

Association of College & Research Libraries
50 E. Huron St. Chicago, IL 60611
800-545-2433, ext. 2523
acrl@ala.org, <http://www.acrl.org>



TO: National Library of Medicine
RE: Next-Generation Data Science Challenges in Health and Biomedicine
DATE: Tuesday, October 31, 2017

Submitted online at <https://www.research.net/r/NLMDataSci>

To Whom It May Concern,

On behalf of the Association of College and Research Libraries (ACRL), I am writing to offer comments on next-generation data science challenges in health and biomedicine.

SECTION 1: Promising directions for new data science research in the context of human health and disease might address such topics as these:

1. Data-driven Discovery - areas such as:

The Association of College and Research Libraries (ACRL) writes in response to the request for information issued September 26, 2017, by the National Library of Medicine on next-generation data science challenges in health and biomedicine.

ACRL is the higher education association for librarians. Representing more than 10,000 academic and research librarians and interested individuals, ACRL (a division of the American Library Association) develops programs, products and services to help academic and research librarians learn, innovate, and lead within the academic community. We enhance the ability of academic library and information professionals to serve the information needs of students and researchers. For example, through a one-day workshop, ACRL presenters travel to campuses across the U.S. and train liaison librarians to enhance their skills with research data management.

The movement of scientific research towards a data-intensive, collaborative approach has been well documented and discussed. The advance of digital technologies has both strengthened the power and reach of data and raised new challenges for the research community. Scientists and research institutions face many challenges while attempting to preserve the vast amounts of data for long-term use, including how best to describe data in a consistent way, how to keep up with evolving data standards, how to consistently and effectively share data while allowing for some restrictions. Additionally, there are often sociological obstacles to data sharing and data reuse, all while coping with the huge increases in the amount of data being created.

Data-driven discovery promises incredible health benefits, predicated on the understanding of the sources of the current information that is captured, identifying gaps and barriers, and improving outcomes for institutions and individual patients.

Among potential opportunities for Data-Driven discovery in health and biomedical fields we can include:

- Increased collaboration between hospitals and healthcare providers to better facilitate patient care
- Text mining and improved natural language processing to identify issues both at point of care and in research
- Improved ontologies and crosswalks between information capture systems
- Increased inter-professional education and collaboration
- Identifying and reducing clinical inefficiencies, over-testing, duplicate testing, and reduce medical errors
- Creating central repositories for secure exchange of data and identification of patients who may benefit from new treatment

These efforts come with specific challenges that must also be kept in the forefront, particularly the recognition of the limitations of current research, data sources, and algorithms with regards to under-representation across various diverse populations including gender, ethnicity, socioeconomic status, disability, sex, and immigration status. Data-driven discovery holds particular opportunities to capture information about these communities from myriad sources in order to overcome continued health disparities seen, but should be approached with the greatest respect for an individual patient's autonomy, as well as acknowledgement of the needs of healthcare professionals to make decisions to provide the best care to their patients.

2. Data-driven Health Improvement - areas such as:

No comment.

SECTION 2: Promising directions for new initiatives relating to open science and research reproducibility might address such topics as these:

3. Advanced Data Management:

In some cases, data-driven discovery and the reproducibility of research are hampered by pay walls and proprietary restrictions on data, algorithms, and analytical software. ACRL has long been committed to open access research and policy. ACRL publishes its own open access peer-reviewed journals: *C&RL*, *C&RL News*, and *RBM*, as well as all of its standards, guidelines and frameworks as open access. In June 2016, ACRL issued a "Policy Statement on Open Access to Scholarship by Academic Librarians" (<http://www.ala.org/acrl/standards/openaccess>).

In other cases, data-driven discovery is hampered by lack of HIPAA compliant methods of sharing research data that enable data sharing while also protecting patient privacy.

We would like to offer recommendations to improve accessibility and sharing of research data. Promising directions for new initiatives that would benefit from increased openness include:

- Dataset crowdsourcing, as promoted by the Center for Open Science. Dataset crowdsourcing is a method of data analysis in which multiple independent analysts investigate the same research question on the same data set in whatever manner they consider to be best. This approach may be appropriate for complex data sets in which a variety of analytic approaches could be used, and when dealing with controversial issues about which researchers and others have different priorities.
- Development of HIPAA compliant storage, as well as a means of verifying and auditing its compliance. The field needs more research on how best to develop and implement infrastructure, procedures, and policies that would facilitate collaboration and large-scale community improvement with secure data access, transfer, and reuse.
- NIH-supported infrastructure that will facilitate required and encouraged data release/sharing in a way that does not increase the burden on an individual PI or institution, and that continues to protect patient privacy. This measure would benefit the scientific process by facilitating peers verification of claims made in papers.
- Strong incentives for researchers to share research articles and data. Currently there are no incentives for researchers to share their research outputs quickly. Current structures encourage researchers to guard their data, and to publish articles in so-called “high impact” journals that take months (and in some cases, years) to approve papers for publication. We need to realign incentives for the sharing of research outputs that support the public’s interest in community health. U.S. federal agencies funding biomedical research should incentivize researchers to share data and articles widely by actively rewarding this behavior in their promotion and funding processes.

4. Intelligent Learning Systems:

No comment

SECTION 3: Promising directions for workforce development and new partnerships might address topics such as these:

5. Workforce Development and Diversity:

Academic librarians are on the forefront of teaching about, facilitating access to, and preserving information and data across extensive resources. As foundational educational partners for

many who will go into healthcare professions, we are ideal partners for workforce development in data management for health and biomedical professionals.

ACRL has created the Information Literacy Framework, an approach to the information seeking, critical assessment, and critical thinking skills that serve researchers, healthcare professionals, and patients in understanding and informed decision-making.

ACRL has developed a curriculum for academic librarians seeking to partner with researchers in both short and long-term management of research data. One major barrier to data sharing by scientists has been a lack of institutional guidance and support. Lack of formal data management processes, insufficient or nonexistent training and tools, and inadequate funding can play into the loss or misuse of research data. The following learning outcomes are based on the ACRL data management curriculum, which has successfully served many librarians in workforce development and new partnerships with a variety of researchers. It may serve the same goal in other professions.

Valuable learning outcomes include:

- Early identification of data within the research process and lifecycle in order to manage and describe data appropriately at the time of collection.
- Learning how to communicate the nuances of disciplinary requirements for data management in order to facilitate partnerships with colleagues who already have expertise in data management.
- Understanding specific existing skills already possessed by information scientists, data scientists, and librarians which transfer to data services.
- Identification of campus partners in research data management in order to create an environment of research data management support.
- Articulating the parts of a data management plan in order to describe its role as a living document within a research project.

6. New Stakeholder Partnerships:

No comment

SECTION 4: Respondents may also propose additional ideas related to health and biomedicine.

7. Additional Ideas:

No comment

SECTION 5: Information about you.

8. Name (Optional)

Mary Ellen K. Davis

9. E-mail Address (Optional)

mdavis@ala.org

10. Role (Optional - Select all that apply)

Researcher

Clinician

Administrator

Information Specialist

Developer

Other (please specify)

11. Affiliation (Optional - Select all that apply)

Academic

Business

Government

Association

Other (please specify)

On behalf of the Association of College and Research Libraries, I urge you to seriously consider these recommendations so that the NIH can best support next-generation data science challenges in health and biomedicine.

If you have any questions about these recommendations, please do not hesitate to reach out to me at mdavis@ala.org or 312-280-3248.



Mary Ellen K. Davis
ACRL Executive Director