AcademyHealth

AcademyHealth Response to Proposed NIH Strategic Plan for Data Science

Do you have comments on the appropriateness of the goals of the plan and of the strategies and implementation tactics proposed to achieve them?

As a member-based organization that serves the research community, AcademyHealth engages in advocacy to support its vision to improve health and the performance of the health system by supporting the production and use of evidence to inform policy and practice. In this spirit, AcademyHealth commends the NIH for setting forth a proposed Strategic Plan for Data Science with its five Overarching Goals and related Strategic Objectives. Our community appreciates the extent to which NIH is responsive to the recent guidance and previous public commentaries from a variety of stakeholders that led to the creation of this proposed plan.

In recognizing the vast and complex problems facing healthcare systems, our collective ability to have an impact on them through research and dissemination of results will depend on strategies to transform data into information and ultimately actionable knowledge.
The overall plan focuses heavily on architecture, infrastructure, and tools. Goal 5 addresses policies focused on privacy and security of data. However, it could be further enhanced by also addressing regulations. Some of the challenges to adoption of multi-institutional data science and FAIR principles are driven by regulation. An assessment of the regulatory barriers to distributed data science applications as well as a commitment to addressing those barriers is needed. Objective 5-1 notes the need to protect confidentiality and intellectual property, but is missing tactics that address those points.

The need for a compelling strategy to harness data has never been greater. The membership of AcademyHealth recognizes that this plan is an important opportunity for the scientific community to embrace new methods, technologies, and dimensions of research that could not have been undertaken before. In our response to this endeavor, AcademyHealth encourages NIH to take a bold view of the plan and place a premium on the objectives and tactics that will enable and support high-risk, high-reward hypotheses generation, testing and research. **The goals of the plan closely align with AcademyHealth's stated priorities of support for policies that enhance the quality, availability, timeliness, transparency, affordability, and governance of data and tools used to produce research and improve healthcare.** As our computational capabilities continue their rapid expansion, our research community should be empowered to undertake projects that leverage and integrate large datasets to unlock hidden answers and associations that will transform health and healthcare.

1666 K Street NW
Suite 1100
Washington, DC 20006
202 292 6700 PHONE
www.academyhealth.org

As a result of our future research being supported by the components of this plan, we can attain the goal of applying evidence-based practices as guiding principle across the spectrum of health care.

We find that the plan has sound principles to support the continuum of research. Our collective community ascribes to the goals underpinned by the FAIR principles for responsible data management although we believe that more could be added to guide achieving them in a complex environment that also addresses the commercial and intellectual property of innovators. We also suggest that the plan acknowledge that these principles will need to evolve and adapt rapidly to meet the conditions of the time.

Data Science Strategies in Support of Health Services Research

As health services research (HSR) is the primary field represented by our community, we have specific findings to share regarding implications that can be anticipated from the plan. The implications for the NIH data strategy to enable cutting edge data and analytic foundations for health system problem formulations addressing domains such as quality, safety, comparative effectiveness and clinical system design is not explicitly addressed. As exascale-level computing emerges and other clinical data expand beyond large-scale EHR data resources, will the infrastructure and policies enable HSR studies to more realistically mimic the speed of care including clinician, patient and system decision making? Recognizing that NIH funds HSR, dissemination and implementation research, and clinical and translational research studies, additional information could be included in the plan to address the nature and opportunity of these studies to rapidly inform health and healthcare.

Major advances to the health services, biomedical and public health research enterprise have been realized by the harnessing of new tools and analytic capacity attributable to data science related innovations. These advances in data science, computational analytics, and artificial intelligence have yielded enhanced data integration platforms, greater computational power, advances in pattern recognition and signal detection, improvements in fraud detection and predictive analytics, and more effective data security processes. Evidence development processes, population health research, and health disparities studies will be advanced as a result of strategic initiatives to enhance these capabilities and we encourage prominent and consistent inclusion of these as important dimensions of the plan. As envisioned, this modernized integrated data platform/system should serve to visibly increase the analytic utility to much broader health and healthcare research enterprise beyond biomedical researchers and related stakeholders.

In addressing specific issues, in support of Objective 1-2, we advocate for a more expansive description of how the accommodation and incorporation of epidemiological data, clinical care processes, social indicators and population health will be addressed

in future data science research. We emphasize that the data sharing objectives should be expanded beyond biomedical researchers to also support policy analysts, funding agencies, professional organizations and the public. However, healthcare delivery systems have business interests in which information released could be used adversely against them by competitors and vendors. The NIH strategic plan for data science should address how these business interests are protected and reward institutions who contribute to the public data and support FAIR principles.

What opportunities do you see for NIH to partner in achieving these goals?
To further advance the analytic utility of the envisioned data platform/system to the greater health, healthcare and biomedical research and policy community at large, greater connectivity to overall Department of Health and Human Services (DHHS) data infrastructure should be considered an essential goal. NIH is a member of the DHHS Data Council, which is the principal internal advisory body for the Secretary of HHS on data policy. NIH should consider the inclusion of well-specified coordination and partnership goals with the Data Council to ensure greater connectivity to all DHHS data sources, thereby enhancing the analytic capacity, utility, interoperability and efficiencies both within NIH and DHHS. In addition, this should be implemented keeping in mind the recommendations of the Commission on Evidence Based Policymaking and its emphasis on data stewardship as part of the infrastructure to advance the use of evidence in decision-making.

AcademyHealth also recommends that a NIH consider establishing an external advisory group to provide ongoing input from non-NIH stakeholders, including HSR investigators in both academia and as well as those embedded in healthcare delivery systems on the implications of data science for training and career development programs and approaches. The current plan places an emphasis on procurement partnerships to efficiently leverage growing and dynamic commercial technical capabilities and does not project collaborations and partnerships involving more balanced approaches to shared data, analysis or priority setting with commercial and non-NIH entities. This may be particularly important to expand the data and analytic support for interventions that rely on community or commercial actions for impact at the health care delivery system level.

Are there additional concepts that should be included in the plan?

Reproduction, Translation and Dissemination of Research Results
To support the reproducibility of research results and to assist in facilitating their dissemination, NIH should consider requiring researchers to submit their algorithms and code to PubMed Central in a standardized format and include an open-source agreement so that industry and other researchers in a range of settings (e.g. academic, health system, contract research organization, etc.) can integrate and advance the use and utility of them.

Importance of Supporting New Forms of Data and Analytic Approaches
Recent legislative efforts have established new authorities and policy consideration for new forms of data to be used in the full range of health research. Real world evidence (RWE), or data regarding the usage or the potential benefits or risks, of a drug, device or other preventive, diagnostic or therapeutic approach derived from sources other than randomized clinical trials, is an important element of the spectrum of data resources that should be considered in the strategy. Among the important aspects of NIH supported research that could enhance the utility of RWE is the establishment of standardized data catalogs, and incorporation into a complex data mart. Further, the plan could consider incorporating concepts to identify common and unique statistical variables and formats across complex data sources to enable integration and connectivity among them. In addition, NIH research strategies could identify the importance of machine learning techniques to support computation of clinical variables to support enhanced and automated evidence development and validation methods.

In developing the data infrastructure to enhance data storage capacity and security, and advance greater utilization of cloud computing, the strategic plan indicates that NIH will leverage what is available in the private sector, either through strategic partnership or procurement, to create a Platform as a Service (PaaS) environment. We encourage that steps be taken when developing this capability to incorporate existing legacy capabilities to ensure timely and efficient integration of all data storage and management resources. AcademyHealth recognizes the emergence of large-scale private data analytics firms that are making significant inroads toward analytics, such as artificial intelligence, in healthcare. The software and computing applications to accurately extract, analyze and archive data from digital health records and an ever growing range of devices is contributing substantially toward a holistic data experience in clinical medicine.

The Research Workforce of the Future and Collaborations
AcademyHealth applauds the emphasis on developing a data science workforce. The strategic plan addresses ways to address gaps in scientific workforce diversity. Further, we encourage NIH to explore impediments to career development that are inculcated within the culture of academic and disciplinary research communities and work to overcome them. We encourage a strong commitment to engaging the trans-NIH workforce equity resources to diversity the peer review process for extramural grants, increase the diversity among the NIH program officer workforce; modify the requirements for diversity of teams in award notices for extramural grants, and deliberately diversify other resource infrastructure assets. In addition, STEM research programs focused on early career stages will be crucial to ensuring that a vibrant community of new investigators with data science skills and capabilities are able to participate. Further, AcademyHealth suggests that concentrated courses designed for current investigators be developed to help them establish a broader range of

computational skills and analytic methods to fully actualize the potential of big data and data science.

Are there performance measures and milestones that could be used to gauge the success of elements of the plan and inform course corrections?

The evaluation goal 2 measures list includes "quantity of datasets deposited…" The potential measures should also address quality and completeness of the data as well as metadata and documentation on how to use it. Measuring only quantity rewards data dumps that may be of little lasting value. An alternative measure would be use of the datasets deposited.

Under evaluation goal 4, it would be advised to add in measures for data-science programs developed or sponsored by NIH and used by the research workforce. Understanding the size, scope and capacity of the data science workforce is important to ensuring the achievement of the stated goals. Regular assessment and reporting on the size of the workforce, as well as evaluation and assessment of efforts to improve diversity would add to the metrics and performance measures of this plan.

Is there anything else you think NIH should consider in developing this strategic plan

AcademyHealth acknowledges the importance of a strategic approach to data science for many reasons. The strategic plan will serve our community as an important guidepost for shaping new initiatives, acquisition of new skills and capabilities, and the unlocking of important research ideas and projects. Execution of the strategic plan will depend upon many communities to engage and support their stakeholders in enhancing the quantitative nature of research in health and health care. Our community stands ready to work with NIH in building these new capabilities, collaborations, and competencies in data science to support health services research and evidence-based healthcare.

We appreciate the opportunity to contribute our views toward this important plan.