

**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 19, 2012**

The meeting of the PubMed Central National Advisory Committee was convened on June 19, 2012, from 9:30 a.m. to 2:30 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 Present**

Patricia Thibodeau, M.L.S., Duke University, Medical Center Library (*PMC Advisory Committee Chair*)

Ivy Anderson, M.L.S., California Digital Library

Martha Bedard, M.S.L.S., University of New Mexico

Christopher Bird, Wellcome Trust

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

Jan Fassler, Ph.D., University of Iowa

Lorraine Haricombe, Ph.D., University of Kansas, Watson Library

Maricel Kann, Ph.D., University of Maryland

Delores Meglio, M.S., Knovel Corporation

Mike Rossner, Ph.D., The Rockefeller University Press

Michael Tanner, Ph.D., Association of Public and Land Grant Universities

David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM, NIH (*PMC Advisory Committee Executive Secretary*)

**Special Guests Present**

Heather Joseph, M.A., Scholarly Publishing and Academic Resources Coalition (SPARC)

**NIH Staff Present**

Dennis Benson, NCBI, NLM

Sarah Post Calhoun, NCBI, NLM Janet Coleman, NCBI, NLM

Jane Davenport, NCBI, NLM  
David Gillikin, NLM  
Marilu Hoepfner, NCBI, NLM  
Betsy Humphreys, OD, NLM  
Chris Kelly, NCBI, NLM  
Andrei Kolotev, NCBI, NLM  
Sergey Krasnov, NCBI, NLM  
Kathy Kwan, NCBI, NLM  
David Landsman, NCBI, NLM  
Chris Maloney, NCBI, NLM  
John Mullican, NCBI, NLM  
Anh Nguyen, NCBI, NLM  
Edwin Sequeira, NCBI, NLM  
Kent Smith, NCBI, NLM  
Neil Thakur, OER  
Bart Trawick, NCBI, NLM

### **I. Welcome and Introduction of New Members and Chair -- Dr. David Lipman**

Ms. Thibodeau called the meeting to order at 9:30 a.m. Committee members and attendees introduced themselves.

### **II. Approval of the June 17, 2011 Meeting Minutes**

The Committee voted to approve the minutes of the June 17, 2011 meeting. In response to a suggestion from Dr. Colamarino, the Committee requested that meeting minutes be sent to committee members soon after the meeting instead of being distributed at the next annual meeting.

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

Ms. Humphreys provided an update on several NLM activities:

#### NLM/NIH Budget

Ms. Humphreys reported that both the President's proposed budget and Senate report language call for an increase for NLM that would enable NCBI to receive the majority of its funding through the base budget instead of obtaining approximately one-third through transfers from other NIH Institutes and Centers. She noted that many in the community do not expect a decision about the NIH budget before December.

#### ClinicalTrials.gov

ClinicalTrials.gov now has records for more than 128,000 clinical trials of FDA-regulated products (drugs, devices and biologics), 6,100 of which include summary results data, Ms. Humphreys said. She noted that ClinicalTrials.gov is a novel source of clinical trial results as it is often a long time before the results are available in publications and some will never appear in publications. She also noted that various groups have been analyzing ClinicalTrials.gov to examine issues pertaining to the design of clinical trials, as well as to determine who is submitting to the database.

#### NLM Journal Article Tag Suite

Ms. Humphreys reported that the National Information Standards Organization (NISO) is reviewing the Journal Article Tag Suite for approval as a standard and is expected to act favorably. The tag suite – commonly known as the NLM Journal DTDs – has had very strong uptake, she said, with 71% of publisher content currently coming into PMC in the DTD.

#### PMC addition of American Psychiatric Association journal back issues

Ms. Humphreys recounted a recent example where PMC was able to make a journal available electronically that may have soon become unavailable. The situation arose when the American Psychiatric Association changed its platforms for electronic content and decided that it would not be converting journals that had been superseded or merged into other journals. NLM learned that 10 years (1992-2001) of the Journal of Psychotherapy Practice and Research might no longer be available and arranged to provide it via PMC.

#### Electronic Health Records interfaces

Ms. Humphreys noted that one of the proposed criteria for certified Electronic Health Record (EHR) providers is supplying a standard interface between EHRs and outside information that can be integrated for decision support. NLM already has one interface designed for use in EHR products; the interface, called Medline Plus Connect, allows health organizations and health IT providers to link patient portals and EHR systems to Medline Plus health information. Ms. Humphreys commented that we may see the information development community design a number of interface products, and there may be demand for interfaces with other NLM information resources, such as the Bookshelf.

#### **IV. eLife, a New OA Journal – Dr. Mark Patterson (via video conference)**

Dr. Patterson described eLife, a new open access journal supported by the Howard Hughes Medical Institute (HHMI), the Max Planck Society, and the Wellcome Trust. The journal, which is expected to launch around the end of 2012, is part of a joint initiative of the funders to improve research communication. Dr. Patterson cited three motivations for developing eLife: to drive more open access, to provide an improved peer review process, and to take better advantage of digital media.

To illustrate the need for more widespread open access, Dr. Patterson noted that only 10% of the content in PubMed is available in full text from PMC. Articles in eLife will be available for free immediately upon publication, and users will have the right to use the content freely, provided that full author attribution is given using the Creative Commons-Attribution license. All content also will be deposited in PMC. The costs of eLife will be totally underwritten by the three funders, at least initially, so that the journal can become established; in time, eLife may charge publication fees and build other revenue streams, Dr. Patterson said.

The scope of eLife will be broad, ranging from basic and theoretical work to translational, applied and clinical research. eLife is intended to be highly selective, with influential work that advances understanding, opens new doors, or has real-world impact, Dr. Patterson said.

eLife is a community-run project with an editorial group that includes an editor-in-chief, two deputy editors, 17 senior editors, and a board of approximately 180 external reviewing editors. The editorial process is designed to provide a speedier review and a better submitter experience. Submitters can send a PDF of their article, which will undergo a quick triage process by senior editors to determine if the article will go on to peer review. A senior editor will then work in consultation with external reviewers, and will assimilate reviewer comments to provide a single set of directions to the author. Once the revised article comes back, the reviewing editor will generally be able to make a decision alone, and very quickly. Because of the extensive amount of work involved for senior editors, they will be paid, whereas external reviewers will not, Dr. Patterson said.

Dr. Patterson highlighted the importance of eLife being a digital journal with no limits on article number or length. There will be a focus on making the format as useful as possible, with encouragement for rich media and inclusion of underlying data. The intent is to extend the traditional workflow by pushing content to more locations and in more formats. Dr. Patterson noted that eLife also wants to make content available through an API and is looking at creation of an API that is both a read and write API, which would allow capture of information about reuse of articles and other advantages. eLife is working with eJournal Press and HighWire, whose open platform has a management system that will provide a lot of flexibility for integration. Dr. Patterson noted that the infrastructure is mostly in place and that eLife expects to begin accepting submissions this month (in June).

#### *Discussion:*

##### Less selective version of eLife & business model

Dr. Lipman noted that there had been talk early on of having an eLife component that would be less selective and asked whether that was still being considered. Dr. Patterson said that would be an attractive next step for eLife and was a definite possibility, but that for now staff would be focusing on the launch of eLife. Dr. Lipman noted the effort involved in reviewing articles and asked whether eLife would work out relationships with other open access journals that might consider publishing submissions not accepted by eLife. Dr. Patterson replied that the initial triage process would be rigorous but fast and that authors would be able to go to another journal without having wasted much time. He said that eLife wants to receive the whole paper but the submission process will be simple for authors. Dr. Patterson estimated that 50% or more of papers that make it to peer review will be published by eLife. For those that are not published, eLife will ask authors if they give their permission to provide contact information to other journals.

Dr. Bourne asked about the business model for eLife given the costs involved. Dr. Patterson said that eLife does not plan to charge for an initial period, perhaps three years, and that the longer term business model will be developed over the next year or so. One possibility is to introduce a journal with a high acceptance rate. Another possibility is charging submission fees. He noted that publication fees would not be sufficient to cover costs. Dr. Lipman commented that if eLife was going to introduce a high-acceptance-rate journal, it might be best to do it in a timely way rather than waiting two or three years.

##### Public communication of science

Dr. Colamarino asked whether eLife would take a leadership role in public communication of science. Dr. Patterson responded that eLife had not identified that as a goal but noted that

every article will include a plain English summary. He also noted that one of their senior member's roles will be to think about ways to add value to the research eLife is publishing. Also, because the research is open access, others can take the content and add value to it.

#### No preferential treatment for sponsor organizations' scientists

Asked whether scientists funded by eLife's three sponsoring organizations would receive preferential treatment or have a special track for submitting papers, Dr. Patterson said "absolutely not." He added that it is essential that eLife is open to everyone and that nobody receives special treatment.

#### eLife acceptance criteria

In response to a committee member's question about the basis for rejecting an article, Dr. Patterson said that while there are some base-level criteria, it is up to the editors to determine the work's significance and there is a large amount of subjectivity. The intent is for the journal to be as inclusive as possible and not limited to advances in biology, he said.

### **V. PMC Update – Dr. Lipman**

#### PMC Statistics

PMC currently has nearly 2.5 million articles, of which less than 50% are from back issue digitization. The total includes more than 200,000 author manuscripts. There are more than 1,080 "full participation" PMC journals; about 270 more journals participate through NIH portfolio agreements, and more than 1,700 journals participate through Selective Deposit (open access articles). PMC usage continues to increase each year, with about 700,000 unique users per day on peak usage days of the week.

#### NIH Public Access Compliance

Of the estimated 336,000 articles published between July 2008 and December 2011 that resulted from NIH funding, 252,000 were deposited in PMC, for an overall compliance rate of 75%. 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. While the compliance rate has moved up significantly from the 19% seen between 2005-2007 when the policy was voluntary, Dr. Lipman reported that it has been relatively static at about 75% for some time. He noted that NIH is considering a more proactive approach, which, if adopted, could boost the compliance rate to 90% or higher.

#### Health Research Alliance (HRA)

HRA is a coalition of 48 private funding agencies, including Howard Hughes and the American Cancer Society, that funds about \$1.5 billion in grants each year. At the encouragement of Heather Joseph and Sophia Colamarino, HRA has taken steps to enable its member organizations to implement public access policies. HRA and SPARC have created policy documents modeled on the NIH Public Access Policy, and HRA expects to have a system ready in the fall. Ms. Joseph noted that HRA has developed a portal that is connected to HRA's grants management system, which enables the member organizations to monitor compliance and connect papers to grants. Ms. Joseph described the system HRA has

developed as essentially "turnkey"; there are template documents, a guide for implementing a policy, and once papers start going through the system, the PMC IDs are fed back into the HRA's grants monitoring system. Twenty HRA member groups have indicated interest in adopting a public access policy in the coming year, Dr. Lipman said. The first deposits to PMC are expected to be at least a year after the policy is announced.

#### Journal growth, impact factor, peer review and reproducibility

Demonstrating a new tool in PubMed that presents a graphic of search results broken out by year, Dr. Lipman showed how the number of articles in traditional journals has been fairly flat while open-access journals (excluding PLoS One) have had linear growth and PLoS One has had tremendous growth. Among the possible factors explaining the growth of PLoS One are: it is open access; it has a broad scope, giving it good name recognition; many funded investigators are publishing there, making it safe; and it has a high acceptance rate for funded investigators and may have a less onerous process (e.g., faster and fewer revisions). Dr. Lipman commented that PLoS One has had a very important role in pushing forward open access, and that a comparable journal by a group with high credibility would further propel growth of open-access articles.

Dr. Lipman briefly discussed the problems with metrics such as impact factor, noting, for example, that 50% of the articles account for almost 90% of the citations and 15% of articles account for 50% of citations. He presented a chart showing that the journals with the highest impact factors, which generally reject the highest percentage of papers, also happen to have the most retractions. He also showed a slide of pharmaceutical products that had serious safety problems that were not identified in published studies but only came to light because of other factors, such as whistleblowers or FDA regulators. These examples, he suggested, illustrate the limitations of peer review in determining accuracy in science.

Dr. Lipman went on to describe some of the work that has identified problems with research reproducibility, including a paper that reported Amgen was able to reproduce published study results only 20% to 25% of the time. He cited a parallel phenomenon that is occurring with cell lines: estimates are that 20-30% of cancer cell lines may be misidentified, and a study of papers published between 2000 and 2004 found nearly 1,000 citations of research with cell lines that were known since the 1960s to be contaminated. Dr. Lipman concluded by discussing the importance of dealing with these issues and the possibility that post-publication efforts might be successful. Dr. Bourne noted that post-publication processes in journals often do not work well and raised the possibility of having an annotation feature in PubMed Central. Dr. Lipman said that there has been interest in having a comment feature in PubMed and that it could be useful for helping users more readily identify important research as well as identifying research that might be problematic. Dr. Lipman is scheduled to present on this issue at an upcoming NIH Steering Committee meeting.

## **VI. The State of Public Access in the U.S. – Heather Joseph**

Ms. Joseph updated the Committee on legislative and executive branch activities related to public access.

#### OSTP public access report

Under the America COMPETES Act, Congress had directed the White House's Office of Science and Technology Policy (OSTP) to convene an interagency working group to study

the possibility of other federal agencies besides NIH implementing a public access policy. The working group was supposed to issue a report in early 2012; however, in late 2011 the group said they wanted to go back to the public for a second set of comments. The intention, Ms. Joseph said, was that the public feedback would complement the working group's efforts. Comments were solicited in late 2011 through early 2012; the report has not yet emerged.

#### Research Works Act

The Research Works Act (H.R. 3699), introduced in December 2011 by Rep. Darrell Issa (R-CA) and co-sponsored by Carolyn Maloney (D-NY), was designed to overturn NIH's public access policy and stop other federal agencies from implementing similar policies. There was a strong public outcry from a number of different communities, including those who support open internet, open science, open government, and open access. As a result, the Act was withdrawn about two months after its introduction, which Ms. Joseph said is quite rare. She added that reaction to the bill had the effect of re-energizing those in Congress who were interested in seeing public access at other federal agencies.

#### Federal Research Public Access Act of 2012 (FRPAA)

FRPAA was introduced Feb. 9, 2012, in both the Senate and the House of Representatives. The bill, which has 32 House cosponsors, requires federal agencies with extramural research budgets over \$100 million to make publically available on the internet the final manuscripts of articles resulting from the agencies' funding. The bill calls for the manuscripts to be made available no later than six months after publication in a peer-reviewed journal, and thus has a shorter embargo than existing law for the NIH Public Access Policy, which requires availability within 12 months of publication. Ms. Joseph noted that there has been a lot of interest in attaching FRPAA to legislation that might move, as it is unlikely to move on its own in an election year.

#### We the People petition

On May 20, 2012, a group of open access advocates (including Ms. Joseph, John Wilbanks, Mike Rossner and Michael Carroll) created a "We the People" petition on the Obama Administration website calling for free, timely internet access to journal articles arising from taxpayer-funded research. The petition garnered the required 25,000 signatures within two weeks. Ms. Joseph commented the profile of the issue has been raised and has reached the science policy levels of the White House. A response to the petition is required, and the expectation is that if any policy is to result there will be news within the next month, Ms. Joseph said.

#### Discussion

In response to a question about whether other federal agencies would follow the NIH model, Neil Thakur pointed out that there are several aspects to the policy discussion: whether it would be mandatory for awardees, the length of an embargo period (NIH is 12 months but FRPAA uses 6 months), and type of license (fair use, full open access, Creative Commons Attribution, etc.), and how/where the content is made available (e.g., a central repository like PMC, a repository developed with other agencies, or publishers' websites).

One committee member commented that many people in the research community may not be aware of pending legislation, and suggested that talking points and other materials for reassuring faculty could be helpful should there be pushback. Ms. Joseph noted that even if there was a directive from the White House, agencies would likely have a year or so to make

public their implementation plans.

The committee discussed how Congressional staff views the NIH policy as a success and is impressed with the heavy usage of PMC. Dr. Lipman noted that some staffers now also recognize that open access publishing is just a different business model and that it is contributing to growth of new journals and businesses. The committee also discussed the most effective testimonials, such as biotech companies testifying that business development will be stifled because they need access to the literature to do their work yet can't afford to purchase all necessary journals. Dr. Colamarino commented that funding organizations such as those in HRA need examples of discoveries that have been derived from open-access datasets to understand what can be accomplished with access to the literature. The Committee also discussed the need for definitions of "open access" and "public access" and the confusion between those two terms.

Citing examples of scientific advances made by high school students, Dr. Bourne commented that this population is becoming more advanced and needs access to the literature. He added that the need for access among high school students fits in well with the U.S.'s interest in science education.

## **VII. Changes in Advisory Committee Membership & Updates from Members**

Dr. Lipman thanked all the members for their valuable work on the Committee and presented plaques to four members who were completing their terms: Mr. Bird, Dr. Blanton, Dr. Kann and Dr. Tanner. Committee members then had the opportunity to provide updates on relevant matters or discuss issues of interest.

### PubMed searching

Dr. Blanton commented that it is sometimes difficult to search PubMed, and, for example, find an older article. He said that sometimes one has to go between Google and PubMed to find an article. Dr. Lipman said that the new timeline feature, which allows users to see subsets of articles from different time periods by mousing over the timeline, should help with finding older articles. He noted that NCBI also will be introducing search tools for older articles that are weighted for relevance and click-throughs. Dr. Lipman explained that it is difficult to make a default search that works well for everybody, and that users often do not take advantage of the ability to modify the default search. He noted that the number of users taking advantage of PubMed search limits has doubled as a result of some redesigned elements, but that while we are seeing improvement, more needs to be done. Ms. Humphreys added that PubMed may not have older articles for some journals.

### Border around journal logos

Dr. Rossner asked whether it would be possible to remove the border that now appears around journal names/logos in PMC. Dr. Krasnov replied affirmatively.

### Alternatives to impact factor

Dr. Rossner asked whether NCBI is looking at any alternatives to impact factor, noting that it is a common request he gets from academic editors. Dr. Lipman said NCBI has been studying areas such as click-throughs to see if retrieval can be improved. He suggested discussing the issue further in a few months after NCBI completes more exploratory work.

### Database mining

Dr. Bourne related his experience at a recent bioinformatics conference where people from a range of disciplines talked about the possibility of a tool that would mine databases and literature overnight and pull out those items that were most relevant to the work one had done the day before. Dr. Lipman commented that it is an exciting area but also very difficult to get right, in part because of the level of noise in the literature and in part because of the need for the system to return the proper yield of highly relevant items.

### Public Access compliance

In response to a question from Mr. Bird about compliance with the Public Access policy, Dr. Thakur said that NIH is working towards a new electronic format for annual reports that is expected to help with compliance. Scientists will use their electronic grants administration accounts and link to My NCBI for their grant reporting. With the current model NIH relies on grantees to tell them what they published as a result of their award and whether they are in compliance. The structured data of the new system will allow NIH to more definitively identify what is out of compliance and to inform grantees. The new system will also allow easier follow up. In addition, NIH is considering a policy regarding delay of funding until grantees come into compliance. If that policy change is adopted, an announcement would be made many months in advance, Dr. Thakur said.

### Licenses

Dr. Bourne noted the lack of standardized license information in PMC, making it difficult, for example, to identify the subset of open-access literature that could be used for commercial purposes. Discussion ensued about how publishers use different licenses and that it would be difficult for PMC to codify until the publishing community agrees on standard licenses. Dr. Thakur commented that from a policy perspective there may be an opportunity for government to specify the licenses that are acceptable for papers subject to public access policies.

### Data sharing

In response to a question from Ms. Anderson about what PMC is doing to support data and data sharing, Dr. Lipman explained that NCBI has three main mechanisms: the standard NCBI databases (GenBank etc.), which are the venue in many cases for authors to submit data related to their papers; PMC, which allows for authors to upload supplementary data; and a new submission portal that will be rolled out over the coming year. The new submission portal will allow submitters to integrate components from complex projects where there might be different types of data (e.g., expression data and sequence data), and submit them together as part of “bioproject,” which can also be linked to the literature. The new portal will also have a “bit bucket” for unstructured data that does not go into other NCBI databases, such as images. Over time, the “bit bucket” will allow metadata to be associated with the submission, and as it evolves, community standards and schema could be specified.

## **VIII. Recent Developments in the U.K. and E.U. – Chris Bird**

Mr. Bird provided an update on open-access developments over the last year in the U.K. and E.U., noting that it has been a busy year with a groundswell of opinion shifting in favor of open access.

## Finch Report

Formally titled “Accessibility, sustainability, excellence: how to expand access to research publications,” the Finch Report was issued by the Working Group on Expanding Access to Published Research Findings on June 18. The working group had been established by Science Minister David Willetts in October 2011 to examine how published outputs of U.K. research could be made more accessible. The key message of the report, Mr. Bird said, is that the U.K. should initiate clear policy direction in favor of gold open access, but that a mixed model will be needed during a lengthy transitional period. Mr. Bird provided further details on the key recommendations of the report, including two controversial elements: moving away from restricted re-use of gold open access content among traditional publishers to use of Creative Commons CC-BY licenses, which permit copying, redistribution, adaptation and commercial uses, and a 12-month embargo period for green open access.

The report is the first in a series of developments expected over the coming weeks. On June 21, the Royal Society will publish its report on Science as an Open Enterprise, and the government is due to publish by the end of June a white paper on the theme of open data. Research Councils U.K. is expected to soon announce its strengthened policy on open access, and the European Commission is expected to publish a statement supporting open access in July, he said.

## Wellcome Trust open-access compliance & policy changes

Mr. Bird presented a slide showing that compliance with the Wellcome open-access mandate has gone from about 20% in 2007 to about 55% in March 2011. In an effort to improve compliance Wellcome will be instituting two sanctions: requiring institutional assurance that all publications associated with a grant are in compliance with the policy before final payment on a grant is made, and ensuring that researchers comply with the open-access mandate for previous and active awards before they can receive renewals or any new awards. In addition, Wellcome will require that where the Trust pays an open-access fee, the licensing terms allow unrestricted re-use, including “commercial” use (CC-BY license); Mr. Bird noted that this policy is in line with the new policy of the U.K. Research Councils.

## Europe PubMed Central

In July 2012, Wellcome plans to announce that U.K. PubMed Central will become Europe PubMed Central. Mr. Bird noted that “significant” European funders would be joining soon.

## Discussion

The committee discussed evidence that six-month embargo periods do not appear to adversely affect journals. Dr. Lipman observed that different approaches will help some publishers and hurt others, but that the key issue may be funding organizations saying what they want.

## **IX. Reuse and Redistribution of PMC Content – Ed Sequeira**

Mr. Sequeira provided an overview of facts about PMC that are relevant to reuse and redistribution of content.

Mr. Sequeira explained that while everything in PMC is free, less than 20% of the articles (465,000 out of 2.45 million) have open-access licenses. However, a growing percentage of

