

**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 Minutes of Meeting – April 19, 2007**

The meeting of the PubMed Central National Advisory Committee was convened on April 19, 2007 in the Board Room of the National Library of Medicine (NLM), Bethesda, Maryland. The meeting was open to the public from 9:30 a.m. to 2:10 p.m. Mr. Robert Kiley presided as Chair.

**Members Present**

Prue Adler, M.S., Association of Research Libraries  
Camila Alire, Ed.D., University of New Mexico  
Shirley Baker, M.A., Washington University  
John Hawley, B.A., American Society for Clinical Investigation  
Robert Kiley, M.Sc., Wellcome Trust  
Sarah Michalak, M.L.S., University of North Carolina, Chapel Hill  
Hemai Parthasarathy, Ph.D., Public Library of Science  
Mary Ryan, MLS, University of Arkansas Medical Sciences  
Mark Sobel, M.D., Ph.D., American Society for Investigative Pathology  
Johannes Velterop, Ph.D., Springer Publishing  
Gary Ward, Ph.D., University of Vermont  
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM,  
NIH, and PubMed Central National Advisory Committee Executive Secretary

**NLM Staff Present**

Jeff Beck, NCBI  
Abraham Becker, NCBI  
Dennis Benson, NCBI  
Janet Coleman, NCBI  
Jane Davenport, NCBI  
Mark Desierto, NCBI

Martha Fishel, NLM  
Marla Fogelman, NCBI  
Al Graeff, NCBI  
Chris Kelly, NCBI  
Andrei Kolotev, NCBI  
Sergey Krasnov, NCBI  
David Landsman, NCBI  
Dawn Lipshultz, NCBI  
Becky Lyon, NLM  
Adeline Manohar, NCBI  
Carol Myers, NCBI  
Jim Ostell, NCBI  
Edwin Sequeira, NCBI  
Jerry Sheehan, NLM  
Kent Smith, NCBI  
Bart Trawick, NCBI  
Kerry Zbicz, NCBI

### **Visitors Present**

Josephine Briggs, Howard Hughes Medical Center  
Laura Brockway, FASEB  
Heather Johnson, SPARC  
Carl Rhodes, Howard Hughes Medical Center  
Erich Staib, Duke University Press  
Crispin Taylor, American Society of Plant Biologists  
Neil Thakur, Office of Director, NIH,  
Nancy Winchester, ASPB

### **I. Call to Order and Opening Remarks**

The meeting was called to order at 9:35 a.m. Dr. Lipman welcomed members of the PubMed Central National Advisory Committee and introductions of members and visitors were made. He made note that neither Dr. Lindberg, Director, nor Ms. Betsy Humphreys, Deputy Director, NLM, were able to attend the meeting.

The new Chair, Mr. Robert Kiley, was introduced. He provided an explanation of the Wellcome Trust and its role in public access and said that PMC plays an important role in meeting its objectives. Minutes from the October 2006 meeting were approved.

### **II. Remarks by NLM**

Mr. Jerry Sheehan, NLM, provided an update on legislation affecting NLM. He reported that the Congressional Continuing Resolution for the current fiscal year contained a specific increase for NCBI. Dr. Zerhouni, Director, NIH, testified before House and Senate Appropriations committees and fielded questions about public access policy and continuing developments in this area. As of April there were no new legislative developments with regard to public access policy.

### **III. PMC Update**

Dr. Lipman reported that PMC currently has about 960,000 articles, two-thirds of which are from the back-issue digitization project. Total items in PMC are over 1 million. Total retrieval for March 2007 is up to 11 million articles and total page views are about 15 million. PMC is seeing constant growth of usage in correlation with size of the database.

PMC International has test sites running in China, Japan, South Africa, Italy and soon Canada. UKPMC and UK manuscript submissions system went live the first week in January. The Korean Association of Medical Journals has shown interest and is pursuing the setup of a site there.

Dr. Parthasarathy asked about the challenges in the back-issue digitization project. Some challenges include identifying content, filling in the time line of missing journals or articles, and quality control. Ms. Martha Fishel reported that developing specifications for citations is especially difficult because a style guide is specific for each journal. In addition, staff must work closely with the contractor for quality assurance. Dr. Alaire asked the percentage of unaccepted citations from the contractor. Ms. Fishel answered that about 25-30% are rejected on the first pass; PMC's acceptance criteria call for better than 99% accuracy.

Ms. Michalak asked if NCBI knows the breakdown of PMC users. Dr. Lipman replied that users of the NCBI site consist of about one-third health professionals and one-third academic with a core group of frequent users. Mr. Kiley asked about the percentage of open access articles in PMC. Dr. Lipman reported that there is a subset of about 50,000 open access articles. In response to another question regarding an analysis of usage by US regions, he said that NCBI tries to avoid such analyses because of potential user identification issues. The NLM/NIH privacy policy was questioned and it was asked if the public could access weblogs. Such information is not public. NIH has denied requests for weblogs containing IP addresses in the past. Dr. Ward asked if the origin of users is known. Dr. Lipman answered that many are from Google and PubMed with a significant number searching PMC directly.

Mr. Kiley asked Dr. Lipman to provide a quick update on the NIH Public Access Policy. He reported that direct submission by authors has remained steady, at about 5%. Elsevier has begun to provide all NIH-funded papers via bulk upload. In this case, authors must still get involved, by verifying their respective articles via the NIH manuscript submission system. Only about 45% of them do so and take the process to completion. Dr. Ward asked how the Elsevier setup differs from affiliate journals. Dr. Lipman explained that there is no formal relationship between PMC and those publishers who send information via bulk upload, and that these bulk uploads contain word processing documents or similar forms of a manuscript, rather than the final edited and tagged articles. Therefore communication with the author is needed to verify and approve the release of every article.

*Break 10:40-10:55*

### **IV. PubLink Policy Review**

Dr. Lipman began the PubLink Policy discussion with a review and history of the policy. The PubLink option was available to publishers who wanted to deposit their articles in PMC, but

have the article viewable only via a link to the publisher site. The option did not, however, bring in many more participants as originally thought. So, in 2004, on the advice of the Advisory Committee, PMC stopped offering the PubLink option for new participants. Most publishers who came into PMC using the PubLink option have since switched to the full participation model, making content directly viewable on PMC. At this time, there are four journals still using the PubLink option, with no plans to switch to full PMC participation. The discussion today is to determine whether PMC should continue to accept permanent PubLink deposits from current PMC participants. This is not a discussion of the situation where a journal applies a longer embargo to its non-research content than to its research articles, and the PubLink approach is used temporarily for the material with the longer embargo.

Dr. Lipman explained that having articles viewable in PMC is more likely to uncover problems that might exist in the archival record. If content is not being used, there is no way to know if there are problems. Also, the cost factor of indexing these articles is an issue. They cost the same to process as any others in PMC, even though the PubLink articles are not viewable on the PMC site.

The committee discussed issues such as quality control, benefit of having the four journals in PMC, participant commitment, and cost to PMC. A grace period was agreed upon with the recommendation that the PubLink option cease.

Ms. Ryan made a motion to discontinue the permanent PubLink option effective July 1, 2008 for archiving new content from all publishers. It was seconded by Dr. Alire. The motion was carried by the committee.

## **V. UKPMC Update**

Mr. Kiley provided an update on the UKPMC project, which supports the open access policies of the Wellcome Trust and other members of the UKPMC Funders Group. UKPMC mirrors PMC and is part of a network of PMC International (PMCI) repositories. A manuscript submission system is also available to enable researchers to deposit their peer-reviewed research papers.

The UKPMC Funders Group together accounts for over 90% of the UK's biomedical research funding. All funders have policies mandating a maximum six month embargo to make articles available via PMC/UKPMC.

UKPMC went live on January 7, 2007 and contains over 700,000 peer reviewed journal articles. Mr. Kiley reported that the database is growing rapidly with over 38,000 new articles since January. Not all PMC content is mirrored, because a number of journals have thus far not permitted their content to be viewed in UKPMC. The manuscript submission system contains entries for 15,700 grant holders from the eight funders. Each grant holder has a login to the system to deposit papers. 485 manuscripts from 198 unique grantees were deposited in the first three months.

Dr. Parthasarathy asked about the mirroring system. When an article is uploaded into the UK manuscript submission system, it is tagged, then sent to US PMC and from there mirrored to UKPMC. It was asked whether an article that is submitted via the NIH public access policy will

be mirrored in UKPMC. Mr. Kiley replied that it will not because the copyright is still held by the publisher and NIH has not been given the right to send these articles to PMCi repositories.

UKPMC development plans include expansion of the system to meet the needs of the UK's biomedical research community. An advisory board has been formed consisting of leading researchers, academics, and representatives from industry.

The first phase of UKPMC research and development includes search and discovery, where local search capabilities will be developed and improved. Contextual linkage to datasets will provide the researcher with live links to the various European Bioinformatics Institute datasets from UKPMC articles. Enhanced reporting tools will enable UKPMC funders to better monitor outputs from grants in order to improve targeting of research spending. An example of the first phase of R&D was shown including the UKPMC Biological Dataset Integration project.

Mr. Kiley provided an update on the Wellcome Trust open access policy. All research papers, funded in whole or in part by the Wellcome Trust, must be made freely accessible from the PMC and UKPMC repositories as soon as possible, and in any event within six months of the journal publisher's official date of final publication. A significant number of commercial and not-for-profit publishers now offer an OA option that is fully compliant with the Trust's requirements, while other publishers allow the author to self archive a version of the final article and make that available within six months. Some publishers have policies that do not allow Wellcome-funded authors to publish. There is approximately 60% policy compliance from biomedical publishers. Mr. Kiley reported that it is too early to tell if Wellcome grantees are adhering to the mandate but there are encouraging signs and more advocacy and awareness are required.

A discussion ensued regarding copyright, trust issues, and compliance in scientific societies and small publishers. Dr. Lipman mentioned that copyright protection does not change regardless of server location: a society's journal site, US PMC, or UKPMC. Monetary repercussions on small societies were also discussed.

*Lunch 12:25-1: 25*

## **VI. International Language Support in PMCi**

Dr. Ostell provided an overview of a new international language project. Currently, PMCi is in English only. Support for other languages includes areas such as user interface, article mark up by contractors, article QA, and search/retrieval.

User interface support can be achieved by substituting phrases in XSLT rendering of web pages. Examples of citations in English, French, Spanish, Russian, and Chinese were shown. Dr. Ostell explained that for fixed parts of the page a language expert is required to edit tables of substitution.

International language support in articles requires contractors who will mark up non-English articles. QA however, must be done by PMC editors with language proficiency. Plans for this area include moving to a new QA system and support of some PMC processing by outside contractors which would allow QA by PMC editors from other countries with language

capability. Canada will start with French language support due to its two official languages. The new system is expected in 3-9 months.

Search support is more complicated because PMCi will contain English and non-English articles and to search together would require translation of articles or queries. This is an area requiring further investigation and an opportunity for others using PMCi to solve problems with their languages.

Lastly, Dr. Ostell informed the committee of a recent PMC feature for inline figure zoom capability. An example was shown with an explanation that the feature is served from a separate Tileshop database with the originals in pieces, enabling zoom capability.

## **VII. Books in PMC**

Dr. Ostell began with a diagram of the infrastructure of NCBI's bibliographic resources today and the future vision. The NCBI Bookshelf contains many books that are not traditional books such as GeneTests/GeneReviews, the Molecular Imaging and Contrast Agent Database, WormBook, and PubMed Help. A PMC article's XML structure was compared with a book XML structure (NLM DTD). While there are differences, Dr. Ostell explained that there are many similarities and a book chapter's structure can be comparable to that of a journal article.

The plan for the future version of Bookshelf includes books and articles in one system, making it easy to send content to PMC and PMCi in one format. Book content is currently being converted to the NLM Book DTD where the structure of the book content is designed to be the same as for the journal article where possible. By mid-October, all books will be converted and available on the PMC site. Soon after, many of them should be available in PMCi as a separate database but using the same structure and software. Dr. Ostell estimated that 40% of Bookshelf titles could be made available in PMCi immediately because the material is in the public domain, and at least another 20% will likely be made available by the respective copyright holders.

Mr. Kiley asked if there are IP issues regarding the translation project. Dr. Ostell answered that the articles will not be translated, only the citations and navigational controls. Dr. Ward asked how NCBI decides what will be included in the Bookshelf as a book, such as GeneReviews and OMIM. Dr. Ostell replied that one consideration is the long-term plan for the project and how it will present best, either in book format or as a database. It was asked if information is more difficult to find in books versus a database. Dr. Lipman answered that the Books database is being discovered by users and NCBI is working with Google to better index the database. This issue is one that will be addressed with the new discovery initiative. Dr. Ward commented that the database seems very useful from a teaching perspective, and suggested that there be pointers to the most frequently used figures. Dr. Velterop asked where the books come from. Dr. Lipman replied that they come from a wide range of sources such as publishers, authors who write monographs directly for the Bookshelf, and some documents produced by US government agencies. It was also asked if NIH-funded research is available in this format. Dr. Lipman replied that most NIH research is in journal article format. There is interest in having some systematic reviews incorporated into a structured book format.

## **VIII. Reports From the Field**

Ms. Ryan reported that many libraries are purging their own backfiles as they become available on PMC and others will be following suit, due to lack of storage space. There is also an organized effort by libraries to store duplicate copies in only a few locations when the information is available on PMC. Ms. Michalak reported that Duke University and the University of North Carolina are incorporating their chemistry and health sciences libraries for similar reasons.

### **Adjournment**

The PubMed Central National Advisory Committee adjourned the public meeting at 2:10 p.m.

### **CERTIFICATION**

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

---

Robert Kiley, Chair      (Date)  
PubMed Central National Advisory  
Committee

---

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