July 27, 2018

Ms. Karen Dunn Kelley
Under Secretary of Economic Affairs
Acting Deputy Secretary
Department of Commerce

Submitted electronically at: Reglations.gov

Dear Acting Secretary Kelly:

The American Medical Informatics Association (AMIA) appreciates the opportunity to provide input regarding the Draft Federal Data Strategy.

Health informatics is a nearly 60-year-old field of study concerned with data collection, analysis, and application within broad domains of health, including healthcare delivery, public health, consumer health, clinical research, and translational research. AMIA is the professional home for more than 5,500 health informatics professionals, representing front-line clinicians, researchers, educators and public health experts who bring meaning to data, manage information and generate new knowledge across the health and health care enterprise. As the voice of the nation's biomedical and health informatics professionals, AMIA members advance health and wellness by moving basic research findings from bench to bedside, and evaluating information and communication technology interventions, innovations, and public policy across care, research and community settings.

AMIA strongly supports development of this strategy and we consider the coordinated management of federal data and information a national imperative. We note that this strategy builds on a long pedigree of federal efforts to manage federal data as an asset, stretching across decades of policy. Among more than a dozen official actions of federal legislation, regulation, and directives are the convictions often expressed by and for the American public that their data should be secure, private, and appropriately leveraged for public benefit. These same actions dictate that administrative data produced by the federal government should be accessible, discoverable, and usable by the public. To advance these convictions, AMIA recommends the Federal Data Strategy articulate how federal agencies should:

1. Collect or create information in a way that supports supplemental uses of downstream information processing and dissemination;
2. Extend the concept of "data as an asset" to grantees and others who receive federal funding so that data generated as part of grants and contracts are more readily findable, accessible, interoperable, and reusable;
3. Develop a framework for agencies to understand, manage, and compare their data assets through a portfolio approach; and
4. Continue efforts to modernize internal IT investments, especially around large-scale computing and shared environments.

July 27, 2018

To achieve these recommendations, especially around supporting downstream use and comparing data assets, there are two critical needs: (1) Publishing data dictionaries; and (2) Harmonizing data dictionaries across agencies. Data dictionaries that identify the element name, data type and meaning for each data point in a database, must be published so both the public and other federal agencies can interpret the data. In addition, these data dictionaries should align across agencies. For example, a field called 'Zip' in one data set, should not be called 'zip code' in another. AMIA recommends that this Strategy initiate a collaborative, inter-agency consensus process to ensure that data elements used by disparate agencies, but which refer to the same concept, are named the same everywhere. These two fundamental issues are essential to the outcomes sought by the Draft Federal Data Strategy.

AMIA appreciates that the Federal Data Strategy begins from well-articulated and comprehensive Principles. Given our members' expertise in the systematic collection, analysis, and application of data for health, AMIA has developed a series of Policy Principles for priority areas within the domain of Health Informatics Policy.[1] One such priority area pertains to Data Sharing.[2]

Below, we offer commentary and amendments to the important concepts that the Federal Data Strategy should consider as it works towards its Year 1 Action Plan. In reviewing the Strategy Principles, we recommend explicit mention of two inter-related concepts, highlight an omission that supports one of our four key recommendations, and offer implementation support for a key principle of ethics.

First, the concepts of data integrity and trust are not explicit across the concepts. We view trust as prerequisite for data to be of value and we view data integrity as a condition important to maintain the meaning of data. We recommend the concept of trust in the accuracy and integrity of data be included as part of the Quality Subprinciple Five, "Ensure Relevance," with the following edit: "*Engender trust in the accuracy and integrity of data through validation that they* are high quality, useful, understandable, timely, and needed." Given our familiarity with policies and principles produced by the National Institutes of Health, we would be remiss in omitting mention of the All of Us Privacy and Trust Principles,[3] which may further inform this work.

Second, we note in our fourth key recommendation that efforts must be made to continually modernize internal IT investments. AMIA's Policy Principles note that, "The advantages of data sharing can only be realized with appropriate levels of investment in underlying infrastructure, including tools for managing, storing, and indexing increasingly large and diverse data sets, as well as, human resources for curating shared data." As such, we recommend that Continuous Improvement Subprinciple Seven, "Demonstrate Responsiveness," be amended with the following:

[1] "AMIA Public Policy Principles and Policy Positions, 2016 – 2017," available at http://bit.ly/2gPB52N
[2] Ibid., Data Sharing in Research Policy Principle (pg. 10)
[3] Precision Medicine Initiative: Privacy and Trust Principles. Available at: https://allofus.nih.gov/about/program-overview/precision-medicine-initiative-privacy-and-trust-principles

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

July 27, 2018

"Improve data sharing and access with ongoing *investment in technical infrastructure and human resources and provide opportunities for continuous* input from users and other stakeholders."

Third, we greatly appreciate the Stewardship Principle 2. Uphold Ethics. A 2018 White Paper[4] articulated that "decisions related to data access and data sharing are fundamental expressions of social and ethical norms in a modern, connected society," and that "Public policy should encourage socially responsible and ethically consistent data access and data sharing." We see these values expressed in Principle 2, and offer the following recommendation to help ensure this principle is adhered to: We recommend establishment of a U.S. Federal Government Interagency Ethics Working Group to consider the social and ethical dimensions of emerging technologies and an explicit focus on data as a social good and data sharing.

Finally, we wish to highlight an important informatics project in the state of Indiana that may serve as a valuable Use Case for the Federal Data Strategy. As part of Indiana University's Grand Challenge Initiative,[5] the "Responding to the Addiction Crisis," includes the Indiana Addiction Data Commons, or IADC.[6] The IADC seeks to provide a more holistic characterization of the opioid crisis in Indiana by giving researchers and health care professionals access to information beyond what can be found in the electronic health record, including state department of health data, criminal justice and treatment resource data, and other kinds of environmental, economic, and demographic data maintained by various public and private sector partners. While this project is in its early stages, it could provide useful information on governance, infrastructure, and data standards that could be useful for the Draft Federal Data Strategy.

We reiterate our support for this plan, and AMIA stands ready to help ensure this effort has the requisite expertise to accomplish these worthy goals. We have provided details to our key recommendations and provided additional information in the enclosed document. Should you have any questions or require additional information, please contact AMIA Vice President for Public Policy Jeffery Smith at jsmith@amia.org or (301) 657-1291 ext. 113. We look forward to continued dialogue.

[4] Redefining our Picture of Health: An AMIA Policy Invitational White Paper. 2018. Available at: https://www.amia.org/sites/default/files/API-2017-White-Paper-Redefining-our-Picture-of-Health.pdf
[5] https://grandchallenges.iu.edu/about/index.html
[6] https://www.regenstrief.org/projects/indiana-addiction-data-commons/

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

July 27, 2018

Sincerely,

Douglas B. Fridsma, MD, PhD, FACP, FACMI
President and CEO
AMIA

(Enclosed: Detailed Comments to Draft Federal Data Strategy)

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

July 27, 2018

Below we articulate rationale and provide additional information for each of the four recommendations made in the body of the letter.

1. Collect or create information in a way that supports supplemental uses of downstream information processing and dissemination.

The E-Government Act of 2002 was foundational in compelling federal agencies towards a more proactive open government stance through "the management and promotion of electronic Government services and processes… and by establishing a broad framework of measures that require using Internet-based information technology to enhance citizen access to Government information and services."[7] The E-Gov Act was a hybrid performance and information access bill that put in place agency and government-wide policies that would enable improved performance, greater access to information, and improved citizen participation with government. In the nearly 20 years since, OMB has magnified the concept of citizen-centric and open government through various memoranda and directives. At the core of these policies is a need for agencies to be deliberate in how they collect and create data so that such data could be used for supplemental purposes.

For this to be the case, the Federal Data Strategy will need to help agencies institute and update their individual policies and procedures. This will be difficult, given the diversity of data and collection activities, but there are both frameworks for which this Strategy could rely[8] and a good repository of tools and resources at: https://project-open-data.cio.gov/. Each agency can take a first step down this path by publishing their data dictionaries for each data set that is not confidential.

2. Extend the concept of "data as an asset" to grantees and others who receive federal funding.

In comments submitted earlier this year to the NIH regarding its Data Science Strategic Plan,[9] AMIA members identified several key recommendations to ensure that the NIH achieves its data science goals. One specific recommendation has relevance for the Federal Data Strategy and is reproduced below:

> AMIA recommends the NIH declare that all data generated through its grants must align with FAIR data practices. On page 3 of the Plan, it is stated that, "this strategic plan commits to ensuring that all data-science activities and products supported by the agency adhere to the FAIR principles…" The NIH must go beyond adherence to FAIR principles and require that grantees also adhere to such principles as a condition of funding. Moreover, the NIH must develop policies that incentivize adherence to FAIR principles, and develop

[7] E-Government Act of 2002, Public Law 107–347, accessed 23 March 2013 (http://www.gpo.gov/fdsys/pkg/PLAW-107publ347/html/PLAW-107publ347.htm)
[8] OMB Memorandum M-13-13
[9] AMIA Response to NIH Data Science Strategy, April 2, 2018. Available at: https://www.amia.org/sites/default/files/AMIA-Response-to-NIH-Data-Science-Strategic-Plan-RFI.pdf

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

> capacity to assist grantees in this endeavor. To do this AMIA recommends NIH make Data Sharing Plans, including Genomic Data Sharing Plans, a "scorable" element of grant applications subject to the existing policies.[10, 11] Further, NIH should include not just data, but software code and algorithms as required elements within the "rigor and reproducibility" section of grant applications.[12] Data sharing has become such an important proximal output of research that we believe the relative value of a proposed project should include consideration of how its data will be shared. These kinds of changes will be force-multipliers for the NIH, as they will encourage broad improvements to how data are collected, shared, and reused. Making Data Sharing Plans scorable would enable those projects that prioritize systematic and strategic data sharing, through use of standards and accepted best-practice, to garner higher scores. By using the peer-review process, we will make incremental improvements to interoperability, while identifying approaches to better data sharing practices over time.

While the focus of this recommendation was in regard to what the NIH could do to ensure grantees think strategically about "data as an asset," AMIA encourages the Federal Data Strategy to include similar policies and guidance for federal agencies to consider for their grantees, contractors, and non-government partners.

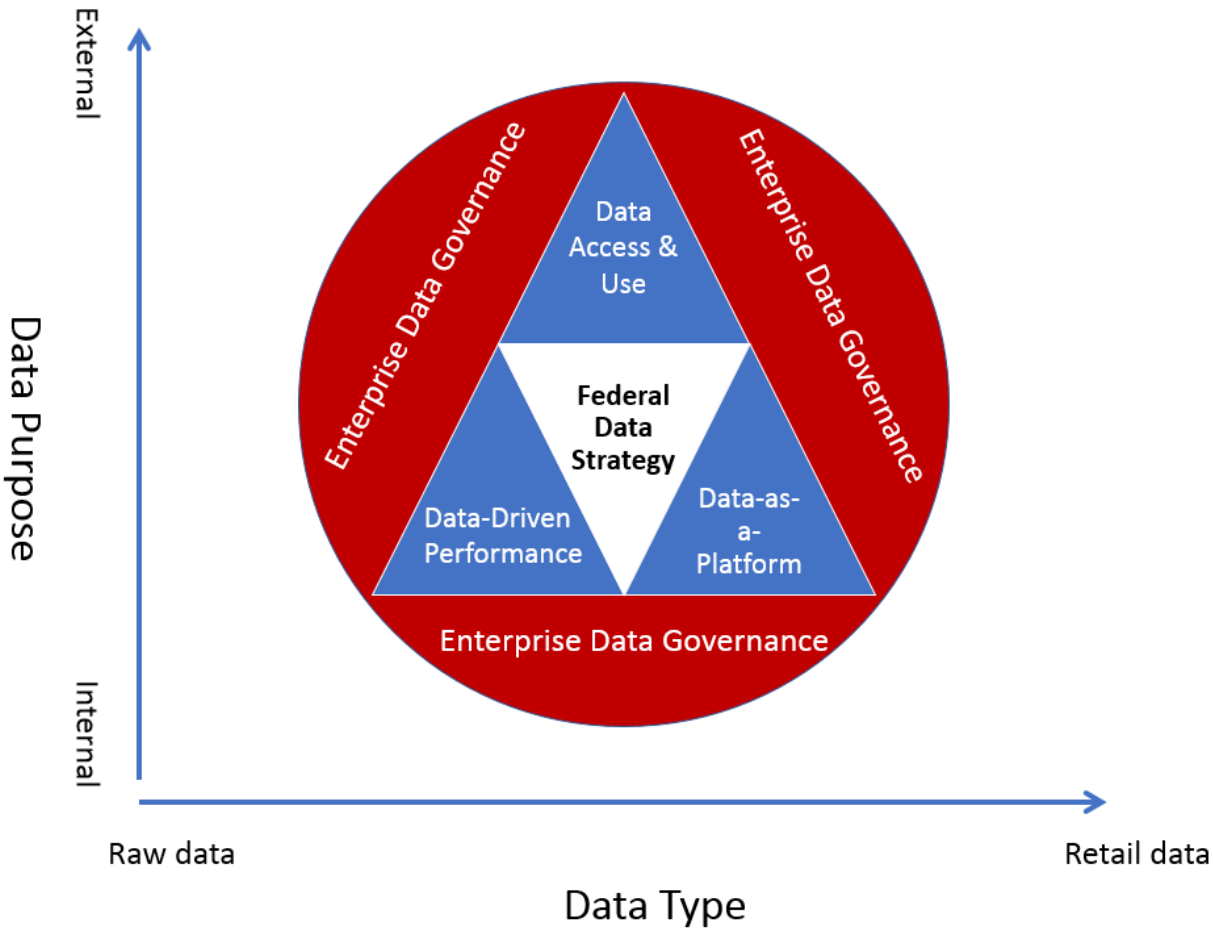
3. Develop a framework for agencies to understand, manage, and compare their data assets through a portfolio approach.

As we understand it, the purpose of a Federal Data Strategy is to have a "coordinated and integrated approach to using data to deliver on mission, serve the public, and steward resources while respecting privacy and confidentiality." Further, we understand the four central tenants of the Federal Data Strategy currently are focused on the quadruple purpose of (1) managing government data as a strategic asset; (2) enabling stakeholders to effectively and efficiently access and use data assets; (3) improving the use of data assets for decision-making and accountability; and (4) facilitating the use of data assets by external stakeholders for commercial ventures, innovation, or for other public uses. A key component of making this a reality will be to gradually migrate all data dictionaries to use common element naming.

For individual agencies to understand, manage, and compare their data assets, we recommend the Data Strategy include a framework that includes scales for data purpose (internal versus external) and data type (e.g. raw, machine readable data and synthesized/retail data on the other) and categories reflective of the Strategy's main components. One consideration for the Data Strategy is to distill its main components to deliver on the following outcomes: (1) improve data access and use by the public; (2) encourage data-driven performance; and (3) enable data-as-a-platform for

[10] National Institutes of Health, "NIH Data Sharing Policy and Implementation Guidance," March 2003 https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
[11] National Institutes of Health, "National Institutes of Health Genomic Data Sharing Policy," August 2014 https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf
[12] National Institutes of Health, "Principles and Guidelines for Reporting Preclinical Research," https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research

commercial venture and innovation. By including a scale of data type and external versus internal use of data, the framework could help agencies view their various kinds of data assets as a portfolio, and the Strategy could set goals for different portfolio types. A mock-up framework is below:



The main takeaway is that agencies will have numerous data assets and various initiatives to make these assets available for public use; innovation and commercial purposes; and performance and accountability. The Data Strategy must have some way to help agencies manage these numerous and diverse data assets as well as be able to dictate an optimal organization of these data assets.

4. Continue efforts to modernize internal IT investments, especially around large-scale computing and shared environments

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

July 27, 2018

In 2017, AMIA responded to another NIH Request for Information regarding "Strategies for NIH Data Management, Sharing and Citation."[13] One of the top recommendations AMIA offered was that the NIH earmark support for data sharing as part of applicable grants' direct costs. We noted that dedicated funding from research sponsors for data curation and donation efforts are necessary to share, collaborate, and advance data sharing capabilities. To our satisfaction, the aforementioned NIH Data Science Strategic Plan included plans to develop "separate funding strategies, review criteria, and management" for databases, knowledgebases, and tool development. We noted that so too will investments in data infrastructure be necessary to facilitate data stewardship, curation, and maintenance.

These same recommendations apply to the Federal Data Strategy. Very little progress can or will be made towards the Strategy's goals unless the underlying infrastructure and human resources are in place to support those goals. Data Commons are used by the FDA, as well as by various Centers and Institutes at the NIH. We would do well to leverage the lessons learned from the use of these shared environments.

[13] AMIA Response to NIH Data Management, Sharing, and Citation RFI, January 2017. https://www.amia.org/sites/default/files/AMIA-Response-to-NIH-Data-Management-Sharing-and-Citation-RFI_0.pdf

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814