

**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 17, 2011

The meeting of the PubMed Central National Advisory Committee was convened on June 17, 2011, from 9:30 a.m. to 3 p.m., in the Board Room of the National Library of Medicine (NLM), Bethesda, Maryland. The meeting was open to the public. Dr. Gary Ward presided as Chair.

Members Present

Gary Ward, Ph.D., University of Vermont (*PMC Advisory Committee Chair*)
Ivy Anderson, M.L.S., California Digital Library
Ronald Blanton, M.D., Case Western Reserve University
Philip Bourne, Ph.D., University of CA, San Diego Supercomputer Center
Sophia Colamarino, Ph.D., Stanford University Medical School
Paul Courant, Ph.D., University of Michigan Library
Jan Fassler, Ph.D., University of Iowa
Cynthia Henderson, M.L.S., Louis Stokes Health Sciences Library, Howard University
Maricel Kann, Ph.D., University of Maryland
Delores Meglio, M.S., Knovel Corporation
Mike Rossner, Ph.D., The Rockefeller University Press
Patricia Thibodeau, M.L.S., Duke University, Medical Center Library
Susan Weintraub, Ph.D., University of Texas Health Science Center
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM,
NIH (*PMC Advisory Committee Executive Secretary*)

Special Guests Present

Carl Leubsdorf, Jr., Solvitor
John Willinsky, Ph.D., Stanford University, School of Education

NIH Staff Present

Dennis Benson, NCBI, NLM
Janet Coleman, NCBI, NLM

Jane Davenport, NCBI, NLM
Kristen Greenland, NLM
Marilu Hoepfner, NCBI, NLM
Betsy Humphreys, OD, NLM
Chris Kelly, NCBI, NLM
Laura Kelly, NCBI, NLM
Andrei Kolotev, NCBI, NLM
Sergey Krasnov, NCBI, NLM
Breena Krick, NCBI, NLM
Kathy Kwan, NCBI, NLM
Chris Maloney, NCBI, NLM
John Mullican, NCBI, NLM
Carol Myers, NCBI, NLM
Jim Ostell, NCBI, NLM
Edwin Sequeira, NCBI, NLM
Kent Smith, NCBI, NLM
Neil Thakur, OER
Bart Trawick, NCBI, NLM

Visitors Present

Heather Joseph, SPARC

I. Call to Order and Opening Remarks – Dr. Gary Ward & Dr. David Lipman

Dr. Ward called the meeting to order at 9:30 a.m. Committee members and attendees introduced themselves. The committee voted to approve the minutes from the last meeting.

Dr. Lipman thanked all the members for their valuable work on the Committee and presented plaques to three members who were completing their terms as Committee members: Ms. Henderson, Dr. Ward and Dr. Weintraub.

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

Ms. Humphreys gave an update on some NLM activities that might be of interest to the Committee. She noted that ClinicalTrials.gov now has records for more than 108,000 clinical trials, and as of 2008, has results data for about 3,800 trials of FDA-regulated products (drugs, devices and biologics). Because a substantial percentage of the results data deposited in ClinicalTrials.gov is available nowhere else at the time it is deposited, it is a valuable source of information for people doing systematic reviews and evidence summaries. She noted that journal editors and others have been promoting the practice of putting the clinical trials number (NCT) in articles when published, allowing connections to be made with other data.

Ms. Humphreys discussed some recent efforts to expand awareness of NLM resources, including development of a Web page that lists and describes all NLM systems and databases that have Application Programming Interface (API) access. NLM also launched, as part of Challenge.gov, a contest to promote innovative software applications using the NLM APIs and databases.

Ms. Humphreys noted that the MedlinePlus Connect API has been well received; the application allows health organizations and health IT providers to link patient portals and electronic health record systems to MedlinePlus. MedlinePlus Connect accepts requests for information on diagnoses (problem codes), medications and laboratory tests, and returns related information from the MedlinePlus resource of health information. The service supports problems code requests based on two standards: the ICD-9-CM International Classification of Diseases), and SNOMED CT (Systematized Nomenclature of Medicine, Clinical Terms). NCBI will be supporting use of SNOMED standards for genetic conditions via an upcoming database – the Genetic Testing Registry – that will provide information on genetic tests.

Ms. Humphreys concluded her remarks by noting that the NIH and NLM budget situation may be difficult in FY2012.

III. PMC Update – Dr. Lipman

PMC Statistics

PMC currently has 2.2 million articles, 55% of which are from back issue digitization. The total includes more than 140,000 author manuscripts. There are more than 850 Full Participation PMC journals; about 270 more journals participate through NIH Portfolio agreements, and more than 1,300 journals participate through Selective Deposit (open access articles). PMC usage continues to increase each year, with about 500,000 unique users each day.

NIH Public Access Compliance

Of the estimated 230,000 articles published between July 2008 and December 2010 that resulted from NIH funding, 168,000 were deposited in PMC, for an overall compliance rate of 73%. Of those deposited articles, approximately 40% were the final published articles that were deposited directly by the publisher under a formal agreement with PMC. The remaining 60% of articles were author manuscripts deposited either directly by the author or by the publisher with the necessary follow up by the author.

Dr. Ward inquired about the compliance monitoring tool that NCBI had described at the last meeting; the tool identifies NIH-funded papers and matches them to grants, principal investigators, and institutions, and allows grantee institutions to track papers that are compliant (appropriately deposited in PMC) and non-compliant. NCBI's Ed Sequeira replied that to allow for privacy of data, NCBI was planning to control use of the tool through the NIH Commons system, but the NIH Commons had to delay the project because of budget cuts.

Figure Searching in PMC

This past year a new feature was added that provides for automatic searching of images (via image captions) when a user does a regular search in PMC. The top four images appear in a side bar – called a “portlet” – on the results page. Users can click the thumbnail image to get the full image and the parent article. Clicking “see more” provides a full list of images that match the search. A similar feature will be

added to PubMed; the search will be run against images in PMC, and the image portlet will be displayed if the images scored high enough in relevance to be deemed of potential interest to the user. Already a feature is in place in PubMed that allows users viewing an article abstract to see a strip of images from the article if it is in PMC; usage analysis shows that this feature has been popular, with many users putting their cursors over the images to get the enlarged view and/or clicking on the images.

Dr. Kann commented that it would be useful for authors to know that image captions were being searched, as it might change the way they wrote their captions.

Dr. Ward asked whether there were other ways to access the images. Dr. Lipman noted that NCBI had initially provided a separate Web site for searching images, but that some publishers raised concerns. The HHS Office of the General Counsel reviewed the issue and determined that there was not a legal problem with providing access to the images through such a Web site, but NCBI worked with the publishers to come up with a solution that was more acceptable to them.

Article-level Usage Data

In response to requests from PLoS, Elsevier and others, NCBI has begun making usage data available to publishers through a password-protected service. Publishers can get usage details, by month, for each of their articles in PMC, including publisher-supplied author manuscripts. The data are available as an Excel file or XML.

IV. Public Access Policies at Autism Speaks and Other Private Funding Agencies – Dr. Sophia Colamarino

Dr. Colamarino discussed why public access is important for non-profit research foundations. She noted that as the former VP of Research at Autism Speaks she needed access to published studies in order to strategically direct resources, and she believes that the organization should not have to pay twice (i.e., to fund the research and to gain access to the published studies). In addition, families need access to published research in order to make informed decisions and to see that progress is being made. Thirdly, the foundation needs access in order to show that the funding they receive is making a difference in progress against the disease. Dr. Colamarino noted that she also was looking for an automated way to track the publications that resulted from Autism Speaks funding.

The above issues led Autism Speaks to pursue a Public Access policy. The policy, initiated in 2008, requires Autism Speaks grantees to ensure that their manuscripts are made freely available via PMC within 12 months of publication.

Dr. Colamarino noted that while other organizations initially did not express interest in the model, there has been more interest in the past year, including contact from two non-profits that wanted to hear about Autism Speaks' experience. In addition, the Health Research Alliance (HRA) invited Dr. Colamarino to provide a webinar to their organizations, which she conducted together with Heather Joseph from the Scholarly Publishing and Academic Resources Coalition. HRA has formed a subcommittee on public access and is creating a public access tool kit that is planned for roll out at their September national meeting. In addition, Dr. Colamarino said the Rockefeller Foundation had a funder's workshop in April

2011 where funders requested a matrix of different open access policies and materials to convince their boards of the utility of such policies.

The Committee discussed how part of the incentive for a funding organization to develop a public access policy is the ability to better track funded research, and that while NCBI's new My Bibliography feature also allows for tracking grantees work, there are additional tracking-related advantages to a public access policy (e.g., being able to click through and read the paper). Dr. Ostell commented that if grantees tagged their papers with grant numbers, NCBI could set up Web pages that allowed organizations such as Autism Speaks to show all the recent publications resulting from the organization's funding. However, it was noted that display of the full text could be delayed by up to 12 months because of embargo periods on display of articles in PMC.

V. Impact of NIH Public Access Policy on Patients & Health Care Providers – Dr. John Willinsky

Dr. Willinsky described a study he conducted under a small “exploratory grant” from the National Science Foundation to investigate the public uptake of the NIH Public Access Policy. He found that there was not much awareness of the Public Access Policy, but that there was a case to be made for the value of health care providers having access. The study was conducted among fewer than 100 health care providers at a hospital and clinic in Oakland, California. The study looked at questions of where the providers were accessing information and their areas of interest.

All of the health care providers (physicians, nurses, nurse practitioners) had access to UpToDate, a paid online service that provides summaries of research with references and links to the research. The study examined whether the health care professionals needed to access the primary research papers if they had access to the summaries at the point of care. The survey found that between one-quarter and one-third of the health care providers had an interest in seeing the primary research.

The study also surveyed patients at the hospital and clinic and found that there was general interest in information if the quality could be improved. Committee members raised questions about some of the survey findings that were suggestive of patients not recognizing the difference between searching reliable sources of information, such as collections of peer-reviewed articles in PubMed, and searching more generally on the internet. Dr. Willinsky explained that the study was aimed at determining levels of interest in research, and that patient education about use of health information is a separate issue that certainly needs to be addressed.

Dr. Willinsky briefly discussed the quality of the information in services such as UpToDate and the ability to drill down and find studies of interest. Dr. Lipman said that NCBI also is interested in the issue of providing information that reflects the gold standard, such as systematic reviews. He noted that PubMed Health is starting to compile information about the most commonly searched areas where it would be useful to provide systematic reviews. The intent of PubMed Health is to provide access to systematic reviews, provide consumer summaries of those reviews, and organize the information in a way that makes it easy for users to find. Dr. Lipman commented that the systematic reviews do not obviate the need for the primary literature.

Dr. Willinsky said one option under consideration for following up his research is a controlled experiment. It would give half of a group of health care providers access to the full Stanford library

holdings and journals via an application that makes it easy to get to the materials, while the other half would get access to a package of summaries such as Ovid, UpToDate, and others. The study would look at a years' worth of behavior to see how much the health care providers use the primary research, the circumstances in which they use it, and the value they think it has provided, and compare those findings to the group using the summaries.

The Committee discussed the proposed study design and suggested various options to consider, such as narrowing the study population to a specialty such as internists or pediatricians in order to get a statistically valid result, and studying a population of health care providers that no longer has access to the literature but previously did.

VI. Annotum: Developing an XML-based Platform for Online Authoring of Journal Articles – Carl Leubsdorf

Mr. Leubsdorf described Annotum, an open-source scholarly authoring and publishing platform based on WordPress that his IT consulting company, Solvitor, is developing. Annotum attempts to address some of the limitations of Google Knol, which is used by PLoS for its PLoS Currents publications. Mr. Leubsdorf noted that he has consulted with Google and with NCBI in creating Annotum.

Goals for the first version of Annotum are to:

- Allow production peer-reviewed journals online
- Allow authors to produce content collaboratively and put it out in a structured format
- Replicate the simplified work flow of PLoS Currents
- Address some of the limitations of the Knol toolset
- Provide flexible hosting options via open source code

Key features of Annotum will include authoring tools for tables, equations, citations, figures; the ability for both public and non-public pre- and post-review comments; attractive templates for presentation in web, pdf and text; and the ability to import and export documents using XML specified by the NLM DTD. The import and export features would allow authors to create content that could be exported to journals in a structured format.

Among the reasons for using WordPress are that it is simple to self-install and can be hosted for free, and there is a rich system of plugins, themes and designs. WordPress has wide adoption: it is used by 14% of all websites, 55% of sites that use a content management system, and 20 million blogs.

Mr. Leubsdorf demonstrated the basic authoring functions of Annotum and the workflow process and answered questions from the Committee. In response to a question about data associated with a paper, Mr. Leubsdorf noted that WordPress allows pointing to data, so that for instance an Excel spread sheet could be embedded and shown as a table. Dr. Lipman commented that because Annotum will provide for output in the NLM DTD, and many journals use the NLM DTD, the process of sending an article to multiple journals is simplified. Mr. Leubsdorf and NCBI staff also discussed the advantages to controlling formatting through tagging, as the underlying structure travels with a document so that, for example, it is always clear that a certain section of text is a citation.

VII. My Bibliography & Author ID – Dr. Bart Trawick

My Bibliography

Dr. Trawick outlined the features of My NCBI, including: saving searches and setting up automatic email updates, changing default display formats, choosing filtering options, using Collections to save information found in NCBI databases, accessing one's activity at NCBI websites over the last 6 months, establishing a bibliography, and tracking compliance with the NIH Public Access Policy. My NCBI has a federated login, allowing users to sign in using their IDs from a large number of groups, such as Google, NIH login, and eRA login.

Dr. Trawick showed the Committee the new My NCBI homepage, which includes a number of enhancements, including the ability to customize the site to suit the user's needs. Among the other enhancements are new functionality for the My Bibliography feature, which allows users to track journal articles (including those not indexed in PubMed), manuscripts submitted to the NIH Manuscript Submissions (NIHMS) system, book chapters, meeting abstracts, talks and presentations, patents, and other materials.

Dr. Trawick demonstrated how My Bibliography is used to manage compliance with the Public Access Policy. Once a user's My NCBI account has been associated with their eRA Commons account (the NIH system used for grant management), the My NCBI system analyzes their citations for compliance. The system also provides a wizard for users to identify citations resulting from NIH funding and for identifying the status of a citation (e.g., whether a citation does not need to be submitted or already has been submitted by a publisher). Users can also begin submission of a manuscript to NIHMS through the site. In addition, the system analyzes a user's bibliography and can provide a view with visual indications of compliance: articles not in compliance are marked with a red dot, those in compliance with a green dot, those that are in process with a yellow dot, and those not affected by the policy with an "N/A."

Other features of My Bibliography include the ability to identify delegates who can manage the user's account and the ability to set different privacy setting for different Collections. The new interface will also show users PubMed citations that are related to their My Bibliography citations. In addition, users can share their bibliography through a URL and also can download it in a variety of formats, including CSV, plain text, and Medline format.

Author ID

Dr. Trawick discussed two approaches that NCBI plans to use in PubMed to distinguish between different authors with the same name. One approach is passive, in that it does not involve the author. In this approach NCBI uses algorithms to identify the citations from an author name that are most closely mapped to the citation being viewed. These educated guesses are based on factors such as similar co-authors and topic areas. When a user clicks on an author name to see other results from that author, the system then gives higher weight to those citations that appear to be more closely related and displays them at the top of the page. The passive system is expected to be operational in the summer of 2011.

An active approach for disambiguation of author names also is planned. In this approach NCBI will rely on authors to identify who they are and what they have published. The system will build off of My

Bibliography and will be interoperable with other author ID systems. NCBI expects to have this system operational in fall 2011.

The Committee raised the issue of researchers claiming articles that they did not actually author. Dr. Lipman said NCBI has not had any significant issues with falsification of submissions in other areas, such as sequence submissions, and doesn't anticipate problems in this area.

VIII. Researcher Profile – Dr. Lipman

Dr. Lipman noted that the work that has been done for My Bibliography and Author ID provides the foundation for a standardized researcher profile that would include the sorts of information that researchers repeatedly need to provide, such as bibliography, affiliation, and a discussion of major contributions with relevant supporting citations. If this information could be standardized, he suggested, it would make it much easier for researchers to do their reporting and apply for jobs and promotions. Committee members agreed that a set of standard specifications would be very useful and a timesaver. Dr. Lipman said that to take this project further there would need to be discussions with major funders, such as NIH, the National Science Foundation, the Department of Energy, and Howard Hughes Medical Institute, as well as with major universities, to see what information they think is necessary.

IX. Discussion – Committee Members

The Committee and other attendees discussed ideas about how to demonstrate the value of public access to funders and other constituents. Dr. Lipman noted that one approach is demonstrating incremental advantages from access to research papers rather than focusing on the potential for breakthroughs. Dr. Colamarino highlighted the idea that open access policies prepare for future possibilities. She noted her experience with a predecessor group to Autism Speaks, where parents created a tissue repository and required researchers to deposit data that resulted from their use of DNA from the repository. This open access policy resulted in a databank that became more valuable than might have been anticipated because of scientific advances in areas such as DNA analysis. Heather Joseph commented that researchers are likely to be interested in public access if it makes their work easier. Dr. Thakur commented that infrastructure improvements that make it easier to analyze the literature would drive innovation.

X. Adjournment

Dr. Ward concluded the meeting by thanking the Committee members and speakers and commending them for their energy and creativity. The meeting adjourned at 3 p.m.

CERTIFICATION

I hereby certify that the foregoing minutes are accurate and complete.

Gary Ward (Date)
Chair, PubMed Central National
Advisory Committee

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