

March 15, 2024

Susan Gregurick, Ph.D.
Director, Office of Data Science Strategy
National Institutes of Health
Department of Health and Human Services

RE: Request for Information (RFI): Inviting Comments on the NIH Strategic Plan for Data Science 2023-2028

Dear Dr. Gregurick:

On behalf of the nearly 40,000 members of the American College of Emergency Physicians (ACEP) and the more than 8,500 members of the Society for Academic Emergency Medicine (SAEM), we appreciate the opportunity to respond to this request for information on the updated National Institutes of Health (NIH) Strategic Plan for Data Science.

The updated Strategic Plan for Data Science seeks to support the NIH mission and the goals of the Department of Health and Human Services (HHS) to increase data sharing, modernize data infrastructure, and develop artificial intelligence (AI) capacity. In the updated Strategic Plan, NIH identifies 5 overarching strategic goals:

- Goal 1: Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing
- Goal 2: Develop Programs to Enhance Human Derived Data for Research
- Goal 3: Provide New Opportunities in Software, Computational Methods, and Artificial Intelligence
- Goal 4: Support for a Federated Biomedical Research Data Infrastructure
- Goal 5: Strengthen a Broad Community in Data Science

NIH seeks comment on the appropriateness of the goals of the plan, the strategies and implementation tactics proposed to achieve them; including potential benefits, drawbacks, or challenges; opportunities for NIH to partner to achieve these goals; emerging research needs and opportunities that should be added to the plan; and, any other topic relevant for NIH to consider in developing this strategic plan.

Goal 1: Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing

NIH proposes that federally funded data repositories should adopt the Office of Science, Technology, and Policy (OSTP) Desirable Characteristics of Data Repositories and should align with community standards such as the Transparency, Responsibility, User focus, Sustainability, and Technology (TRUST) principles; Findable, Accessible, Interoperable, and Reusable (FAIR) datasets; and Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) principles. NIH will promote data repository interoperability by creating a FAIR-enabled data ecosystem that will break down data silos and promote greater findability and accessibility of data, thereby preventing unnecessary duplication of efforts and maximizing NIH investments. NIH identifies 3 main objectives in achieving this goal:

Objective 1-1: Support the Biomedical Community to Manage, Share, and Sustain Data

The NIH Policy for Data Management and Sharing established requirements that emphasize the importance of good data management practices and established the expectations for maximizing the appropriate sharing of scientific data generated from NIH-funded or -conducted research, with justified limitations or exceptions. By requiring researchers to anticipate their needs for managing and sharing scientific data, NIH will ensure that researchers develop data management and sharing plans that include where and how their scientific data will be shared and any anticipated limitations. NIH aims to shift the biomedical research culture into one in which data sharing and data reuse are the rule rather than the exception.

NIH will establish and enhance guidelines, processes, data sharing tools and training in data management and sharing and will explore funding and governance models to ensure a sustainable NIH data sharing infrastructure, for researchers, NIH staff, and for data stewards and librarians, including those at low resourced institutions.

ACEP and SAEM support incentivizing researchers to prospectively plan the management and dissemination of research data in order to make data accessible for the greater use of all research. We acknowledge that these data sharing plans will be useful in both driving new innovations and discouraging dataset silos, decreasing disparities across individual institutions. We commend the NIH's leadership in developing good data management practices and promoting the development of coordinated systems.

Objective 1-2: Enhance FAIR Data and Greater Data Harmonization

To further enhance the data ecosystem, NIH will encourage the use of community agreed upon standard schemas and metadata, enhance automated ontologies and automated curation processes, and create capabilities for greater data discovery and interoperability across multiple data repositories and knowledge bases. This objective will inform future activities under the NIH Plan for Public Access to Research Results. To implement this objective, NIH will enhance abilities to improve data and metadata quality, including Data QA/Q; encourage usage of open and standardized schemas, ontologies, and data formats; enhance automated processes for ontology use and for enhanced data curation including enrichment of metadata; and, create a minimal set of consistent and computable common data elements, or concepts, with consistent data models.

Objective 1-3: Strengthen NIH's Data Repository and Knowledgebase Ecosystem

To accelerate research, and to maintain participants' data protections, NIH seeks to streamline and semi-automate controlled access processes by developing, testing, and deploying the use of emerging technological advancements where feasible and appropriate. In addition, NIH seeks to develop common approaches and infrastructure for addressing data management incidents across NIH supported repositories. To implement this objective, NIH will:

- Enhance data repositories and knowledge bases that promote equitable access to all in alignment with the OSTP memo about Desirable Characteristics of Data Repositories for Federally Funded Research.
- Enhance FAIR, CARE, and TRUST capabilities that ensure secure and effective data management and promote data governance and data sovereignty.
- Support methods and programs with tribal communities to develop tribal data governance and sharing that recognize tribal rights in data.
- Promote shared data management practices, utilize open metrics for impact including enhancing data citation practices, and provide guidance on data preservation and long-term data archiving.
- Develop a comprehensive, coherent, and acceptable sustainability framework for identifying and supporting the portfolio of the most widely used and impactful NIH data resources.
- Develop a single policy framework that governs controlled data access repositories and standardized language for institutions and researchers.
- Streamline controlled data access processes across NIH repositories, including greater use of automation.
- Develop a common approach and infrastructure for addressing data management incidents across controlled access data repositories.
- Develop a single approach to help investigators find and appraise the relevance of controlled access data in NIH repositories, which enable metadata sharing.
- Enhance the visibility and use of NIH intramural research datasets and data resources.
- Develop methods to promote computational interoperability across data repositories and knowledge bases.

ACEP and SAEM agree with the necessity of enhancing data repositories and knowledge base ecosystems in order to support the storage and dissemination of individual datasets. We commend the NIH's efforts to ensure that shared datasets are available for use by the larger scientific community, thereby maximizing the advancement of medical treatments and improving health outcomes.

Goal 2: Develop Programs to Enhance Human Derived Data for Research

NIH will work with federal agencies, medical institutions, and health IT developers and vendors, where appropriate, to bridge the technology or data gaps between health care settings and clinical research to enable the biomedical and

behavioral research community to take full advantage of the multitude of health-derived data. To enable researchers to gather and integrate data of interest to address health related questions, NIH will improve access to data repositories that hold participant-derived data and will enhance abilities to link real-world data from multiple sources, with appropriate informed consent from the participants. NIH will support approaches to leverage or build on existing programs, bring new partnerships together to enhance clinical data science, and support cross-training between clinician researchers, data scientists, and other technical experts/stakeholders. NIH's overarching goal is to increase the use and utility of health care-derived data for research, with proper security and privacy safeguards.

ACEP and SAEM strongly support the concept of increasing the availability and access to human derived data for clinical and scientific research. Having a national database of emergency medicine data and outcomes could represent a great step forward in emergency care research.

Objective 2-1: Improve Access to and Use of Clinical and Real-World Data

NIH will enhance methods for informed consent in cases where data are combined from multiple sources and/or combined over longitudinal studies, with additional considerations for populations with health disparities; create, test, validate, and adopt methods to enable researchers to use multi-modal and digital data combined from multiple sources including through partnerships with other agencies, where appropriate; establish and promote standards for new types of health data, such as data captured from home health care devices; enable federated frameworks that will allow sensitive data to be utilized in clinical research, including fostering data linkages and interoperability across existing NIH supported real-world data platforms; develop ethical, governance, and policy frameworks to guide data linkages in different use case scenarios; and leverage existing agreements and infrastructure to create avenues for researchers to use and access health care and administrative datasets, enhancing participant awareness and consent of data use, especially for vulnerable populations.

Real-time data platforms such as the creation of dashboards monitoring hospital capacity and emergency department (ED) overcrowding in addition to other quality metrics could be a big step forward in public health. ACEP and SAEM support the creation of such platforms.

Objective 2-2: Adopt Health IT Standards for Research

To further advance NIH's goal to bridge the gap between health care settings and applied and clinical research, NIH will strengthen the use of ontologies with vocabularies and terminologies (e.g., SNOMED, LOINC) and exchange standards such as Fast Healthcare Interoperability Resources (FHIR). NIH will partner with health data standards bodies and organizations and other federal agencies that work with health data standards.

ACEP and SAEM encourage the adoption of health IT standards to support consistency in data transfer between healthcare settings and clinical research. ACEP has previously supported adherence to the Office of the National Coordinator (ONC) terminology, code sets, and syntax standards for data elements in ED and emergency medical services (EMS) information systems, with the intention of enabling interoperable data exchange with other electronic health record (EHR), health information exchange (HIE), and public health databases. We acknowledge the importance of exchange standards such as FHIR in advancing tools for clinical research and data exchange.

Objective 2-3: Enhance the Adoption of Social and Environmental Determinants of Health for Health Equity

NIH acknowledges the health disparities persisting despite technological advances and seek to broaden and enhance consensus-driven social determinant of health (SDoH) and environmental determinant of health (EDoH) standards for data capture and integration across a variety of systems. NIH will engage communities and stakeholders to develop demonstration projects, real world pilots and use cases to identify and implement SDoH and EDoH data and common data elements for specific diseases/conditions.

Emergency departments are an important part of the health safety net for marginalized populations. ACEP and SAEM encourage the NIH to include the emergency medicine community in these stakeholder discussions.

Using data science to develop a better understanding of SDoH and EDoH to produce more effective interventions resonates deeply with the mission to support policies that affect the SDoH and EDoH of our communities. The enhanced adoption of SDoH and EDoH is critical for the specialty of emergency medicine as we are in a unique

position to capture these data for the population, and we look forward to taking an active role in identifying the stakeholders in collaboration with NIH and other institutes.

Objective 2-4: Cross-disciplinary Training to Empower Clinical Data Science

NIH recognizes that to maintain and enhance clinical research informatics as a career path requires not only clinical training but also training in informatics, analytics, ethics, and standards. Thus, they will support cross-training between data scientists, clinical researchers, and nurses engaged in research at various stages of the academic tracks; develop training on consent practices and ethical use of data that go beyond legal and regulatory requirements with special considerations for linked/merged data and data from underrepresented communities; develop trainings on data sharing, management, transparency, provenance, and data quality for clinical research; and, create networking opportunities for clinical and data science researchers to develop collaborations, build teams, and learn from experts on these topics.

ACEP and SAEM agree and strongly support the training of emergency and acute care providers in data science and informatics. Training programs directed at emergency medicine physician scholars should be developed for this purpose. We believe that interdisciplinary data science training and networking would set the stage for powerful collaboration within the field.

ACEP and SAEM urge NIH to consider developing training programs such T32, K12, K08 programs specifically targeted towards emergency physician scholars for this purpose.

Goal 3: Provide New Opportunities in Software, Computational Methods, and Artificial Intelligence

To enhance the robustness and utility of data analysis and processing methods, NIH proposes to take advantage of new innovations in open and FAIR software and algorithms. NIH will support partnerships to codesign emerging capabilities including new methods in AI including generative AI, computational image analysis, and machine vision; new infrastructures such as quantum information sciences; automated workflows new tools for researchers to leverage data in a transparent, explainable, fair, and ethical manner; and, new ways of enabling communities to develop software through collaborative projects.

Objective 3-1: New Opportunities to Enhance Artificial Intelligence, Including Ethical AI for Biomedicine

Important goals for NIH are to enhance AI methodology and technologies that expand on the unique opportunities for biomedical and health research and to ensure that AI/ML capabilities are equitably beneficial across populations in the United States and globally. NIH activities align with and support the Blueprint for an AI Bill of Rights to ensure that AI algorithms and systems are used and designed in an equitable way. In partnership with the NIH ICOs, the agency will support emerging technologies and AI to integrate multiple streams of data including genomic, nutritional, sensor-based, social and behavioral, exposure, and community-level data to develop explanatory theoretical models, to inform prevention efforts, and to address health disparities. Support to increase opportunities for research communities to include concepts of diversity, equity, and inclusion in the development of trustworthy AI-enabled infrastructures and training is a priority. NIH will implement this objective by:

- Developing socio-technical solutions, including guidelines and principles, for ethical AI and to redress biases in training sets (containing ground truth) and algorithms, and support their effective assessment, validation, and adoption;
- Establishing and operationalizing community engagement for diverse, equitable and inclusive data, methods, and sources for AI.
- Supporting research in the development, validation, and use of synthetic clinical datasets for AI training and applications, when appropriate;
- Developing tools and training opportunities to help researchers create and prepare data that are FAIR and AI-Ready, including ontologies, schema, and data quality measures;
- Supporting the development of AI models, with appropriate metadata (model cards) that are explainable, transparent, and FAIR;
- Leveraging new technologies and methods for foundational models to accelerate biomedical and behavioral research;

- Supporting opportunities to develop new AI technologies that will enable the translation of data to knowledge, including AI tools to enable data cleaning, harmonization, integration, and metadata collection; and,
- Enhancing NIH capabilities in AI through partnerships across federal agencies and communities to develop new methods in AI.

ACEP and SAEM support this initiative and agree with the importance of developing guidelines for ethical AI and redressing biases in training sets. We acknowledge the utility of developing new AI technologies and methods and applaud the emphasis on developing AI/ML to benefit all populations. Though there is great potential for AI methodologies and technologies to enhance biomedical research, it is important to recognize the structural barriers in place in lower-income and rural communities that may limit access to these advancements. We therefore support efforts to ensure diverse, equitable, and inclusive data, methods, and sources for AI.

Objective 3-2: Develop Cutting Edge Software Technologies

NIH is poised to take advantage of the integration of real-world devices, the increased scale of computational resources and significant automation in software and algorithms to advance biomedical discoveries and innovation. They will adopt and adapt emerging and specialized methods, algorithms, tools, software, and workflows for biomedical and behavioral scientific discovery; enhance tools and workflows for greater automation, while maintaining robust ethical standards; leverage advances in computational methodology and studies to create new opportunities for ethical and social science research; and, investigate the potential of digital twinning approaches to organs, systems, individuals, and populations.

ACEP and SAEM support the NIH in leading the advancement of biomedical discoveries and innovation via the development of cutting-edge software technologies. We appreciate the necessity of developing multi-dimensional data integration in tandem with respecting privacy and ethical concerns.

Objective 3-3: Supporting FAIR Software Sustainability

FAIR software principles foster practices to ensure that research software is sustained by larger biomedical research communities over time. NIH recently issued best practices for software sharing that align with the FAIR software principles. They seek to enable investigators to adapt and enhance software and tools to take advantage of new technologies and computing paradigms and to optimize research software for robustness and ultimately to increase software sustainability by enhancing community-focused software development and dissemination, improve visualization tools to support the scale and variety of modern biomedical data, establishing metrics and best practices for software sustainability and integrate these into software development lifecycle, facilitating research activities for software engineers and biomedical and computational researchers to collaborate, developing mentorship programs that pair experienced software engineers with early-career researchers and software developers, and exploring innovative models for public-private partnerships to support software and data innovation and sustainability.

ACEP and SAEM support these various NIH-led initiatives aimed at fostering collaborative, innovative, and FAIR software development with the ultimate goal of software sustainability. We commend NIH's foresight in prioritizing sustainability.

Goal 4: Support for a Federated Biomedical Research Data Infrastructure

NIH believes that they have a responsibility to integrate new data ecosystems and resources into a federated data infrastructure that leverages ideas from industry and cutting-edge research with the goals of providing: 1) easier access to and use of data across multiple Institutes supported data platforms, 2) economies of scale for NIH to support and maintain shared tools and capabilities, 3) opportunities for communities to collaboratively develop and share new methods and workflows, and 4) oversight by the community for greater transparency and autonomy of data use. In collaboration with the NIH ICOs, the agency will support development of innovative data sharing platforms, data analytics, and their integration.

Objective 4-1: Develop, Test, Validate, and Implement Ways to Federate NIH Data and Infrastructure

NIH seeks to develop comparable databases on health outcomes, risk factors, and determinants of health disparities through a common and coordinated data access process with shared policy and governance. NIH's vision is to build

a connected and federated data ecosystem to ensure that data repositories can be used together rather than in isolation. NIH will support and enhance a federated biomedical data research infrastructure that will create, test, validate, and implement a set of sharable services (e.g., common search capabilities, application programming interfaces (APIs), identify and access management (IAM) services, and workspaces/sandboxes).

ACEP and SAEM strongly agree with this vision. We recommend that NIH consider the creation of a central data repository similar to PubMed for data storage. As emphasized in the strategic plan, an important component of this objective is to improve efficiencies and maximize the ability of researchers to find, access, and use data. At present, a fragmented approach is developing, which considerably limits data accessibility.

Goal 5: Strengthen a Broad Community in Data Science

NIH seeks to help create a vibrant, innovative, and inclusive data science community to help ensure that data science advances in biomedical research can benefit all populations.

ACEP and SAEM agree with this initiative in principle. To achieve this, NIH should commit to the training of emergency physicians in informatics and data science management as well as the creation of emergency medicine-specific clinical databases for the purposes of improving systems of clinical care in our nation's emergency departments.

We appreciate the opportunity to provide comments on this RFI. If you have any questions, please contact Erin Grossmann, ACEP's Manager of Regulatory and External Affairs, at egrossmann@acep.org, or Melissa McMillian, SAEM's Senior Director of Foundation and Business Management, at mmcmillian@saem.org.

Sincerely,



Aisha T. Terry, MD, MPH, FACEP
ACEP President



Wendy C. Coates, MD
SAEM President