

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 14, 2002

**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 14, 2002

The meeting of the PubMed Central National Advisory Committee was convened on January 14, 2002 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
Frank Davidoff, M.D., *Annals of Internal Medicine*
Tony Delamothe, BMJ Publishing Group
Uta Francke, M.D., Stanford University Medical Center
Paul Ginsparg, Ph.D., Los Alamos National Labs
Michael Homan, Mayo Medical Center Libraries
Samuel Kaplan, Ph.D., University of Texas Medical School at Houston
Elizabeth Marincola, American Society of Cell Biology
Suzanne McInerney, Health Writer/Patient Advocate
James G. Neal, Johns Hopkins University
Richard J. Roberts, Ph.D., New England Biolabs
Vitek Tracz, Current Science Group
Harold Varmus, M.D., Memorial Sloan Kettering Cancer Center
James F. Williams, University of Colorado at Boulder
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM, NIH,
PubMed Central National Advisory Committee Executive Secretary

Invited Guests Present

Mark D. Doyle, Ph.D., American Physical Society

Eugene Garfield, PhD., *The Scientist*

Howard Ratner, *Nature*

Maxine Rockoff, New York Academy of Medicine

NLM Senior Staff Present

Donald A.B. Lindberg, Director, NLM

Kent Smith, Deputy Director, 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 6, 2002 and November 4, 2002 were set for the next two meetings. Dr. Lipman then introduced invited guests and Committee members, welcoming new members and thanking departing members for both their time and support.

II. Remarks by NLM Director

Dr. Lindberg thanked Dr. Lederberg for the opportunity to address the Committee. He began by discussing recent trends in electronic database usage, highlighting the rapid increase in use of NLM electronic databases such as MEDLINE and PubMed (bibliographic databases) and noting growth in MEDLINEplus, NLM's health information service. He elaborated on the expanding relationship between MEDLINEplus and other online health information providers, such as WebMD, and mentioned that WebMD was now posting online scientific meeting summaries/reports geared towards health practitioners. Discussion ensued about the appropriateness of disclosing non-peer-reviewed literature to clinicians.

Dr. Lindberg also referenced NLM's recent partnership with the Library of Congress and discussed how the two institutions, each having a similar mission, were sharing approaches for building digital collections of print material. As an example, Dr. Lindberg mentioned NLM's "Turning the Pages" exhibition" and the Library of Congress's American Memory project. Both projects serve as a gateway to source materials relating to the history and culture of the United States.

Lastly, Dr. Lindberg mentioned NLM's plans for constructing a new building that would be located on the NIH campus. He emphasized that the new building would provide ample space for NLM's and NCBI's expanding activities as well as space for multi-purpose public areas.

III. PubMed Central Update

Dr. Lipman reminded the Committee that the redesigned PMC system, which is more robust and scalable than the original system, has been up and running for the past six months. He then introduced Mr. Sequeira who, with the help of various members of the PMC development and operations team, provided the Committee with an update on PMC's status and overview of new features. He demonstrated for the Committee the enhanced capacity to search and retrieve relevant information from within the PMC archive as well as the addition of a number of new links, such as a link between a correction and the original article; from an article to other articles in PMC that cite it; from a citation in the reference section to its full text in PMC; and from an article to related records in the Entrez databases. Also mentioned was the development of a full-text searching feature for PMC.

Dr. Trawick briefly demonstrated NCBI's new Bookshelf project, a growing collection of biomedical books adapted for the Web based on the concept that the textbook will provide accessible background material that a user can explore to understand unfamiliar concepts found in a PubMed or a PMC search result. Bookshelf may be accessed from the "and links to:" pull-down menu in PMC if the article retrieved is directly cited in the book. Once retrieved, book sections are listed in order of relevance to the search term. Each book "page" has a navigation bar that allows other parts of the book to be found and all figures, tables, and boxes are hyperlinked where cited in the text. Books can also be searched from the search box found at the bottom of the navigation bar.

The Committee noted progress on the Bookshelf project and much discussion ensued about the possible addition of certain features, such as pagination matching the original text and highlighting within the retrieved text those words or phrases searched. Dr. Lederberg suggested that NCBI consider adding the "Textbook of Military Medicine" published by the Office of the Surgeon General, U.S. Department of the Army. Dr. Lipman agreed to look into this option. He also pointed out that the Bookshelf team was in the process of converting NCBI's Genes and Disease Web pages into a Web-based book that would be available through Bookshelf. This led to a discussion on the public benefit of NCBI posting additional scientific monographs, an option currently under consideration.

The Committee then convened for lunch from 12:00-1:00

Mr. Beck informed the Committee that all material deposited in PMC is now stored in a common format, the PMC XML DTD. Using a single scheme to mark up the elements of every article allows for a more effective and portable archive. Currently, journals can provide data in their own DTD, and the PMC team makes the translation into the PMC format. NCBI would like to make this new format public, and so decided to have it reviewed by a consulting group specializing in related technologies. Based on their evaluation, the current DTD will be revised along a number of different lines. The most significant change will be to split the DTD into two subsets: one for archival purposes that will accommodate conversion from any publisher-supplied DTD and a second for anyone wishing to use the PMC DTD for original markup. The benefit of making such changes is that the new DTD should better accommodate material received from new sources and therefore, be less subject to constant tweaking and patching.

In order to encourage wider publisher participation in PMC, an option was recently introduced that requires full submission of published material to PMC, but permits viewing of full-text at the publisher's site. Specifically, the publisher agrees to furnish the NLM with full-text articles in a format that conforms to an established DTD. NCBI transforms this data into a PMC DTD and loads the data to create the PMC archive. From this common archival format, indices would be built to facilitate end-user searching of journal articles. However, articles retrieved in searches have links to the publisher's site where the user would then be able to view the full-text. Mr. Sequeira informed the Committee that the software required to perform this function has been fully implemented and that these "Publink" articles include many of the features that have been developed for standard PMC articles. PMC also indicates whether an article is restricted or available free of charge from the publishers site. Mr. Sequeira closed by inviting the Committee to review the handout for an updated listing of those journals that are in production, near production, in process, and informally committed.

The Committee was pleased with the rapid progress NCBI had made in launching this new option. Members wondered whether "restricted" was the appropriate term for articles available only from a publisher's site. They suggested that NCBI consider more positive wording as well as listing an approximate release date for those article designated as restricted. Dr. Lipman agreed to implement alternate phrasing, with the exact wording to be determined by PMC team members.

IV. NLM's Plans for Archiving Electronic Journal Literature

Ms. Betsy Humphreys provided an overview of the NLM's long-range plans for archiving electronic journal literature. She began by defining NLM's statutory functions, which include acquiring and preserving library materials pertinent to medicine and making these materials available to the public. When carrying out these functions, relevant NLM policy is to first preserve the content of journals indexed by NLM and to encourage publication in permanent media to reduce future preservation costs. NLM's Long Range Plan for 2000-2005 further states that NLM will take a leadership role in both organizing and providing permanent access to electronic information. To accomplish this, NLM first had to define what conditions make digital information permanently accessible, such as deposition of the material in more than one autonomous repository; access at each repository; and the transfer of content to other repositories. Ms. Humphreys pointed out that the establishment and maintenance of the PMC archive will provide the necessary infrastructure to allow NLM to fulfill its mission of not only preserving library material but of organizing and providing permanent access to electronic information.

A Committee member inquired whether NLM currently supported any "data repositories," and, if so, were the grantees required to store and provide access to their data through alternate or mirror sites. Ms. Humphreys replied that the establishment of "mirror sites" was not a requisite, although encouraged where possible for a variety of reasons. She used Dr. Roberts' REBASE repository as an example. Dr. Roberts pointed out that although he was not required to establish an alternative site, that for security and accessibility reasons, he did have his files backed-up at another location.

V. Archiving Scanned Images of Print Material

Dr. Mark Doyle of the American Physical Society (APS) demonstrated the Physical Review Online Archive, or PROLA, for the Committee. The PROLA archive contains every paper in every APS journal that has been published since 1893. PROLA includes scanned images of all articles; full-text searching; references and forward citations; comments, replies, and errata linked to and from articles; hyperlinks to referenced papers published by APS as well as other participating publishers; and links to supplemental data sets. Each year the most recent content is updated by transferring material into PROLA from current journal files. Papers published in the current year, along with papers published over the past three years, are held separately and can be accessed only by subscribers. Non-subscribers can view the abstracts and table of contents and can obtain copies of individual article for a fee. Currently, Cornell University Library maintains a mirror copy of PROLA that can be accessed by anyone in accordance with APS's subscription scheme. APS has also signed an agreement with the Library of Congress to provide a long-term repository for the files found in PROLA.

The Committee thanked Dr. Doyle for taking the time to attend the PMC meeting and for presenting on PROLA. Committee members were impressed with the archive and had a number of questions relating to the management of and costs associated with archiving new content versus back content. Dr. Lederberg pointed out to Committee members that electronically archiving past content from participating journals might be a way to get new publishers to sign on to PMC, but felt it would be up to the NLM to expend the resources necessary to get the back content up and running. Dr. Lipman agreed and will approach current PMC publisher members about their interest in archiving past content to see if this is a project worth pursuing. He also agreed to put together an estimate of what it would cost NLM to undertake such a project and to approach non-participating publishers to see if there was interest.

VI. Conclusion

Dr. Lederberg 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 3:45 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