

**DEPARTMENT OF HEALTH AND HUMAN SERVICES**

**NATIONAL INSTITUTES OF HEALTH**

**PUBMED CENTRAL NATIONAL ADVISORY COMMITTEE**

**NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION**

**NATIONAL LIBRARY OF MEDICINE**

**January 16, 2003**

**NLM Board Room  
National Center for Biotechnology Information  
National Library of Medicine  
8600 Rockville Pike  
Bethesda, Maryland 20894**

**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**

Since the mission of NIH is to conduct and support medical research and to disseminate the results of that research widely to the public and the scientific community, it will make use of electronic publishing technology to fulfill this role by establishing and maintaining PubMed Central. This new service is a Web-based repository, housed at the NCBI, that will archive, organize, and distribute peer-reviewed reports from journals in the life sciences, as well as reports that have been screened but not formally peer reviewed. The Committee shall advise the Director, NIH, the Director, NLM, and the Director, NCBI, concerning the content and operation of the PubMed Central repository. Specifically, it is charged to establish criteria to certify groups submitting materials to the system, monitoring the operation of the system, and ensuring that PubMed Central evolves and remains responsive to the needs of researchers, publishers, librarians and the general public.

**SUMMARY MINUTES OF MEETING –JANUARY 16, 2003**

The meeting of the PubMed Central National Advisory Committee was convened on January 16, 2003 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 4:00 p.m. Dr. Joshua Lederberg presided as Chair.

**Members Present**

Joshua Lederberg, Ph.D., The Rockefeller University, PubMed Central National  
Advisory Committee Chairman  
Patrick O. Brown, M.D., Ph.D., Stanford University  
Nicholas Cozzarelli, Ph.D., University of California at Berkeley  
Michael Eisen, Ph.D., University of California at Berkeley  
Anthony Delamothe, BMA House  
Paul Ginsparg, Ph.D., Cornell University  
Michael Homan, Mayo Medical Center Libraries  
Richard Johnson, The Scholarly Publishing & Academic Resources Coalition (SPARC)  
Heather D. Joseph, BioOne  
Samuel Kaplan, Ph.D., University of Texas Medical School at Houston  
Elizabeth Marincola, American Society for Cell Biology  
Richard J. Roberts, Ph.D., New England Biolabs  
Sarah Thomas, Ph.D., Cornell University  
Linda A. Watson, University of Virginia  
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM, NIH,  
PubMed Central National Advisory Committee Executive Secretary

## **NLM Senior Staff Present**

Donald A.B. Lindberg, M.D., Director, NLM

Kent Smith, Deputy Director, NLM

Donald King, M.D., Deputy Director for Research and Education, NLM

Betsy Humphreys, Associate Director for Library Operations, NLM

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

Dr. Lederberg welcomed members of the PubMed Central National Advisory Committee. The Committee officially adopted the minutes from the previous meeting. Tentative dates of May 19-21, 2003 and November 4 or 17-20, 2003 were discussed for the upcoming two meetings. Dates will be set at a later time. Dr. Lipman then introduced Committee members and guests, welcoming new members and thanking departing members for both their time and contribution to PMC.

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

Dr. Lindberg thanked Dr. Lederberg for the opportunity to address the Committee. He began by thanking the Committee with an emphasis on the importance of advisory committees to the NIH. He credited Dr. Lipman's leadership for the progress made on the PubMed Central (PMC) database. He then discussed a suggestion made by Dr. Richard Roberts for a formal statement put forth by the National Library of Medicine (NLM) regarding a long-term commitment to PMC and a commitment to the preservation of data for PMC contributors. Dr. Lindberg is in favor a public statement of support.

Discussion ensued regarding a commitment by NLM to PMC and to providing an archive for PMC participants. Many committee members believe a public statement will be beneficial, especially for obtaining more journals for inclusion in the database.

In that context, there was discussion of a financial model for open access publishing, and the way page charges, which are part of the research process, can be handled. Dr. Lederberg emphasized two important issues in this statement – preservation and expenses. He believes that including alternative economic models could provide assistance to journals in paying publishing expenses. Committee members added that there are no granting agencies that will not pay reasonable publication costs.

Dr. Lipman would like to provide Dr. Zerhouni, the NIH Director, a summary of information on the role that NIH can play in open access publishing. A statement coming from NIH would clarify to scientists that nothing in NIH policy gives a preference to any particular type of publishing methods. Dr. Lederberg suggested that the NIH statement include a statement that all granting agencies will pay reasonable publishing costs. Dr. Lindberg suggested that the major benefit of PMC participation is preservation, an expensive process. Dr. Lipman mentioned that outside groups and organizations could be more influential than the PMC Committee regarding open access publishing.

Jane Griffith, NLM's Assistant Director for Policy Development, informed the Committee about the Emoluments Clause of the U.S. Constitution as it applies to Special Government Employees (SGE's). PMC Advisory Committee members are considered SGE's and as such must abide by the rules laid out in the Emoluments Clause. Ms. Griffith asked members to contact her if they

have any questions, problems, or concerns with the process of reimbursement from foreign governments. Members should also contact her if there is any question whether activities are in violation of the Emoluments Clause.

### **III. PubMed Central Update**

Dr. Lipman reminded the Committee that PMC is now integrated within the Entrez retrieval system and encouraged feedback from members regarding its performance. There are over 100 journals participating in PMC at this time with others in process. The Journal of Clinical Investigation is ready to be released. The PMC group is also working on three Cold Spring Harbor journals and about a dozen others.

Over half of the non-BMC journals are participating in PMC's PubLink option which takes users to the publisher site for full text but provides other PMC features. Five journals at this time are submitting data directly to PMC via the PMC DTD with more anticipated for the future. Dr. Lipman mentioned that getting publishers to attend to data quality continues to be a challenge but it is improving over time. There are now more than 80,000 items in the PMC database with PNAS and the ASM journals constituting the majority of the database due to the amount of their back issue material. The system averages 100,000 unique users per month.

Significant changes in PMC include its integration into Entrez which allows linking, cross-database searching, and other navigational features. Full-text searching is also a new feature, as is the identification of organism names and various database accession numbers in the full text. This, in turn, provides bi-directional links between PMC articles and the corresponding taxonomy, sequence, or structure database records. Another new feature provides a list of PubMed citations for the references in one or more PMC articles. New advanced search features allow searching of selected parts of articles, such as the methods section, or figure and table captions. Internal software changes have been made for better archiving and file management, providing faster article retrieval.

#### *NLM Archiving DTD*

Jeff Beck provided an overview of the newly completed NLM Archiving DTD which stemmed from a major redesign of the PMC XML DTD. The PMC team met with outside consultants and collected recommendations. The redesign of the PMC DTD was due to the desire for a modularized DTD so that many DTDs could be built using the same elements, giving both flexibility and consistency. The resulting product is a modular DTD suite, a set of building blocks from which any number of specific DTDs can be created with relatively little effort.

The "NLM Archiving and Interchange DTD Suite" has been created from a collection of all of the elements and attributes needed for journal articles. The "NLM Journal Archiving DTD" has been created from the DTD Suite. This is PMC's Archiving DTD which is now being used to convert data from multiple sources for the archive DTD. The Journal Article DTD, a subset of the Archiving DTD, is an authoring DTD used for original markup of the full text of a journal and is more prescriptive, requiring that particular kinds of data are always tagged the same way. The Archiving DTD and Suite are publicly available at this time and the Journal Article DTD is being tested for release soon. Mr. Beck then showed the Committee the documentation that accompanies the new DTD suite.

The Committee was enthusiastic about the new DTD Suite and the ease with which journals should be able to use it. Questions were raised regarding the customization and use of the DTD by current and new PMC participants.

Dr. Lipman discussed how to maintain control of the DTD Suite so that NCBI can map, update, and make changes as needed. However, NCBI is willing to allow others to use and modify it for their purposes as with other NCBI software and information. Dr. Lipman asked the Committee for input on how to make NCBI's investment in the DTD Suite as effective as possible for the publishing community.

The interest of the Library of Congress (LoC) archiving material was mentioned by the Committee. NLM staff has been in contact with the LoC and has suggested use of the NLM DTD Suite. Endorsement by the Library of Congress would clearly bolster publisher confidence in the Suite.

The concern of PMC participation was brought up by the committee with the expressed desire of improvement in recruiting journals. Now that barriers have been lifted with respect to content location, and PubLink is an option, there should be more involvement in PMC. Rather than merely being seen as a means to give information away for free, members stated that PMC should be seen as a beneficial public service due to open access, back issue scanning, and the archiving aspects of participation. Committee members want a more proactive approach to broadcasting this message to journals. The NLM is providing some assistance in this area. In March, NLM is participating in a meeting in conjunction with the American Medical Publisher's Association, where a presentation on archiving will use PMC as a case study. Also, NLM is approaching electronic journals indexed in MEDLINE about archiving strategy, and suggesting participation in PMC as a means for archiving their materials. Committee members reiterated that a more effective approach would be to get high profile scientists to represent PMC and speak on a colleague-to-colleague level to strategically targeted journals. Dr. Lipman agreed to look into society journals and find scientists on their boards willing to advocate participation in PMC. Committee members made suggestions of journals that would be good candidates to target.

### ***The Committee convened for lunch from 12:35-1:00***

#### ***PMC Demonstration***

Dr. Lipman provided a brief demonstration of the new interface and functionality of PMC as part of the Entrez retrieval system. He illustrated the increased retrieval speed, the "free full text" message associated with articles located in PMC, and a new "Smart Search" filter. Also illustrated were new links to the Taxonomy database based on organisms extracted from the full text of the article. Links to the Structure database and accession numbers relevant to the article were demonstrated as well as navigational and retrieval aspects of linking search results. Committee members were interested to know how indexing and information linking was implemented in Entrez. NCBI's E-Utilities were mentioned in response to a question about automated access to the linking information for research purposes. The Committee was asked for feedback and ideas on how to make NCBI retrieval tools and cross-database searching more powerful.

#### **IV. Back Issue Scanning Project**

Martha Fishel, from NLM's Library Operations Division, provided the Committee with an update on the NLM Back Issue Scanning Project proposed last January. Basic terms for publishers to participate are that the NLM will cover the cost of cover to cover scanning back to the first issue in return for permanent rights to archive and distribute the material freely through PMC within copyright restrictions. The contractor and contract award were discussed, along with vendor evaluation criteria and vendor workflow. OCR text will be generated from the scanned pages to support full-text searching. Article headers and abstracts will be keyed as XML-tagged records which will be included in both PMC and PubMed. Each journal has contributed whatever disposable print copies they had available and NLM is using contacts to get the rest of what is needed. Each journal will receive a complete electronic copy of its material.

The goal of the project is to scan 52 journals, with a total of 564,800 articles. The first articles should be available online beginning in July in PDF format and will be accessible through a table of contents for each journal issue. The benefits of such a project include full back files for 52 titles, enhanced service to the scientific and research community and content preserved in a readily accessible format. The Committee was enthusiastic about the progress of the project and commented that this additional archival content in PMC could prompt greater participation by publishers.

#### **V. Uses of Usage Statistics**

Dr. Paul Ginsparg of Cornell University presented information on analysis and utilization of web usage statistics. The data presented by Dr. Ginsparg were collected from logs of full text downloads from his physics web site over a period of six years. Dr. Ginsparg discussed factors that can influence the statistics and conclusions of analysis. Analysis must take into account the possible causes of spikes in usage such as new submissions, study reports presented at society meetings, referrals to an article from another source, and type of article such as review articles. Paying attention to signals in data can provide information such as interest in certain types of articles.

Dr. Lipman expanded upon usage analysis of PMC which showed that usage was based more on information provided rather than the journal name. Relevance to a query seemed to be the deciding factor for viewing an abstract. This contradicts the idea that users give preference to 'name' journals, although the number of articles printed by a journal has an impact on the number of citations in the search results and therefore the number of hits to the journal.

Committee members expressed an interest regarding whether hit rates for particular papers are greater at PMC or the publisher's site. Difficulty could arise if PMC use negatively affects a publisher's site usage, and thereby discourages journal participation. Dr. Lipman commented that this is a possible factor in publisher participation but expressed doubt because users are given the option of going to the publisher site directly at every step of the query and retrieval process. Thus far, there is no indication that usage has declined on publisher sites because of PMC participation. Dr. Lederberg suggested that a flag of credit be provided to the publisher on their articles such as "courtesy of". The Committee felt that this type of credit could alleviate the concern of brand identity by some journals.

*Update of Books project*

Dr. Lipman gave a brief overview of the status of the Books site usage. At this time there are 18 books in the database with about 25 more in process.

Jo McEntyre gave a summary of the Monographs series from the Books site which consists of live reviews of scientists' work. There are approximately five monographs lined up at this time.

**VI. Conclusion**

Dr. Lederberg expressed the need for more emphasis on developing grassroots support for both open access and NLM in general. More thoughts on promotion will be discussed at the next meeting.

Drs. Lederberg and Lipman thanked both the Committee members and invited guests for their valuable time and input.

**VII. Adjournment**

The PubMed Central National Advisory Committee adjourned the public meeting at 4:15 p.m.

**CERTIFICATION**

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

---

(date)  
Joshua Lederberg, Ph.D., Chair  
PubMed Central National Advisory Committee

---

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