



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

April 2, 2018

Cindy Caughman, MPH
Scientific Data Council
National Institutes of Health
DataScienceRFI@mail.nih.gov

Re: Request for Information: Soliciting Input for the National Institutes of Health Strategic Plan for Data Science, NOT-OD-19-134

Dear Ms. Caughman:

The Association of American Medical Colleges (AAMC) appreciates the opportunity to offer comments related to the National Institutes of Health (NIH) Strategic Plan for Data Science. Founded in 1876 and based in Washington, D.C., the AAMC is a not-for-profit association dedicated to transforming health care through innovative medical education, cutting-edge patient care, and groundbreaking medical research. Its members are all 151 accredited U.S. and 17 accredited Canadian medical schools; nearly 400 major teaching hospitals and health systems, including 51 Department of Veterans Affairs medical centers; and more than 80 academic societies. Through these institutions and organizations, the AAMC serves the leaders of America's medical schools and teaching hospitals and their more than 173,000 full-time faculty members, 89,000 medical students, 129,000 resident physicians, and more than 60,000 graduate students and postdoctoral researchers in the biomedical sciences.

The AAMC recognizes the essential role of data science in the biomedical research enterprise, and the need to create an overarching strategic plan for this growing field of inquiry. We are pleased to offer the following comments in response to the request for information.

Goal 1: Support a Highly Efficient and Effective Biomedical Research Data Infrastructure

The AAMC strongly supports NIH's leadership role in developing "practical and effective policies and principles related to the storage, use, and security of biomedical data," particularly to reduce redundancies, promote coordinated systems, and create appropriate options for data storage. We recognize that the NIH has had significant success in developing and supporting large databases in genomics and other fields by fostering standards and consistent policies that have had broad support across the research community, and by ensuring the research community has access to those resources. We encourage the NIH to engage with the community and relevant advisory groups while developing cloud storage in the NIH Data Commons.

Goal 2: Promote Modernization of the Data-Resources Ecosystem

The AAMC supports the modernization of data repositories and other tools to support the storage and sharing of individual datasets, so that shared datasets are useful and usable by the broader scientific community and don't exist in individual silos. The AAMC applauds NIH's substantial effort in the development and curation of a variety of data repositories. We encourage the NIH to continue to engage the scientific community in development and integration of existing study specific, discipline specific, or generalist data repositories.

We are also pleased that the NIH is involved in the AAMC's "Credit for Data Sharing" initiative, a collaborative project with the *New England Journal of Medicine* and the MRCT Center, which will identify a validated, systematic pathway to link datasets to publications, allow academic researchers to obtain credit for shared datasets, and hopefully incentivize and promote data sharing in accordance with FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

The AAMC believes that it is critical to address issues of data security when developing infrastructure for clinical and observational data, to protect the privacy and guard against re-identification of human research participants and patients. While we support the integration of these data types into biomedical data science, we strongly urge the development of a security framework to govern the use of these datasets, to include data use agreements and the implementation of new technologies to support this goal. We recognize that overly rigid or inflexible security requirements can inhibit research and innovation without commensurately improving patient protection. The strategic plan should express NIH's dedication to addressing privacy and security of data, and ensuring balanced and appropriate protection while facilitating research. We also agree that NIH should harmonize with or refer to global data security standards as appropriate.

Goal 3: Support the Development and Dissemination of Advanced Data Management, Analytics, and Visualization Tools

The AAMC agrees that optimizing the value of biomedical research data will require new methods of data management, analysis, and visualization. Related to these additional elements of working with large-scale or complex biomedical data, institutions have frequently raised the issue of cost as a central concern, realizing that this process may be expensive and burdensome, and that additional resources should be committed both at the level of the institution as well as by funding organizations. As noted in the *NIH Strategic Plan for Data Science*, funding data resources using the same approaches as research projects creates challenges during the review process. The agency should consider dedicated funding opportunities for tool development and database support.

Goal 4: Enhance Workforce Development for Biomedical Data Science

The AAMC is in strong agreement on the need to incorporate data science approaches into biomedical training, and the development of a workforce with the appropriate skills to build, support, and effectively utilize large datasets. We recommend that the NIH advance a comprehensive strategy for implementing training, increasing awareness, and incentivizing development of such programs, through the use of training grants or other existing mechanisms. The NIH should also ensure as much as possible that the training is current and fits existing needs and data types. The AAMC recommends that the NIH work closely with the extramural community to produce training modules as well as collect and disseminate best practices and curricula.

Goal 5: Enact Appropriate Policies to Promote Stewardship and Sustainability

The AAMC shares the NIH's commitment to the laudable goals of FAIR data, and appreciates the agency's intent to involve the research community in the development of policies for governance and data stewardship. We again note that meeting these specifications is not without cost, including time and support staff. We look forward to further guidance to clarify NIH's expectations for data curation, quality, and maintenance at the level of the individual investigator and institution. We also strongly support the creation of NIH-wide policies or harmonization across Institutes and Centers (ICs) whenever possible. The AAMC encourages the NIH to work with academic institutions before implementing this strategic plan, to collect feedback and promote the development of policies, standards, and programs

that are compatible or can be implemented along with existing practices whenever possible. We would be happy to assist in these efforts.

Consistent with movement in the direction of evidence-based policymaking, the AAMC supports the inclusion of robust evaluation measures in strategic planning and appreciates that the NIH has included this component in the *NIH Strategic Plan for Data Science*. We urge the agency to ensure that the measures of progress are reflective of the stated goal. For example, in the workforce evaluation, we encourage the incorporation of skills-based metrics and measurement of whether the new training is achieving longer-term outcomes, such as having a well-trained and diverse biomedical science workforce that pursues data science as a career or has sufficient knowledge to engage with data scientists.

The AAMC is again appreciative of the NIH's commitment to engaging the relevant stakeholders in the strategic planning process for data science, and would be happy to engage with our member institutions and work with the agency as it moves forward. Please feel free to contact me or my colleague, Heather Pierce, Senior Director for Science Policy and Regulatory Counsel at hpierce@aamc.org or (202) 478-9926 with any questions about these comments.

Sincerely,

A handwritten signature in blue ink that reads "Ross E. McKinney, Jr., MD". The signature is stylized and includes a circular mark at the end.

Ross E. McKinney, Jr., MD
Chief Scientific Officer