PSP and STM submission in response to NIH RFI: Strategic Plan for the National Library of Medicine, National Institutes of Health

Name *(Optional)*

John Tagler and Michael Mabe

Organization *(Optional)*

Association of American Publishers Professional and Scholarly Publishing Division and International Association of STM Publishers

Role *(Select all that apply)*

- ☐ Researcher

- ☐ Health Care Provider

- ☐ Public Health Professional

- ☐ Emergency Preparedness and Response Personnel

- ☐ Library or Information Professional (unchecked)
- ☐ Historian

- ☐ Educator

- ☑ Publisher

- ☐ Data Resource or Tool Developer

- ☐ Student
- ☐ General Public

- ☐ Other

Organization Type *(Select all that apply)*

- ☐ Academic Institution

- ☐ Library

☑ Association

☑ Non-profit Organization

☐ Government

☐ Medical Office or Clinic

☐ Hospital

☑ Publisher

☐ Commercial Organization

☐ Non-academic Research Institution

☐ Other

Please provide input on topics within any or all of the four themes below.

1) Role of NLM in advancing data science, open science, and biomedical informatics

Identify what you consider an audacious goal in this area – a challenge that may be daunting but would represent a huge leap forward were it to be achieved. Include any proposals for the steps and elements needed to reach that goal.

There are significant opportunities for NLM to work closely with ongoing efforts to advance data science, open science, and biomedical informatics in international consortia, non-governmental organizations, and the private sector. Such collaboration provides an opportunity to leverage both NLM's significant resources and connections in the stakeholder community and ongoing investments to advance these goals, reducing costs and maximizing impact.

One of the most significant opportunities for multi-stakeholder collaboration involves research data. NLM should work with international collaborative projects, like the Research Data Alliance, to develop best practices for storing primary research data and establish links to/from their uses.

The NLM may be an ideal repository for such material, given its reach within the biomedical community, especially where appropriate resources do not already exist. For example, clinicaltrials.gov could be expanded to include study protocols and fields for registering data sharing. NLM could work with AHRQ on the Systematic Review Database Repository (SRDR) by hosting the database where systematic reviewers could deposit trial-level data. More broadly, NLM could consider serving as a site for other data repositories to ensure the availability of these data for sharing, linking to resources when they are available elsewhere to prevent competition and duplication of effort. One consideration in hosting such data or objects should be making available clear copyright and license metadata elements to ensure legal use of hosted material.

Various industries and stakeholders, including publishers, are actively working to create standards and products that can, for example, support coordinated approaches and best practices regarding sharing and archiving of primary data; develop improved techniques for de-identification of individual patient data; develop options for efficient sharing of data, including standards for best-practices in data sharing; study,

explore, and begin to use new ways to evaluate the quality, reach, and impact of scientific journals and individual articles or content pieces; develop methods and standards to better view and use content (e.g., tables on small devices); and develop acceptable rules and techniques for enabling research tools within electronic health records, both for primary investigations and for connecting to related literature and data.

The most important thing NLM does in this area, from your perspective.

Research areas that are most critical for NLM to conduct or support.

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

It is our hope that NLM will learn from its experience with publications and work in partnership with the private sector and other stakeholders to more efficiently deploy scarce resources and avoid crowding out innovation in emerging technologies and services that can support data communication. Direct government competition would undermine the development of a healthy community in which to share biomedical information.

2) Role of NLM in advancing biomedical discovery and translational science

Identify what you consider an audacious goal in this area – a challenge that may be daunting but would represent a huge leap forward were it to be achieved. Include input on the barriers to and benefits of achieving the goal.

The most important thing NLM does in this area, from your perspective.

Research areas that are most critical for NLM to conduct or support.

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

NLM, as part of NIH, can best support the research community by ensuring as much of the limited funds available to NIH go to support groundbreaking research, rather than duplicative investments in infrastructure. Especially in a rapidly changing technological environment, NLM's goal should be to seek every opportunity to collaborate with other stakeholders, including publishers, devoted to the quality, integrity, and accessibility of the research literature to best capitalize on limited funds in the research community. For example, using open APIs of CHORUS, NLM could link to content on publisher websites, ensuring research is available in the context of the journal in which it is published, saving money on infrastructure, storage, and reformatting.

NLM was established to "assist the advancement of medical and related sciences and to aid the dissemination and exchange of scientific and other information important to the progress of medicine and to the public health" of the United States, and it has consistently worked to fulfill its mission. In this era of constrained budgets, activities that take it away from that mission and divert resources from its goals or those of NIH or HHS could undermine the excellent work that stakeholder communities depend upon. The

strategic plan should start from an assessment of all activities in view of NLM's core mission and each activity's value for the research community, as well as communities of practice and the general public. As one example, NLM should carefully assess whether creating international mirror or enhanced sites for PMC takes taxpayer money away from research itself. In addition, NLM should think strategically about the impact of PMC's plans to host repository sites for non-medical research. To better achieve its strategic goals in the context of constrained funding and technological developments throughout the research space, NLM may want to study, explore, and disseminate ways of using new technologies to accomplish the goals of repositories, including virtual repositories such as CHORUS.

3) Role of NLM in supporting the health of nation: clinical systems, public health systems and services, personal health

Identify what you consider an audacious goal in this area – a challenge that may be daunting but would represent a huge leap forward were it to be achieved. Include input on the barriers to and benefits of achieving the goal.

The most important thing NLM does in this area, from your perspective.

Research areas that are most critical for NLM to conduct or support.

Healthcare systems and public health arenas in which NLM participation is most critical.

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

> NLM is an excellent resource for the healthcare community, and it can do even more to integrate itself as a part of the larger community that develops and delivers biomedical and public-health information to various audiences. This larger community is international, multi-disciplinary, and includes private and public entities. Publishers stand ready to collaborate with NLM and its service components to find solutions to enable health professionals to integrate data and knowledge from biomedical research into effective practice. Such collaboration would also help ensure the best possible use and development of limited funds available to NIH, and ensure the best value for health professionals.

> For example, the Emergency Access Initiative (EAI) is a partnership of the NLM, the National Network of Libraries of Medicine, and the Professional/Scholarly Publishing Division of the Association of American Publishers and other publishers to provide free access to full-text information to health care professionals and libraries affected by disasters. This collaborative effort between NLM and publishers is a valuable service to the health care profession and the public, and an exemplar of what can be achieved through public-private partnerships. The strategic plan should include a vision for encouraging and expanding such collaborations.

4) Role of NLM in building collections to support discovery and health in the 21st century

Identify what you consider an audacious goal in this area – a challenge that may be daunting but would represent a huge leap forward were it to be achieved. Include input on the barriers to and benefits of achieving the goal.

> NLM's tools provide many services to the research community and NLM should seek to make them even more valuable to the stakeholder community. Additional functionality could be developed by NLM itself,

or by improving transparency or APIs for others to build new tools that could be of use to the research community (or public). For example, APIs could enable researchers to filter articles by institution(s), funding source(s), city, state, and country. More transparency could also help researchers better understand advancements and active areas of research in their field, for example, by consistently sharing usage statistics with publishers so that they can provide complete information about which articles are being read and where. To support the sustainability of scholarly communication, NLM should consistently share COUNTER-compliant distributed usage statistics for manuscripts it hosts to enable improved altmetrics and support a better understanding of the usage of the research literature for publishers, their customers, and the broader public. It appears the PMC provides such data to some publishers but not to others, and NLM should ensure practices are consistently applied without bias.

NLM should seek opportunities to advance its goals through collaboration and leveraging existing resources in order to save money and increase efficiency. One such opportunity for advancing public access goals is via the CHORUS service (www.chorusaccess.org), which uses existing infrastructure, tools, and services to enable sharing, access, discoverability, reporting, and preservation. As an additional benefit, such solutions could reduce the administrative burden on researchers. Current requirements for authors to deposit to and check work on PMC, rather than establishing links to the publisher's sites, creates extra work for researchers and adds costs to taxpayers for converting documents in order to make them available through PMC. Such costs and burdens could be avoided through collaborative approaches.

Various industries and stakeholders, including publishers, are actively working to, among other things: explore new strategies for discovery and personalization of searching; including Google-like algorithms for weighting and filtering search results; developing new ways for searching scientific content other than text – with search techniques for figures, images, databases, videos, etc.; and improving discovery of related content (e.g., beginning with the information on one patient, should explore techniques for finding similar cases and relevant data)

Rather than pursuing any such functionality on its own, NLM should work with other stakeholders to minimize costs to the government and support innovation in this space. While these are technologies that may be able to improve NLM's services, direct competition would undermine the development of a healthy innovation ecosystem.

The most important thing NLM does in this area, from your perspective.

Publishers recognize the value that NLM's discovery services like MEDLINE provide to health and life sciences researchers in in these disciplines. The MEDLINE content indexing services, especially the assignment of MeSH terms and making these terms available to other search and discovery services, enhances the discoverability and dissemination of the scholarly literature.

However, NLM could improve its approach to the approval of journals for MEDLINE indexing, by making the process more consistent, transparent, and engaging in two-way communication with publishers. According to NLM: "… the selection is highly dependent on the judgment of Committee members and the Director" (*http://www.nlm.nih.gov/pubs/factsheets/jsel.html*). MEDLINE has declined to consider journals whose subject areas it deems not relevant, even when these journals publish a significant number of articles reporting on NIH funded research and whose authors are required to deposit their accepted manuscripts to PMC. MEDLINE journal indexing should be more aligned with the scope of NIH funding in the biomedical arena. A journal should be included when that journal meets a fair and consistent approval process and agrees to the MEDLINE terms (http://www.nlm.nih.gov/bsd/policy/ejournals.html).

A similar but distinct issue exists with respect to PubMed. PubMed citations are limited to the following: 1) MEDLINE indexed journals, 2) journals/manuscripts deposited in PMC, and 3) NCBI Bookshelf (http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html). Because of the restricted nature of the

source of citations, some journals that publish a significant number of articles reporting on NIH funded research, whose authors are required to deposit their accepted manuscripts to PMC, are not indexed in PubMed. This creates an anti-competitive advantage for journals hosted on PMC. It also creates the potential for confusion by users of NLM resources, who may not be aware that not all materials available through the NLM are subject to the same standards. This practice should change.

Finally, some publishers and journals seem to have a "fast-track" to MEDLINE indexing, as was the case for *eLife*, while others seem to languish. For the integrity of the system, it is critical that NLM develops a more frequent, transparent, and consistent approach to approval of journals for MEDLINE indexing.

Overall, NLM should consider whether their goal is to provide indexing of quality content or whether being a repository to all content is the primary goal, and align its criterion for indexing accordingly. Such an analysis should also be performed in the context of the efforts to support public access policies of other agencies, which take NLM away from its mission of focusing on biomedical research.

Research areas that are most critical for NLM to conduct or support.

New data types or data collections anticipated over the next 10 years.

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

We continue to be concerned about the NLM activities that are duplicative or in competition with those provided by other elements of the research ecosystem, including scholarly societies and academic publishers. There are opportunities for NLM to work more collaboratively with publishers to ensure high-quality peer-reviewed medical and related scientific articles are widely available and accessible, while improving the operation of the scholarly communication ecosystem. NLM's goals, in addition to supporting the wide availability of material for the research community, should include minimizing administrative burdens on researchers, avoiding the fragmentation of the scholarly record and identifiers thereof, and avoiding unnecessary investment in infrastructure that can be more cost-effectively provided through other means.

Current practices require, according to some reports, 23 steps and several emails for researchers to submit articles to PMC under the NIH public health mandate. Even where publishers submit to PMC on behalf of the researchers, a byzantine process is involved for authors that they sometimes confuse with the publishing process itself. If NLM were better integrated with existing systems, this burden on researchers could be reduced or eliminated. There are collaborative solutions developed by publishers and other stakeholders, which identify funders (FundRef), disambiguate researchers (ORCID), and can provide access to the full text of articles (CHORUS). NLM should work to integrate itself more into the scholarly communication community to capitalize on opportunities to minimize burdensome administrative requirements on researchers, who should be spending time on research rather than administration.

These issues also isolate content in NLM and fragment the research literature, making it more difficult for researchers to ensure they are finding the authoritative version of content. For example, PMC's practice of creating and requiring PMCIDs, rather than using DOIs and other scholarly communication standards, undermines the utility of NLM's content. For example, using standard DOIs would better connect researchers to particular articles and capitalize on CrossRef interlinking. Using ORCID would help connect

researchers to potential collaborators and ensure the fidelity of authorship. To better serve the research community, NLM should work to integrate with widely accepted and used tools, or standards promoted by NISO, CrossRef, and other organizations, rather than creating proprietary formats and indicators.

NLM should avoid "mission-creep" that reduces opportunities for innovation in the private sector. The creation of derivative works and the provision of enhanced services that undermine the value of copyright and duplicate work of publishers (e.g., e.g. the image database, PubMed Commons) go beyond NLM's mission and crowd out other investments. They also divert Federal funds that could otherwise support research. In as much as any of these activities are continued, they should be broadly assessed for their compliance with statute that required NIH to implement its public access policy "in a manner consistent with copyright law" (Omnibus Appropriations Act, 2009, Division F Section 217 of PL 111-8). Future activities that would create enhanced or derivative works from the copyrighted content that PMC hosts should only be undertaken after close consultation with, and approval by, legitimate rightsholders. All such activities, including the current NIH public access policy, should also undergo regular assessment for cost-effectiveness and impact on non-governmental actors.