November 1st, 2017

Dr. Patricia Brennan
Director, National Library of Medicine

Dear Dr. Brennan,

The Alliance for Nursing Informatics (ANI) advances nursing informatics leadership, practice, education, policy and research through a unified voice of nursing informatics organizations. ANI has reviewed the Request for Information (RFI) for: *The Next-Generation Data Science Challenges in Health and Biomedicine.*

We offer our comments as nursing stakeholders aligned with the information requested for each of the three focal areas identified by the NLM. Please find each focal area and information requested by NLM in bold and italics, followed by our response below.

1. **Promising directions for new data science research in the context of health and biomedicine.** *Input might address such topics as Data Driven Discovery and Data Driven Health Improvement.*

With a cross-disciplinary focus, ANI supports expanding the reach of information technology and artificial intelligence (AI) for the improvement of individual and population health across the health and illness spectrum.

Research areas the most critical for NLM to conduct or support.

ANI recommends that the following research areas are a priority for the NLM:
- Support the collection and cataloging of multisource and multi center open platform simulation tools for testing Information Technology (IT) and Data Science initiatives before use with patients/citizens, clinicians, and public health and safety personnel.
- Development of open platform software tools for Data Science initiatives that aid in scientific discovery and clinical applications in prevention and management of disease; as well as promotion of health and well-being.
- Further development and application of advanced statistical methodologies to capture meaningful insights from the large volume and variety of multi-level data sources relevant to precision prevention and care.
- Support all phases of the Data Science scope to support data-driven initiatives that use the full range of health care provider and patient data in the continuum of care, as well as additional data from the environment wherein patients are connected.
- Integrate public health, build environmental, healthcare, and ecological data for better understanding of holistic determinants of human health and well-being.
- Provide funding for systematic review and meta-analyses as a key method for knowledge translation to clinical care.

Other comments, suggestions, or considerations, keeping in mind that the aim is to build the NLM of the future.

ANI highlights the following additional areas and domains for building the NLM of the future:

- It is important to highlight the need for innovative information technologies that are inclusive of clinical conditions, domains, and the continuum of care. ANI endorses NLM's strategic role in advancing the adoption of Data Sciences research across the spectrum of disease, health and wellness states.
- Data Science approaches should continuously seek to expand from a focus on a single disease in isolation towards multi-level precision health discovery, simultaneously integrating a variety of data from genetic predisposition to population-level determinants of health to develop insights.
- Continue to support strategies to ensure that all phases of Data Science research receive the appropriate resources for knowledge discovery solutions that protect confidentiality and privacy.
- Advance cutting-edge methodologies and initiatives that other NIH Institutes do not focus on to ensure that all phases of Data Science research are feasible and trustful, such as: data generation, characterization, management, storage, analysis, visualization, integration and use of large, heterogeneous data sets that have relevance to health, precision-health, and biomedicines.

2. **Promising directions for new initiatives relating to open science and research reproducibility**. *Input might address such topics as Advanced Data Management and Intelligent and Learning Systems for Health.*

   ANI supports the following activities and new initiatives:

   - Establish and curate public de-identified datasets from a wide range of healthcare, administrative, and other related databases that can be used for a multitude of purposes from knowledge discovery, reproducing research results, improving data science methods to predictive analytics. Availability of datasets would decrease the time from research proposal to publications and translation of results into practice. One of the biggest challenges for researchers involved in data science projects is acquiring data. The process to acquire data can be long and arduous delaying much needed research that significantly impacts the ability to improve health and wellness.
   - Provide free platforms for individuals to voluntarily upload symptom, disease, and citizen science data.
   - Support open source software to be used in developing best-practice use to develop insights from large volume and variety data sources. Provide quality assurance of and training on the open source software supported.
   - Increase research reproducibility by suggesting or requiring that de-identified datasets obtained using federal government funds be archived within the NLM. These datasets would then allow other researchers resources to test similar or improved research methods and pool data from multiple sources for increased precision, interpretation, and generalizability of data.

3. **Promising directions for workforce development and new partnerships.** *Input might address such topics as Workforce Development and Diversity and New Stakeholder Partnerships.*

ANI suggests the following approaches to improve workforce development:

- Improve workforce development by providing continuing professional education to achieve role-based informatics competencies, including nurses in all professional roles. To support this goal, we recommend that the NLM's role include serving as a digital library for all free, open-source informatics training courses that currently exist. In this role, the NLM could perform curation of course content to guide self-learning based on competency assessments. ANI commends the NLM's support of nurses in the area of informatics training and workforce development and encourages NLM to expand these important activities. Data-driven and evidence-based identification of informatics-related educational needs of the healthcare workforce research is needed.
- Expand workforce training in NLM genetic and genomic resources toward population and precision health initiatives.
- Expand clinician scientist training on the use and integration of contextual determinants of health, to include data captured from myriad sources and at various levels of analysis for enriched clinical research insights (e.g. environmental sensors, wearables, neighborhood poverty, census data).

ANI appreciates the opportunity to contribute to the conversation on the Next-Generation Data Science Challenges in Health and Biomedicine aligned with the NLM's efforts to strengthen and expand the NIH's scope of its investments in data science. We believe that the priorities ANI has endorsed in this letter can have a significant impact on improving data and information access to enhancing health and wellness.

Sincerely,

Charlotte Weaver, PhD, RN, MSPH, FHIMSS, FAAN
ANI Co-chair
E-mail: caweaver2011@gmail.com

Mary Beth Mitchell, MSN, RN, BC, CPHIMS
ANI Co-chair
Email: marybethmitchell@texashealth.org