Association of
American Medical Colleges
655 K Street, N.W., Suite 100, Washington, D.C. 20001-2399
T 202 828 0400 F 202 828 1125
www.aamc.org

March 15, 2024

National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Re: Request for Information (RFI): Inviting Comments on the National Institutes of Health (NIH) Strategic Plan for Data Science 2023-2028 (NOT-OD-24-037)

Submitted online at https://datascience.nih.gov/rfi-strategic-plan.

The Association of American Medical Colleges (AAMC) appreciates the opportunity to provide feedback to the National Institutes of Health (NIH) on its Strategic Plan for Data Science.

The AAMC is a nonprofit association dedicated to improving the health of people everywhere through medical education, health care, medical research, and community collaborations. Its members are all 158 U.S. medical schools accredited by the Liaison Committee on Medical Education; 12 accredited Canadian medical schools; approximately 400 academic health systems and teaching hospitals, including Department of Veterans Affairs medical centers; and more than 70 academic societies. Through these institutions and organizations, the AAMC leads and serves America's medical schools, academic health systems and teaching hospitals, and the millions of individuals across academic medicine, including more than 193,000 full-time faculty members, 96,000 medical students, 153,000 resident physicians, and 60,000 graduate students and postdoctoral researchers in the biomedical sciences. Following a 2022 merger, the Alliance of Academic Health Centers and the Alliance of Academic Health Centers International broadened participation in the AAMC by U.S. and international academic health centers.

The AAMC recognizes the increasing importance of data science within the research enterprise, with data-driven biomedical research that involves the storage and generation of ever greater volumes of data, and more sophisticated algorithms and software for its analysis. AAMC's member institutions work with data across the spectrum of biomedical research, from basic discovery to clinical trials and electronic health record data. We recognize that this process takes significant investment and infrastructure on the part of both institutions and the NIH, and that shared data is more likely to be reused if it is properly curated, standardized, and aligned with the FAIR Guiding Principles of

Findability, Accessibility, Interoperability, and Reusability.[1] The work of the NIH Office of Data Science Strategy (ODSS) is a critical component of supporting these efforts.

We are pleased to provide comments on the following aspects of the ODSS strategic plan as requested in the RFI:

The appropriateness of the goals of the plan, the strategies and implementation tactics proposed to achieve them, including potential benefits, drawbacks or challenges

The AAMC is supportive of the goals that have been proposed in the ODSS strategic plan. In particular, as AAMC has previously commented[2] to the NIH, we are strongly encouraged by the agency's efforts to increase access to the data resulting from federally funded research (Goal 1). Increased data sharing advances biomedical research by enabling further validation of scientific results, facilitating reuse of hard-to-generate data, catalyzing new research, and generally promoting more responsible stewardship of federal resources.

We also realize that implementation of the new data management and sharing policy, which went into effect in early 2023, will require the use of NIH-supported and -managed resources (Goal 3, Goal 4). We note that AAMC previously submitted relevant comments[3] to the White House Office of Science and Technology Policy on desirable characteristics for data repositories. We also appreciate the thought that ODSS has placed into operationalizing this new policy and in the plans to build and disseminate tools and training and share additional guidelines as needed. Community engagement will be a key component for NIH to understand and respond to the particular challenges faced by low resourced institutions in meeting data sharing obligations. We commend the NIH's goal of shifting the culture of biomedical research such that data sharing and reuse are "the rule rather than the expectation," but note this cannot happen without the appropriate support from NIH. The plans proposed by the ODSS here will be critical in helping institutions to make this culture change.

We would also like to call out the importance of the goals in the strategic plan centered on enhancing the use of human health data in research (Goal 2) and broadening the data science workforce (Goal 5). These are both topics which are integral to the research enterprise in academic medicine, where medical schools and teaching hospitals are training the next generation of researchers and physicians and conducting cutting-edge work on research in clinical settings.

---

[1] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).
[2] AAMC Comments to NIH re: NOT-OD-20-013 "Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental Draft Guidance." https://www.aamc.org/media/40536/download?attachment
[3] AAMC Comments to White House OSTP re: Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research (85 FR 3085). https://www.aamc.org/media/42891/download?attachment

Opportunities for NIH to partner to achieve these goals

The success of this strategic plan will be greatly enhanced by engagement and partnerships with the research community and other relevant organizations. As NIH has noted throughout the plan, there are many community principles (FAIR, CARE and TRUST Principles), standards, and organizations around research data. Ensuring that NIH's efforts in data science incorporate existing standards that have been widely adopted and accepted by researchers is critical. We agree with ODSS's assertion that "a key to success in data interoperability is the development and use of agreed upon data standards, standardized terminologies, and CDEs (Common Data Elements)." We also note that many organizations mentioned, such as the Research Data Alliance and biomedical societies, are continually exploring new areas of data and technology, and recommend that NIH maintain ongoing engagement with these groups.

AAMC would also like to emphasize the importance of continued collaboration across the federal government, both with other science agencies and in coordination through the White House Office of Science and Technology Policy when appropriate, to ensure harmonization in policies and guidelines and to reduce regulatory burden for institutions which receive federal funding.

Specifically for the NIH's plan to "enhance programs that provide for credit and incentives for sharing data, including working with publishers, academic institutions, and other funding organizations and agencies," we recommend to the agency AAMC's Credit for Data Sharing initiative[4] and the resulting publication[5], which outlines a system for leveraging existing initiatives and infrastructure to track the use, reuse and impact of scientific data through the consistent adoption of unique identifiers. The stakeholders involved in this project represent multiple sectors, fitting with NIH's plans in this area, and many of them continue to work on the development of metrics to measure data sharing, reuse, and impact.

Where there are additional challenges in maintaining, managing, and sharing data, AAMC would be happy to connect NIH with AAMC member institutions, which have expertise across all areas listed in the strategic plan, and have the additional advantage of working across the entirety of the research spectrum, from managing basic science data to data collection and use in clinical settings. We would also like to note the leadership of AAMC's Center for Health Justice in health equity, tools for engagement with and demonstrating trustworthiness to underserved communities,[6] and research on artificial intelligence and machine learning,[7] all of which align closely with the stated goal to

---

[4] AAMC Credit for Data Sharing Initiative. https://www.aamc.org/what-we-do/mission-areas/medical-research/data-sharing

[5] Credit Data Generators for Data Reuse. Pierce, H.H., Dev, A., Statham, E., Bierer, B.E. *Nature.* 2019 June; 570 (7759): 30-32. https://www.nature.com/articles/d41586-019-01715-4

[6] AAMC Center for Health Justice: Principles of Trustworthiness. https://www.aamchealthjustice.org/our-work/trustworthiness/trustworthiness-toolkit

[7] AAMC Center for Health Justice: Foundations of Responsible NLP Use for Maternal Health Equity. https://www.aamchealthjustice.org/our-work/maternal-health-equity/foundations-nlp

"increase and improve opportunities for community engagement and partnership with underserved communities to collectively build tools and frameworks for biomedical and clinical data science uses."

AAMC would be glad to provide any additional comments to the NIH Office of Data Science Strategy as this plan is finalized and implemented. Please feel free to contact me or my colleague Anurupa Dev, PhD, Director of Science Policy and Strategy (adev@aamc.org) with any questions about these comments.

Sincerely,

Heather H. Pierce, JD, MPH
Acting Chief Scientific Officer
Senior Director for Science Policy and Regulatory Counsel

cc: David J. Skorton, MD, AAMC President and Chief Executive Officer