Association of College & Resea5.ulfh1

researchers to share data. Deposited data that are not accessible due to their commercial potential should automatically be made openly accessible if the researchers cannot realize commercial viability within a reasonable time frame.

Data should regularly be exported and backed up to offline tape storage in a secure location to prevent loss during natural or man-made disasters, or in the event of large scale cyber attack.

Finally, the NIH Data Commons needs to ensure the security of highly sensitive human subjects data, including those data from small clinical trials, where an underrepresented population is put at risk due to the potential for re-identification.

Objective 1-2 | Connect NIH Data Systems

We strongly support aggregating and connecting NIH data and systems and Data Commons as described in Objective 1-2. These connections will add value to data in NIH systems. However, the implementation onnections to non-

further recommend compliance with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and Object Reuse and Exchange (OAI-ORE) specification, as well as allowing APIs to access Data Commons. These will facilitate further potential connections that the NIH and its stakeholders can not yet envision. The NIH should not be the only arbiter of which NIH data connections are possible, and allowing these levels of access in the inf

(rather thMM9O917(o(T)-f e(1)-liMM)gT1O

appropriate computing skills and data skills to researchers, whether this would be part of the curriculum for new professionals, or for continuing education and technical development for more established researchers.

Objective 3-3 | Improve Discovery and Cataloging Resources

The NIH clearly understands the importance of good indexing and a good discovery layer.

It would be useful if this section of the linked out to examples of approaches for making data findable and accessible as it does not quite clarify how the search and analysis workspaces for authenticated users improves findability or accessibility in the Data Commons pilot.

More tools for interaction might be attractive to active users, but discoverability relies on understanding use cases, and how different segments of the targeted populations will search the Data Commons, and matching that with how data producers describe the research data and how data curators normalize and refine the metadata. We agree in the importance of collaboration to this approach towards a community-driven process for identifying and implementing optimal standards to improve indexing, understandability, reuse, and citation of datasets. To this end, we suggest that the NIH fund user studies as well as applied research in information seeking behavior with the explicit goal of improving the Data Commons.

Goal 3 Evaluation

The Evaluation metric described in the Strategic Plan suggests counting new tools developed or tools adopted from other fields, but these implementation tactics should also include plans for disseminating such tools, or for developing a community to update the code for the tools.

GOAL 4 Enhance Workforce Development for Biomedical Data Science

trends and address diversity gaps in its workforce is indeed promising, but lessons that NIH can take

(<u>https://nyupress.org/books/9781479837243/</u>) indicate that quantitative analyses also risk forcing individuals into demographic categories that do not match their identities. For example, when trying to understand gender ratios of professionals in medicine and life science, it would be a mistake to rely on data that only provided research subjects with binary male/female options for gender, or even to

established researchers and program managers. The NIH Data Fellows Program also sounds like an effective method of promoting interdisciplinarity, in addition to the benefits described in 4-1.

Objective 4-2 | Expand the National Research Workforce

We strongly support the diversity enhancing efforts in data science, like the example (<u>https://www.ncbi.nlm.nih.gov/pubmed/28439180</u>) cited in the . We also support workforce development through data management education for students as well as training for established professionals. Data management services in academic libraries are growing, as are efforts to increase data management and data science skills of new professionals entering the workforce. ACRL currently offers relevant professional development through its Research Data Management Roadshow (<u>http://www.ala.org/acrl/conferences/roadshows/rdmroadshow</u>). Furthermore, librarians at Purdue, Virginia Tech, James Madison University, and University of Illinois at Chicago, to name but a few teach for-credit graduate level courses in data management and visualization. Courses such as these should be explored for inclusion in curriculum for certain professions working with life science data. Finally, there are many established graduate level courses to teach the medical community to use GenBank and Protein Data Bank. Partnering with experienced educators, trainers, and facilitators already teaching these tools, methods, and skills would be an efficient and effective means of disseminating advanced data management, analytics, and visualization tools.

Objective 4-3 | Engage a Broader Community

ACRL supports NIH's commitment to community engagement through expanded access to non-research academic organizations, community colleges, and citizen scientists.

GOAL 5 Enact Appropriate Policies to Promote Stewardship and Sustainability

Objective 5-1 | Develop Policies for a FAIR Data Ecosystem

A fairly comprehensive study by Cox et al. (<u>https://doi.org/10.1002/asi.23781</u> libraries have provided leadership in [research data management], particularly in advocacy and policy development. These areas of governance and policy development are a particularly strong area for libraries, and this

values and th

This is a dynamic area, but the Cox study cited in our response to current and projected data management activities in academic libraries.

above gives a sense of

Section 2: Opportunities for NIH to partner in achieving these goals

There is significant overlap in the Strategic Plan with the goals of the Institute of Museum and Library Services National Digital Platform (<u>https://www.imls.gov/issues/national-issues/national-digital-platform</u>). This program addresses digital capability and capacity of libraries and museums across the US It is the combination of software applications, social and technical infrastructure, and staff expertise that provide digital content, collections, and related services to users in the US Given the differences in user communities, the overlap should not be construed as redundancy, but just as the research data themselves should not be siloed, nor should the strategic plans of various federal granting agencies when they are address

strategic planning in data science for research libraries (http://d-

scholarship.pitt.edu/33891/67 0.333 s feabove gives a sense oood % E510.80nseDQ(em)-6drut p://d

independent ISO 16363 audit within four years. These performance measures will inform course corrections to ensure the reliability, commitment and readiness of NIH for long-term data preservation.

Section 5: Any other topic the respondent feels is relevant for NIH to consider in developing this strategic plan

Section 6: Respondent information

Name Mary Elen K. Davis

Email Address mdavis@ala.org

Type of Organization

Academic institution Scientific research organization Private sector Health professional Professional society Advocacy group

Patient community Government agency Member of the public Other

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 <u>mdavis@ala.org</u> or 312-280-3248.

MaryEllert