Comment 1

Current NLM elements that are of the most, or least, value to the research community (including biomedical, clinical, behavioral, health services, public health, and historical researchers) and future capabilities that will be needed to support evolving scientific and technological activities and needs.

Publishers recognize the value that NLM provides to health and life sciences researchers in providing discovery services 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.

At the same time, 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 are required for researchers to submit articles to PMC under the NIH public health mandate. 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's tools provide many services to the research community, but could be expanded. 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.

Finally, NLM, as part of NIH, should 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 should 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, it should avoid activities that take it away from that mission and divert resources from its goals or those of NIH or HHS. All activities should be assessed in view of the core mission and their value for the research community, as well as communities of practice and the general public. In particular, NLM should carefully assess whether creating international mirror or enhanced sites for PMC takes taxpayer money away from research itself. In addition, NLM should consider whether PMC's mission should include hosting repository sites for non-medical research, as has been suggested in some agency public access plans, Instead, NLM should study, explore, and disseminate ways of using new technologies to accomplish the goals of repositories, including virtual repositories such as CHORUS.

Comment 2

Current NLM elements that are of the most, or least, value to health professionals (e.g., those working in health care, emergency response, toxicology, environmental health, and public health) and future capabilities that will be needed to enable health professionals to integrate data and knowledge from biomedical research into effective practice.

The comments in Comment 1 are relevant to the health professional community, and should be referred to for additional comments about NLM's value to researchers and the public.

To ensure the best possible use and development of limited funds available to NIH, and to ensure the best value for health professionals, NLM should 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.

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 an example of a valuable service to the health care profession and the public that should be continued, and an exemplar of what can be achieved through public-private partnerships.

Comment 3

Current NLM elements that are of most, or least, value to patients and the public (including students, teachers, and the media) and future capabilities that will be needed to ensure a trusted source for rapid dissemination of health knowledge into the public domain.

The comments in Comments 1 and 2 also included comments on NLM elements relevant to patients and the public, and should be referred to for additional comments about NLM's value to patients and the public.

To ensure the best possible use and development of limited funds available to NIH, and to make sure as much information is available to the public as possible, NLM should 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 ensure a trusted source for rapid dissemination of health knowledge into the public domain.

Comment 4

Current NLM elements that are of most, or least, value to other libraries, publishers, organizations, companies, and individuals who use NLM data, software tools, and systems in developing and providing value-added or complementary services and products and future capabilities that would facilitate the development of products and services that make use of NLM resources.

NLM and publishers share many of the goals articulated in NLM's mission, particularly those related to the sharing and communication of the latest biomedical discoveries and data, ensuring the widest possible dissemination of knowledge, and leading the development and adoption of new information technologies, all in support of the needs of researchers, practitioners, and the public who want and need the latest biomedical information.

In the past, publishers and the NLM have sometimes been partners in advancing this mission. Members of the Association of American Publishers' Professional and Scholarly Publishing division have identified synergies in the NLM's services with their own and have pursued collaboration with NLM across a range of areas. At the same time, publishers have been frustrated by NLM's approach to managing PMC. Specifically, NLM has imposed policies that are counter to the best interests of scholarly communication and sought to expand into areas where the private sector is already developing value-added or complementary services or products. Most troubling are activities that conflict with the spirit of copyright, including the practice of creating enhanced or derivative versions of published works that compete directly with publisher content. As NLM seeks to develop and improve services that are of benefit to the broader community, it would be beneficial for NLM to broadly engage stakeholders and to pursue collaborative over unilateral solutions where possible. The NLM publisher panel is one possible venue for such collaboration with publishers, but more opportunities need to be provided for genuine two-way conversation regarding outstanding issues.

Broadly speaking, our recommendations fall into two categories: (1) develop ways to improve existing services so that they are as useful and beneficial as possible and (2) avoid investing in duplicative systems and services in competition with non-governmental organizations. The latter not only ensures cost-effective approaches, but also is critical to support NLM's goal to support the development of value-added or complementary services and products to those that they provide.

We also highlight several areas in which NLM practices could be improved by clarifying existing guidance and rules and improving two-way communication with stakeholders. Examples include the provision of usage statistics and preferential indexing in MEDLINE. We urge the NIH Director (ACD) Working Group on the National Library of Medicine (NLM) to review practices to ensure that they are as consistent and transparent as possible.

Below are a few areas where NLM might be able to apply these twin areas of recommendation:

NLM's API:

NLM's mission is to aid the dissemination of information. Improving discovery, use, and tracking of content in NLM would increase the content's utility. In addition to implementing COUNTER standards as recommended earlier, NLM should develop and offer comprehensive API's for PMC and other databases which would include information about the usage of articles and other information. Such efforts would not only improve the utility and

understanding of NLM's content, but would also improve the transparency and accountability of the taxpayer investment in NLM and its services.

In addition, NLM should use industry standard identifiers to content such as the DOI. The entire community benefits from analytics about the information in the library, and having well-defined DOIs is critical to such analysis as well as designing new products to improve the discovery, understanding, and use of the literature and data. If proprietary identifiers remain in use, well-defined and assigned PMIDs and PMCIDs, together with the PMID-DOI converter and API, must continue to be supported.

The indexing process in MEDLINE and PubMed:

MEDLINE is an invaluable service to libraries and publishers, as well as to researchers and the general public. NLM's journal selection in MEDLINE has served an important role in identifying journals that produce quality content and uphold standards of editorial work. The value of the MEDLINE selection and approval is a service to authors and readers that should be enhanced and promoted, especially with the emergence of many predatory journals.

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 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. 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.

Supporting a sustainable system of peer-reviewed biomedical articles:

NLM and PMC contain a large collection of publisher peer-reviewed articles, all of which has been published by organizations outside the government who make considerable investments in the quality, integrity, dissemination, and preservation of the scientific record. Although NIH has presented PMC as a public-private partnership, collaboration on the development and operation of PMC has been minimal. There are many opportunities where NLM and publishers could work together on solutions that advance the goals of providing access without undermining the important value that the journals that publish the material, and the publishers that invest in those journals, provide to the community.

As one example, content on NLM unnecessarily competes with authoritative versions on publisher platforms. While publishers appreciate that, in most cases, links are available from content on PMC to a journal version, these links are often hidden and the PMC version has increasingly been enhanced to make it look like a published article.

This requires significant and unnecessary investment by NLM in reformatting and enhancing manuscripts when articles are already available on journal sites. We recommend that the link to the final article on the publisher's site should be improved and made more prominent and the copyright holder and terms of use for content should be clearly identified. This will ensure readers can find authoritative versions with corrections, retractions, and related information. Similarly, references in content on PMC only link to other articles within PMC. Use of the DOIs already supplied to PMC would allow NLM to link all the references in hosted content and create a richer experience for users. NLM should therefore work to eliminate such self-referential linking to direct within PMC and point instead to authoritative content on publisher site

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, which is yet another area where NLM should ensure practices are consistently applied without bias. In addition, PMC should take concrete steps to prevent commercial re-use of manuscripts archived in PMC that is not authorized by the copyright holders of these works.

NLM should also 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 should also undergo regular assessment for cost-effectiveness and impact on non-governmental actors.

Finally, although the NIH's public access policy has been in place for approximately five years, it has never undergone a thorough and unbiased review to determine its impact on innovation and advancement. It would be worthwhile for NLM to perform such an assessment, including the costs and impact, both positive and negative, on the scholarly communication system. In doing so, it should also consider the requirement in the February 2013 OSTP memo that public access policies include a petition mechanism for adjustments to policies, including the length of exclusivity allowed to journals to recoup the costs of investment in peer-reviewed articles. Especially if NLM is considering expanding the content it covers beyond biomedicine, it is critical to recognize that a one-size-fits all 12-month embargo period is not suitable for all journals nor for all publishers or disciplines, and allow changes as necessary.

Improving transparency:

In considering how NLM can improve the cost-effectiveness and utilization of its resources, it is critical to get a complete picture of the operations, costs, and utilization of the various divisions and services that NLM provides. NLM could work on improved public repowering that includes transparency of costs; completeness of NIH funding coverage; quality control to ensure compliance of terms; and reporting of metrics to information providers and users that include individual articles, users (who and where), institutions, funding sources, authors, and their affiliations.

Reducing systemic costs through collaborative approaches:

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 a side 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:

- 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.

Comment 5

How NLM could be better positioned to help address the broader and growing challenges associated with:

- Biomedical informatics, "big data", and data science;
- Electronic health records;
- Digital publications; or
- Other emerging challenges/elements warranting special consideration.

NLM is well-positioned to contribute significantly to the challenges mentioned in the question, as well as many others, based on its role in making biomedical information available and discoverable. However, as mentioned earlier, NLM can best leverage its position by working in collaboration with all stakeholders, including publishers, and making data about its services and their usage statistics more widely available. Standards have evolved and are being established in a variety of areas and NLM should engage as a partner within this effort – sometimes leading and other times joining. For example, NLM should work with NISO and other standards organizations in a collaborative and/or consultative manner on standards for biomedical information, digital publication, data sharing, and electronic health records. NLM should also look for opportunities to adopt standards rather than developing proprietary formats for information.

Research data: As NLM looks to advance its mission to support data science in biomedical fields, it should pay close attention to the opportunities for collaboration both to contain costs and to advance its mission. It is our hope that NLM will learn from its experience with publications and avoid the kinds of competitive activities that will unnecessarily consume resources and potentially crowd out innovation in the emerging technologies and services that can support data communication. 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. For example, clinicaltrails.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. 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:

- 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
- determine standards and best practices to share and archive primary data
- 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 content (e.g., tables on small devices)
- develop acceptable techniques for enabling research tools within electronic health records, both for primary investigations and for connecting to related literature and data

NLM should work with other stakeholders to minimize costs to the government and support innovation in this space. Direct government competition would undermine the development of a healthy community in which to share biomedical information.