

**DEPARTMENT OF HEALTH AND HUMAN SERVICES
NATIONAL INSTITUTES OF HEALTH
NATIONAL LIBRARY OF MEDICINE
NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE**

Function of the PubMed Central National Advisory Committee

PubMed Central was established to support NIH's mission of disseminating the results of biomedical research widely to the public and to the scientific community. PubMed Central employs electronic publishing technology to archive, index and distribute peer-reviewed journal literature in the life sciences. The PubMed Central National Advisory Committee shall advise the Director, NIH, the Director, NLM, and the Director, NCBI, on the content and operation of the PubMed Central repository. Specifically, the Committee is charged to establish criteria to certify groups submitting materials to the system, monitoring its operation, and ensuring that PubMed Central evolves and remains responsive to the needs of researchers, publishers, librarians and the general public.

Summary of Meeting – June 10, 2014

The meeting of the PubMed Central National Advisory Committee was convened on June 10, 2014, from 9:30 a.m. to 3:00 p.m., in the Board Room of the National Library of Medicine (NLM), Bethesda, Maryland. The meeting was open to the public. Ms. Patricia Thibodeau presided as Chair.

Members Participating

Patricia Thibodeau, M.L.S., Duke University, Medical Center Library (*PMC Advisory Committee Chair*)
Martha Bedard, M.S.L.S., University of New Mexico
Sophia Colamarino, Ph.D., Stanford University Medical School
Barbara Dewey, M.A., Pennsylvania State University
Bevin Engelward, Sc.D., Ph.D., Massachusetts Institute of Technology
Lorraine Haricombe, Ph.D., University of Kansas, Watson Library
Victor Jongeneel, Ph.D., University of Illinois at Urbana-Champaign
Victor McCrary, Ph.D., Morgan State University
Randall Morse, Ph.D., Wadsworth Center, Molecular Genetics Program
Sharon Terry, M.A., Genetic Alliance
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM,
NIH (*PMC Advisory Committee Executive Secretary*)

Invited Participants/Consultants Present

Heather Joseph, M.A., Scholarly Publishing and Academic Resources Coalition (SPARC)
Philip Bourne, Ph.D., OD, NIH
Victoria Stodden, Ph.D., Columbia University

NIH Staff Present

Joyce Backus, LO, NLM
Jeff Beck, NCBI, NLM
Dennis Benson, NCBI, NLM
Devon Bourexis, NCBI, NLM

Sarah Post Calhoun, NCBI, NLM
Janet Coleman, NCBI, NLM
Jane Davenport, NCBI, NLM
Martha Fishel, LO, NLM
Marla Fogelman, NCBI, NLM
Katherine Funk, NCBI, NLM
Mary Herron, LO, NLM
Marilu Hoepfner, NCBI, NLM
Betsy Humphreys, OD, NLM
Chris Kelly, NCBI, NLM
Sergey Krasnov, NCBI, NLM
Kathy Kwan, NCBI, NLM
Fiona McNabb, NCBI, NLM
Anh Nguyen, NCBI, NLM
Rebecca Orris, NCBI, NLM
Nicole Pettenati, NCBI, NLM
Laura Randall, NCBI, NLM
Jeffrey Reznick, LO, NLM
Edwin Sequeira, NCBI, NLM
Kent Smith, NCBI, NLM
Pierce Smith, NCBI, NLM
Rebecca Stanger, LO, NLM
Neil Thakur, OD, NIH
Bart Trawick, NCBI, NLM
Frances Truong, LO, NLM
Erin Zellers, NCBI, NLM

Others Present

Gabriella Karger, Karger Publishing

I. Welcome and Introductions -- David Lipman

Ms. Thibodeau called the meeting to order at 9:30 a.m. Dr. Lipman thanked the members for their service on the Committee and voiced his appreciation for the work done by the three members who are completing, or recently completed, their terms on the Committee: Dr. Bourne, Dr. Colamarino, and Ms. Thibodeau. Dr. Lipman noted that Dr. Lorraine Haricombe will be assuming the position of Committee Chair.

The Committee members and attendees introduced themselves.

II. Approval of the June 27, 2013 Meeting Minutes

The Committee voted to approve the minutes of the June 27, 2013 meeting.

III. Report from the NLM Director's Office – Betsy Humphreys

Ms. Humphreys updated the Committee on the FY2014 NLM budget and ClinicalTrials.gov, which now has summary results for about 12,800 studies and more than 168,000 registered trials. She noted that

NLM is working on responsive design throughout the library so that its websites respond to whatever type of device (laptop, phone, etc.) is being used.

Ms. Humphreys also said she would email members with information about the videocast of a symposium held May 14, 2014, that reflected on the last 30 years at NLM, as well as information about how to submit comments regarding NLM's direction for the next 30 years. NLM also is seeking comments on the priorities and focus of the National Network of Libraries of Medicine.

IV. Initiatives Related to Federal Public Access Policies – Heather Joseph

Ms. Joseph updated the committee on the various legislative and other activities related to public access.

America COMPETES Act

There was an attempt by the House Science Committee to insert language into the FIRST (Frontiers in Innovation, Research, Science, and Technology) bill – the House version of the America COMPETES Act reauthorization – that would have extended the allowable embargo on public access articles from the current 12 months to 24 months. Ms. Joseph noted that the provision would have been detrimental to existing public access (PA) policies and any new PA policies coming out of other agencies. However, there was bipartisan support to remove the provision, and an amendment doing so was passed by unanimous voice vote during markup. The Senate bill for reauthorization of COMPETES is expected to be introduced in July.

OSTP public access directive

In February 2013, the White House's Office of Science and Technology Policy (OSTP) issued a policy memo directing Federal agencies with more than \$100 million in R&D expenditures to develop plans to make publicly available the published results from federally funded research, as well as data resulting from unclassified federally funded research. The directive affects more than 20 federal agencies and departments. The memo set an August 2013 deadline for agencies to submit draft plans.

Proposed plans for agency policies have been through one round of approval from OSTP and the Office of Management and Budget. OSTP has indicated that policies will start to be released one by one in a few weeks. Ms. Joseph noted that many of the agencies are providing plans only and have not yet drafted policy language. Some of the agencies are planning on either emulating the NIH policy or depositing manuscripts in PMC, while others are understood to be developing their own plans using initiatives such as the publishers' CHORUS (Clearinghouse for the Open Research of the United States) platform.

California public access bill

A bill providing for a state-wide PA policy is being reintroduced into the Committee on Governmental Organization. Ms. Joseph said her understanding is that the bill has been reworked to reflect NIH policy in terms of deposition requirements and embargo times, and also to limit the scope to publications in the biomedical area that result from state funding.

Discussion

Ms. Joseph noted that an outgrowth of congressional and OSTP discussions has been questions about the possibility of full open licenses to make NIH PA articles more usable, as well as usage parameters for datasets. Committee members asked whether it would be useful to have a statement from the Committee

on the issue. Ms. Joseph replied affirmatively, saying it would be helpful to have an indication of the parameters that would improve the utility of articles and information about the likely outcomes. Dr. Lipman offered to provide the Committee with information about the use of the subset of articles in PMC that has explicit Creative Commons licenses allowing unrestricted use.

V. NIH Data Policy – Philip Bourne

Dr. Bourne, who recently joined NIH as Associate Director for Data Science, outlined some of the data-related issues he believes are important to address, such as understanding how existing data is being used, new business models for handling data (e.g., public-private partnerships), opportunities for efficiencies in data resources as well as in training, and alternative methods for review of data-centric grants.

Dr. Bourne focused most of his presentation on the idea of an NIH Data Commons, which he described as being a research “sandbox,” an environment for collaboration, and “the equivalent of an extramural NCBI.” The intent of the Commons would be to facilitate data sharing in a sustainable way that would result in improved scientific discovery, knowledge, usability, quality, security/privacy, and standards.

A Data Commons would likely involve use of the cloud, but not be exclusively cloud-based. He noted that one business model of interest is that researchers would get some sort of credit for putting their research objects into the cloud.

Dr. Lipman noted that he and Dr. Bourne share the view that a practical approach would be to begin by creating a service that, even if initially minimal, is clearly useful and solves an obvious problem. For example, one could create a simple way for investigators to sign up to receive a terabyte of storage space, using a system that facilitates the community providing standards for the data.

Discussion

Discussion ensued about the training that students would need going forward and the benefits of having scientists who are biologists and are trained in areas with quantitative and theoretical frameworks, such as structural biology, evolutionary biology, population genetics and statistical genetics.

The Committee also discussed the desirability of having the data commons include a mechanism that would permit identification of content that users find particularly useful so that search results could be prioritized. Dr. Bourne commented that the community needs to drive all such aspects of the project.

VI. PMC Update – David Lipman

Dr. Lipman updated the Committee on PMC usage statistics, a proposed new policy for accepting journals into PMC, international PMC sites, and other federal agencies that might use PMC to meet the Office of Science and Technology Policy’s directive regarding public access to articles and data.

PMC usage statistics

Dr. Lipman reported that PMC currently has more than 3.1 million articles, of which more than 350,000 are NIH author manuscripts. There are more than 1,500 “full participation” PMC journals; additional journals participate through NIH portfolio agreements (where they deposit only NIH papers) and through selective deposit programs (where they deposit open access articles). PMC usage continues to

increase each year, with about 1.2 million unique users per day on peak usage days of the week. The number of articles retrieved also has continued to rapidly increase, with approximately 55 million articles currently retrieved each month. The compliance rate for NIH's Public Access policy is currently 83%.

Proposed new policy for accepting journals for PMC

Dr. Lipman described a proposed new process for reviewing journals that request inclusion in PubMed Central. The current policy is that journals need to qualify for the NLM collection and meet certain technical standards. Because of the growth in the number of journals applying to PubMed Central, there is a need to update the policy to deal with the increased numbers of journals and to maintain quality standards. After eliminating out of scope journals, approximately 350 journals per year request inclusion in PMC.

Under the new process, NLM's Library Operations Division will continue to conduct a screening to assess whether the journal is in scope and meets certain other requirements. In addition, NLM will ask independent outside consultants for input about the suitability of each journal applying for PMC. Library Operations will then make the final decision about which journals will be included in PMC.

After discussion the Committee unanimously agreed to approve the new review process for PMC journals.

PMC international sites

Dr. Lipman updated the Committee on PMC international sites. The first international site was UK PubMed Central (UKPMC), which was launched in 2007 and subsequently expanded to become Europe PMC. Dr. Lipman noted that this site has been very successful, in part because of the commitment from Wellcome Trust and the other participating research funders. Europe PMC receives all of its final published articles from the U.S. PMC archive; it also processes author manuscripts of journal articles funded by the Europe PMC sponsoring agencies and makes them available to PMC and PMC Canada, the second international PMC site.

PMC Canada became operational in 2009, and like PMC Europe, receives its final published content through U.S. PMC; similarly, it directly takes in author manuscripts funded by the Canadian Institutes of Health Research and makes them available to the U.S. and European PMC sites. Both Europe PMC and PMC Canada have their own websites and retrieval systems, and a core server coordinates content between the sites. Dr. Lipman noted that Canada has not been as active on a mandatory deposition policy as Europe.

A third planned international site is in South Korea. Dr. Lipman noted that South Korea has a law relating to mandatory deposition of government-funded papers, which should help support the success of its PMC site. The Korean Association of Medical Journal Editors (KAMJE) makes articles freely available, and PMC already has content from about 50 South Korean medical society journals.

NCBI has gotten permission from most open access publishers in PMC to provide their content to the South Korean PMC. Once the system is running successfully, NCBI will ask other PMC publishers if they are willing to have their content distributed there as well.

Other federal agencies using PMC for Public Access

Dr. Lipman reported that NCBI has been in discussion with a number of agencies following the White House Office of Science and Technology Policy's directive that U.S. government agencies with annual extramural research and development expenditures over \$100 million produce a plan to ensure that articles and data resulting from research they fund be made freely accessible online. NCBI currently has signed agreements with the Centers for Disease Control and Prevention (CDC) and the Department of Veterans Affairs (VA) for using PMC to handle their covered publications. Discussions also are underway with the Agency for Healthcare Research and Quality (AHRQ), the HHS Assistant Secretary for Preparedness and Response (ASPR), the Food and Drug Administration (FDA), and the National Institute of Standards and Technology (NIST). Dr. Lipman noted that most of the agencies that ultimately use PMC would likely have their content as part of the main corpus of articles in PMC, but that PMC would have a "storefront" to make it easy for users to limit their retrieval to articles from specific agencies.

Discussion

Topics of discussion in the Q&A following Dr. Lipman's presentation included:

- Metrics for assessing the impact on article retrieval of a shorter delay period for access to NIH-funded articles
- Balancing the desire to provide access to new and innovative journal publications while ensuring PMC journals meet a certain quality level
- The possibility of talking with the Department of Defense about having their funded research included in PMC

VI. New Wellcome Trust-NLM Digitization Project – Jeffrey Reznick & Martha Fishel

Dr. Reznick and Ms. Fishel described NLM's collaboration with the Wellcome Library and Wellcome Trust for a new project that builds on the Medical Journal Backfiles Digitization Project of 2004-2010. An MOU for the 3-year PMC "Backfiles II" scanning project was signed in April 2014, formalizing a gift to NLM of £750,000 (\$1.2 million) for the project. The new effort will involve article-level scanning of hundreds of thousands of pages of articles, with cover-to-cover capture of everything in the journals, including articles, covers, tables of content, ads, and administrative matter.

Unlike the previous project, which sourced materials from all over the country, this project will use journals from NLM's own collection. Some of the key differences from the last project include:

- The journals will not be taken apart
- The journals will be returned to NLM
- A U.S. vendor will be used (a vendor in India was used last time)
- The images likely will be captured in color instead of black and white

The journals to be digitized are primarily in the area of mental health and include *Mental Health*, *Mental Hygiene*, *Journal of Psychological Medicine*, and *Mental Pathology*. Other journals were selected for their general relevance, including *Indian Medical Gazette*, *British and Foreign Medico-Chirurgical Review*, and *Transactions of the Epidemiology Society of London*.

VII. Depositing Supplemental Data for PMC Articles – Katie Funk

Ms. Funk updated the Committee on PMC’s policy regarding deposit of supplemental data. She noted that PMC had reviewed its policy and concluded that, at least for the time being, it would continue to require that supplemental data be submitted for PMC articles. She explained that while there has been growth in repositories such as Dryad and Figshare, pointing to supplemental data in such repositories via a url is not ideal, and there is value in keeping data with the article it is supporting. There will continue to be exceptions, such as very large datasets that cannot be accommodated in PMC. Ms. Funk also noted that PMC has been looking at ways it can make it easier for publishers so that they might not have to submit to multiple repositories. Dr. Lipman added that NCBI would continue to monitor the situation as repositories and data policies evolve.

VIII. Updates from Committee Members

Committee members informed the group about various other committees, meetings and/or activities that might be of interest to the PMC Committee, including:

- An Institute of Medicine committee that will conduct a study to develop guiding principles and a framework for the responsible sharing of clinical trial data
- Data sharing policy discussions by the Global Alliance for Genomics & Health, the International Rare Diseases Research Consortium, the National Patient-Centered Clinical Research Network, and others
- Work by the National Data Service, an international federation of data providers, data aggregators, community-specific federations, publishers, and cyberinfrastructure providers that is building on data archiving and sharing efforts underway within specific communities and linking them together with a common set of tools

IX. Adjournment

Ms. Thibodeau concluded the meeting by thanking the Committee members and speakers. The meeting adjourned at 2:30 p.m.

Patricia Thibodeau, M.L.S., M.B.A. Date
Chair, PubMed Central National Advisory
Committee

David J. Lipman, M.D. Date
Director, National Center for Biotechnology
Information, NLM