March 15, 2024

Office of Data Science Strategy
National Institutes of Health

RE: NOT-OD-24-037 – Request for Information (RFI): Inviting Comments on the National Institutes of Health (NIH) Strategic Plan for Data Science

Comments submitted electronically via https://datascience.nih.gov/rfi-strategic-plan

The Federation of American Societies for Experimental Biology (FASEB), a coalition of 23 scientific societies representing over 110,000 researchers in the biological and biomedical sciences, appreciates the opportunity to comment on the draft *NIH Strategic Plan for Data Science 2023 – 2028*. For over a decade, FASEB has actively engaged in initiatives related to data management, sharing, and reuse within the biological and biomedical sciences, both through science policy efforts and the strategic investment in FASEB DataWorks!, an initiative to support research teams in the sharing of best practices to foster data sharing and reuse.

In 2018, FASEB submitted extensive comments in response to the RFI seeking input on NIH's inaugural Strategic Plan for Data Science, and we are pleased with the agency's progress towards the goals and objectives of that initial plan as noted in Appendix I of the updated plan including finalization and implementation of a new data management and sharing policy for research grants submitted after January 25, 2023. Although the landscape within which this updated plan is presented is significantly different, many of the core issues highlighted in FASEB's 2018 comments – namely the need for minimal data standards, consideration of the long-tail of biological research data, the need for NIH-wide coordination, and engagement of extramural stakeholders to foster culture change regarding data sharing and reuse in the biological and biomedical sciences – still apply and are reflected in the updated plan.

For the purposes of the current RFI, FASEB's comments focus on Goal 1, Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing, and Goal 5, Strengthen a Broad Community in Data Science, and the objectives contained within each. In addition to the specific comments on the draft goals, we offer an overarching suggestion to improve the community's understanding of the NIH data management and sharing ecosystem.

Goal 1: Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing
Appropriateness of the goal and proposed implementation tactics
NIH's adoption of an updated agency-wide policy for data management and sharing to fulfill the 2013 Office of Science and Technology Policy (OSTP) memorandum to increase access to the results of federally funded scientific research is a key update to the policy landscape within which the proposed strategic plan and goals will operate. However, the first year of this policy's implementation highlighted a critical need for resources – both within NIH and in the extramural research community – to ensure uniform understanding and enforcement of the policy. FASEB was pleased to see that the objectives

Full members: American Physiological Society • American Society for Biochemistry and Molecular Biology • American Society for Pharmacology and Experimental Therapeutics • American Society for Investigative Pathology • The American Association of Immunologists • American Association for Anatomy • Society for Developmental Biology • Association of Biomolecular Resource Facilities • The American Society for Bone and Mineral Research • Society for the Study of Reproduction • Endocrine Society • Genetics Society of America • The Histochemical Society • Society for Glycobiology • Association for Molecular Pathology • Society for Redox Biology and Medicine • Society For Experimental Biology and Medicine • American Aging Association • Society for Leukocyte Biology • American Federation for Medical Research • Shock Society • **Associate members:** American Society of Human Genetics • American Society for Nutrition

within Goal 1 seek to close these gaps. The three objectives within this goal highlight the need for NIH to support the biomedical community in managing, sharing, and sustaining data. "Support" can take several forms, including financial, infrastructure, and professional acknowledgement, all of which are required to foster the culture change needed for effective data management and sharing among NIH-funded researchers.

There were a lot of lessons learned during the first year of implementation of the NIH Policy for Data Management and Sharing, many of which resulted in key clarifications for applicants (e.g., NOT-OD-23-161, NIH Application Instruction Updates – Data Management and Sharing (DMS) Costs) but also many questions that risk relegating the policy to one of compliance over long-term utility of federally funded science. While the community appreciates the flexibility of the Data Management and Sharing Policy to meet the range of science supported by NIH, there is still a lot of confusion and variability regarding what information constitutes data. Similarly, the lack of metadata standards risks a lot of effort being put into datasets that are poorly curated and thus have limited secondary utility and thus reduced return on research investment. To facilitate this critical step, FASEB strongly recommends that this goal area clearly articulates the need for NIH to issue Notices of Funding Opportunities (NOFOs) to support workshops or related convenings to establish metadata standards that are broadly applicable to NIH-funded research as well as those for more specific research domains.

In addition to the critical need for the research community to coalesce around shared metadata standards, NIH similarly must coordinate agency-wide resources to ensure uniform implementation and enforcement of the Data Management and Sharing Policy. A common concern raised across FASEB's constituents is a lack of consistency in the guidance received from I/C to I/C regarding plan implementation. This highlights a need for both internal agency resources to ensure comparable understanding of the intent of the policy as well as resources to help the community fulfill policy requirements. Resources such as sharing.nih.gov and templates developed in partnership with the Federal Demonstration Project are helpful but reinforce a checklist approach rather than true culture change.

Opportunities for NIH Partnership
Objective 1-1, Support the Biomedical Community to Manage, Share, and Sustain Data, is very aligned with what FASEB is trying to achieve with its DataWorks! initiative. Launched in September 2021, DataWorks! aims to support researchers by providing a multi-faceted approach to develop and hone core competencies in data management, sharing, and reuse among biological and biomedical researchers. Current offerings include (1) DataWorks! Salons – conversation spaces through which researchers can learn about existing resources and exchange best practices for data management at the individual, team, and network level; (2) DataWorks! Prize – a mechanism to reward research teams for adopting novel strategies to integrate data sharing and reuse to advance scientific knowledge; and (3) DataWorks! Help Desk – a knowledgebase resource to provide researcher-to-researcher guidance for navigating and adopting data sharing and reuse policies and practices. A fourth component – DataWorks! Community – has not launched but reflects the need to identify and develop core data management competencies within research teams.

FASEB appreciates the strong support from the NIH Office of Data Science Strategy (ODSS) to offer financial rewards as well as a forum within which award-wining strategies can be presented to the broader research community. This mechanism and other incentive-based approaches requires the investment of resources – both financial and intellectual capital – to ensure they meet the desired audience(s) and goals.

Without an understanding of existing practices pertaining to data sharing and reuse, it will be difficult to assess the impact of the NIH Data Management and Sharing Policy. To understand existing practices within the FASEB community, the DataWorks! team collaborated with SciCrunch to measure mentions of research resources, databases, and repositories (RDRs) and utilize them as proxies of data sharing and reuse activity.[1] An effective metric, such as RDR mentions, will require incentivization and coordination across funding agencies, publishers, and scientific societies to foster the necessary culture change.

Goal 5: Strengthen a Broad Community in Data Science

As a long-standing proponent of training and professional development opportunities to ensure a biological and biomedical workforce equipped to utilize a range of resources to tackle critical scientific challenges, FASEB was pleased by the inclusion of a goal devoted to fostering data science skills across the research community. The essence of Goal 5 is to expand both the community engaged in data science and the resources available to foster their career development and progression. However, while FASEB supports the core objectives and proposed implementation tactics for this goal, we are concerned that many will be seen as unfunded mandates to existing programs or lacking financial backing necessary to promote culture change in the extramural research community. Therefore, as noted in our comments for Goal 1, we urge NIH to issue NOFOs (parent or supplemental) to ensure appropriate resources are available for these professional development opportunities.

Improving Community Awareness of NIH Data Science Resources and Investments

To facilitate community understanding and awareness of the NIH-wide efforts related to data science, FASEB recommends inclusion of an illustration or schema of the current NIH data science ecosystem. For many investigators, compliance with the NIH Data Management and Sharing Policy may be their first and/or only interaction with the agency's data science efforts. Better understanding of agency-wide goals in data science and how compliance measures such as the Data Management and Sharing Policy fit into this larger ecosystem will help foster the culture change necessary to pursue and fulfill larger government-wide goals, such as those in the 2022 OSTP memorandum, *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*. Similarly, we encourage the NIH ODSS to provide regular updates on progress towards this strategic plan to the broader research community, once finalized and adopted.

FASEB applauds NIH, particularly ODSS, for the progress achieved towards the goals of the agency's first strategic plan for data science strategy. Overall, the proposed strategic plan for the next five years (2023 – 2028) represents a reasonable transition from establishing data science infrastructure to equipping the research community with the tools and resources to use it in the most effective manner and enable new discoveries through data reuse. There are many community partners – FASEB included – ready and willing to assist with this vision; however, all will require designated financial support to provide the infrastructure and human resources to foster this culture change.

Sincerely,

Mary-Ann Bjornsti, PhD
FASEB President

---

[1] Piekniewska, A., Haak, L.L., Henderson, D., McNeill, K., Bandrowski, A., and Seger, Y. (2024). Establishing an early indicator for data sharing and reuse. *Learned Publishing*: 37(1): 22-29.