe the draft provisions fall short on acknowledging and grappling with these realities of data sharing. As the NIH considers updating its data sharing policies, it has a unique and important opportunity to lead the way on responsible data sharing by articulating the tradeoffs between maximizing the value of scientific data and protecting the rights and interests of research subjects, and by providing guidance on best practices for responsible data sharing given those tradeoffs. Both of these measures will, in turn, enhance public trust in the data sharing enterprise.**

In what follows we elaborate on these points, highlighting more specific areas we urge the NIH to address in future iterations of data sharing and management policies. In response to the specific areas of focus in the RFI, we begin by addressing, in sections I and II, the definition of scientific data and the proposed requirements for data management and sharing plans. Section III raises several additional points for the NIH to consider.

## **I. The Definition of Scientific Data**

**The definition of "scientific data" provided requires clarification.** According to the proposal, scientific data is "the recorded factual material commonly accepted in the scientific community as necessary to validate and replicate research findings including, but not limited to, data used to support scholarly publications," and "may include certain individual-level and summary or aggregate data, as well as metadata." The definition offered is very broad but, at the same time, excludes specific sources of information, which could lead to confusion about how best to interpret and implement it. For example, the

definition of scientific data explicitly excludes laboratory notebooks. However, the information contained in laboratory notebooks might reasonably be seen as “necessary to validate and replicate research findings.” Indeed, given its breadth, the proposed definition of scientific data seems open to the interpretation that almost *all* data collected for a study counts, and therefore should be shared. We are concerned that a definition of scientific data that is open in this way to multiple local interpretations increases the likelihood that more identifiable information will be shared than is intended by the policy.

Furthermore, we note that the proposed definition of “scientific data” is presumably meant to include data from both quantitative and qualitative research. If this is the case, then the draft provisions fail to acknowledge that there are key differences between qualitative and quantitative research methods and data. For instance, qualitative research data often contains more identifiable information than quantitative research data. PRIM&R is unaware of any established standards for making qualitative data widely available in a way that protects the rights and interests of research subjects. Indeed, qualitative researchers would argue that there are strong ethical reasons *not* to share primary datasets. The one-size-fits-all model proposed may not be appropriate for all research data. The NIH should address this concern, at the very least clarifying whether and when its policies apply to both quantitative and qualitative data, and if so, acknowledging the unique challenges associated with the latter.

Finally, we urge the NIH to consider and clarify whether and how its definition of scientific data applies to data that is generated after a grant ends. We can imagine circumstances in which a researcher re-analyzes a project’s data after the end of the grant that initially funded the project, and finds something that fits the proposed definition of scientific data. Will that data be covered by the NIH’s policy? What is NIH’s “reach through” in such circumstances? The agency should address how it plans to oversee any sharing requirements when new data is generated after the funding period has ended.

## **II. The Requirements for Data Management and Sharing Plans**

We believe the NIH’s proposed requirements for data management and sharing plans cover many important elements of such plans. The agency notes several times that data management and sharing plans should provide for the broadest use of data, “consistent with privacy, security, informed consent, and proprietary issues.” However, the agency provides very little guidance about what those issues are or how they should be addressed in data sharing plans. We urge the NIH to more fully acknowledge and address the risks and complexities associated with data sharing. Below we provide several examples.

First, the proposed provisions suggest that use of “persistent unique identifiers” for scientific data would be acceptable as an indexing tool for making shared data discoverable. However, persistent unique identifiers may actually facilitate reidentification. For example, some government projects require the use of the Global Unique Identifier (GUID Tool). In these circumstances, a given subject retains the same GUID, which enables the triangulation of data from unrelated studies and poses privacy and confidentiality risks. We suggest the NIH reconsider whether it should endorse use of persistent unique identifiers or instead suggest other indexing tools that might be more appropriate and less susceptible to re-identification efforts.

Second, we are also concerned about the requirement that broad sharing be consistent with informed consent. The revised Common Rule requires consent documents to include a statement about what will happen to any identifiable private information collected during the course of research—specifically, whether or not the information might be stripped of identifiers and distributed or used for future research without consent. This means that for data collected under the new rule, it will be relatively clear what it means to share it “consistent with consent.” However, information collected prior to the January 21, 2019 compliance date for the revised rule is not subject to those new consent requirements, and existing consent forms vary with respect to how, or even whether, they address data sharing. In these circumstances, it is not clear what it means to say that sharing should be done as broadly as possible, consistent with consent.

**We again encourage the NIH to lead by providing guidance for IRBs and other stakeholders on how to determine what retrospective uses of existing data, including data sharing, would be ethically appropriate when consent is not specific about, or is silent on, future uses.** We urge the NIH to be explicit, in its own policies, about the series of considerations that come into play when making these decisions, such as the characteristics of the study population, the sensitivity of the data, the likelihood of reidentification, and the scientific utility and value of the data itself. The NIH should also remind stakeholders that these issues may need to be reviewed on a case by case basis to reach a decision about how best to share data while protecting research subjects’ rights.

More generally, the **NIH should provide more guidance not just on how appropriate informed consent can facilitate responsible data sharing, but also on best practices for sharing data in ways that are consistent with privacy and confidentiality standards.** For example, the draft policy currently does not mention the HIPAA Privacy Rule, with which much data sharing must, of course, be consistent for certain health related research. The research community would also benefit from guidance on when it is reasonable to place restrictions on data use and sharing.

Finally, **the NIH should encourage, if not require, data management and sharing plans to include provisions about how research subjects will be informed about the limitations of current technologies to completely de-identify or anonymize their data while preserving that data's utility for research.** This sort of transparency about data sharing and the tradeoffs involved demonstrates respect for research subjects and may enhance the public's trust in the data sharing enterprise. In the same spirit of transparency and fostering trust, we further urge the NIH to encourage the creators of data management and sharing plans to incorporate input from research subjects and/or the public on those plans' assessment of risks and benefits. This may not necessarily require soliciting input on each project's plan, but rather institutions seeking input on the risks and benefits associated with particular categories of data or types of research.

Given the inherent complexity of all of these issues, we suggest eliminating the proposed two-page cap for data management and sharing plans.

### **III. Other considerations**

**We encourage the NIH to consider how it can ensure that institutions who have promised to share data have the resources to do it well.** The proposed provisions focus almost exclusively on the requirement to create a data management and sharing plan, and on what should be included in the plan. But plans are only as effective as their implementation. We are concerned that institutions with fewer resources to dedicate to data sharing, and/or less experience with data sharing, may write good plans, but may be unable to execute them successfully, leaving people and their data vulnerable. This concern is particularly acute given the agency's expectation that, where possible, scientific data be digitized, a resource intensive process.

These concerns may be partially addressed by the NIH encouraging grantees to request the appropriate amount of resources to facilitate data sharing in a safe and ethical manner—though less experienced institutions may need help from the NIH understanding what the costs are. Furthermore, the NIH should consider other ways it can support institutions with fewer resources for safe data sharing, so that this policy does not for them constitute an unfunded mandate.

Relatedly, we suggest that any future policy expand on the compliance and enforcement provisions proposed. Although other sections of the proposal emphasize that data sharing must be consistent with privacy and confidentiality considerations and informed consent, there is no discussion of what penalties might be levied if research subjects' rights are

violated in the course of data sharing—for instance, in the event that private information about them gets in the wrong hands—and how to determine who should be held responsible for such violations. As data sharing becomes more prevalent, the public will increasingly demand consequences when their data are not shared with adequate attention to protections. **The NIH can demonstrate its commitment to the public’s interest by detailing the consequences when data is shared inappropriately, beyond just rescinding funding.**

In addition, although the proposed policy mentions the utility of data repositories, it doesn’t address the current proliferation of repositories, each with their own rules and procedures. Not only does this state of affairs lead to confusion, it weakens the overall utility of the data sharing enterprise. Effective use of existing data to advance science requires an accessible set of data repositories structured in a rational and coherent way. The agency endorses use of repositories that meet “community-based standards,” but it is unclear, without further explication, what the NIH has in mind—for instance, whether the agency means the FAIR data principles. **We urge the NIH to use its policies to encourage standardization across data repositories, and to articulate a gold standard for how data should be managed and shared to maximize utility.**

\*\*\*\*

Ultimately, **it will be important for the NIH to justify the relative risks and benefits of its data sharing policies**, and we hope we have provided some useful input accordingly. But it is worth pointing out that we, collectively, still have a lot to learn about data sharing and its risks and benefits. We are at the early stages of broad data sharing efforts, and technologies for both sharing and protecting data are evolving rapidly. Until we fully understand the risks and benefits of data sharing, we urge the NIH, in its leadership role, to continue to monitor both the utilization of data sharing strategies and the barriers to their use, to learn from the successes and failures of methods used to protect people’s privacy and enhance their welfare, and to incorporate what is learned into its communications with the research community, and into its own policies.

Thank you again for the opportunity to comment on this important issue. My PRIM&R colleagues and I are available to discuss our comments further, should that be of interest. We look forward to the next stage of policymaking in this area. Please feel free to contact me at 617.303.1872 or [ehurley@primr.org](mailto:ehurley@primr.org).

Respectfully submitted,

A handwritten signature in black ink, reading "Elisa A. Hurley". The signature is written in a cursive style with a long, sweeping tail on the letter "y".

Elisa A. Hurley, PhD  
Executive Director

cc: PRIM&R Public Policy Committee, PRIM&R Board of Directors