

**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 26, 2006

The meeting of the PubMed Central National Advisory Committee was convened on April 26, 2006 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. Dr. Samuel Kaplan presided as Chair.

Members Present

Camila Alire, Ed.D., University of New Mexico
Shirley Baker, M.A., Washington University
Heather Joseph, M.A., SPARC
Daniel Greenstein, Ph.D., University of California
Samuel Kaplan, Ph.D., Houston Medical School
Robert Kiley, M.Sc., Wellcome Trust
Isaac Kohane, M.D., Ph.D., Harvard Medical School
Debra Lappin, J.D., Princeton Partners Ltd.
Hemai Parthasarathy, Ph.D., Public Library of Science
Bob Roehr, B.A., Self-Employed
Mary Ryan, MLS, University of Arkansas Medical Sciences
Anthony So, M.D., Duke University
John Wilbanks, B.S., Science Commons
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, IEB, NCBI
Dennis Benson, Branch Chief, IRB, NCBI

Jane Davenport, IEB, NCBI
Brooke Dine, IEB, NCBI
Mark Desierto, IEB, NCBI
Jason Eshleman, IEB, NCBI
Martha Fishel, PSD, LO, NLM
Marla Fogelman, IEB, NCBI
Al Graeff, NCBI
Demian Hess, IEB, NCBI
Betsy Humphreys, Deputy Director, NLM
Jennifer Jentsch, IEB, NCBI
Laura Kelly, IEB, NCBI
Donald King, Deputy Director, Research and Education, NLM
Sheldon Kotzin, Chief, BSD, NLM
Sergey Krasnov, IEB, NCBI
Dawn Lipshultz, NCBI
Adeline Manohar, IEB, NCBI
Jim Ostell, Branch Chief, IEB, NCBI
Edwin Sequeira, IEB, NCBI
Kent Smith, NCBI
Bart Trawick, IEB, NCBI
Deborah Zarin, NLM

Visitors Present

Rebecca Chasen, Oxford Journals
Norman Frankel, Director, American Medical Association, Allied Publishing Activities
Mark Leader, ASCB
Manon Ress, CP Tech
Michael Rogawski, NINDS, NIH
Carl Rhodes, Howard Hughes Medical Center
Crispin Taylor, Executive Director, American Society of Plant Biologists
Neil Thakur, NIH, OD

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. The new Chair, Dr. Samuel Kaplan, was introduced and the agenda reviewed. Minutes from the October 2005 meeting were approved. The next PMC Advisory Committee meeting will take place on October 26, 2006. The following meeting is tentatively scheduled for April 19, 2007.

II. Remarks by NLM Deputy Director

Ms. Betsy Humphreys told the group that 2006 is the 170th anniversary of NLM and the 50th anniversary of the National Library of Medicine Act which transferred the Armed Forces Medical Library to DHHS. A long range plan is in the works for NLM, which includes meetings of four panel groups to help guide NLM's future projects and services. The NLM Long Range Plan includes continued expansion of digital archiving and public

access. She reported that the Library of Congress and British Medical Library are adopting and endorsing the NLM DTD as the standard for archiving digital content.

Ms. Humphreys shared a recent statistic that 16% of PubMed articles are linked to free full text via the LinkOut system.

Ms. Humphreys also reported statistics of outreach by the National Network of Libraries of Medicine and NLM directly over the last five years. She attributed an increase in usage of consumer sites to this outreach. NLM is also reviewing how improvements can be made in training and linking its resources. She then thanked the group for their time in serving on the committee.

Dr. Lipman reported on a new NIH endeavor, the Whole Genome Association (WGA) resource, in which NCBI will provide a combination of genotype and clinical phenotype information. The genome project originally developed tools for understanding the genetic basis of disease such as the HapMap and genome-wide approaches using various types of chip technologies. Through a project called Genome Association Information Network (GAIN), a collaboration between Pfizer, Affimetrix, and the Foundation for NIH, funding will be available for acquiring genotyping data for existing clinical studies.

NCBI has developed a database for storing the clinical and genetic information from GAIN studies for research use. NIH ICs will provide authorization for those who request access to proprietary information and the sponsoring institutes will decide what type of information will be included in the database. Dr. Lipman explained that ongoing discussions are addressing the policy aspects related to access to the combined phenotypic and genotypic data.

Dr. So asked about data handling in the HapMap project and licensing issues with the WGA project. Dr. Lipman replied that at this time, policies are still being formulated and they are trying to prevent trivial patenting of information. Some pre-computed data associations will be provided by NCBI to try to make as much information publicly available as soon as possible. Dr. Kaplan asked if researchers can publish with data extrapolated from the WGA resource. Dr. Ostell answered that there is a nine month period where those involved in the studies have exclusive rights to publications.

IV. ClinicalTrials.gov

Dr. Deborah Zarin provided a presentation on ClinicalTrials.gov, a project of the Lister Hill Center for Biomedical Communications, NLM. ClinicalTrials.gov contains 28,900 registered studies including 13,700 open for recruitment and 15,200 closed. Trials come from various providers: 13,200 from Federal agencies, 8,700 from universities/foundations, and 7,000 from industry. There are 8.6 million page views and 470,000 unique visitors per month to the website.

Dr. Zarin demonstrated examples of the ClinicalTrials.gov website. Various examples of trial summaries were shown and the information found within each record such as references and links to PubMed and MEDLINEplus. Unpublished results are not

accepted.

Data are entered by data providers through the Protocol Registration System, a web-based data entry system. As of October 2004, registrants from anywhere in the world could register their trials in ClinicalTrials.

Recent events have resulted in a large number of trial registrations. The International Committee of Medical Journal Editors (ICMJE), a group of medical journal editors, required registration of trials for new studies by July 2005, in order for any report of a trial to be accepted by one of the journals. The ICMJE deadline for registration of ongoing trials was September 2005.

An FDA report on implementation of FDAMA 113, the law that requires registration of trials of drug treatments for life threatening illnesses, resulted in an increase in compliance from 2002 to 2004. Reasons for non-compliance were cited as a lack of awareness, a business decision to cease trials and unwillingness to comply.

Quality assurance includes the Protocol Registration System which holds organizational accounts and performs automated validation. A QA team manually reviews the records for internal consistency and duplicate registrations from multisponsored or multisite trials. Evidence of IRB approval is required such as contact information or an IRB approval letter. An analysis of some key data elements shows that many companies fail to define primary outcome measures for a trial. Similarly, many industry submissions for interventional trials do not identify the specific drug being studied.

Dr. Camila Alire asked if the database is being marketed to libraries and the general public. Dr. Zarin answered that the NLM, MLA, and PLA have recently promoted ClinicalTrials.gov. Dr. Alire asked if it is assumed that Google searchers are consumers and added that there may be a disadvantage to minority groups without access to computers. Ms. Humphreys reported that an effort is underway for more outreach to community groups, public health resources, and clinicians. Dr. Isaac Kohane asked if leading journals are checking for reporting of trials in the database. Dr. Zarin replied that journals are checking more frequently due to the recent ICMJE deadline. Dr. Kohane's question of whether a controlled vocabulary was used was answered in the affirmative. He also asked about a Google AdSense program in which keywords can be bought. It was answered that NLM cannot do this due to federal regulations. Ms. Lappin asked if usage trends are tracked and asked about a role for ClinicalTrials.gov in negative clinical trial results. Dr. Zarin replied that having access to knowledge of studies in general can be helpful for researchers, but there is a need for reporting results. Mr. Roehr asked about the 5% of non-registered NIH trials. Dr. Zarin answered that there was some confusion regarding policy, recent deadlines, and NLM is making an effort to close the gap.

Break 10:55 -11:10

V. PMC Update and NIH Public Access Policy Status

Dr. Lipman apprised the group on current PMC highlights. Three publishers, Blackwell,

Springer, and Oxford University Press offer authors a hybrid publication model. Authors may publish in the traditional way, where only subscribers have access for some time after publication, or they may pay an open access fee, in which case their articles will be available for free immediately. All three publishers have agreed to deposit the final published version of open access articles in PMC. Blackwell and Springer have already started to do so. The publisher deposits relieve the corresponding authors from having to deposit their manuscripts in PMC under the NIH or Wellcome Trust public access policies.

A set of detailed tagging guidelines for the NLM Journal Publishing DTD has been released. A new piece of software, Style Checker, is being used to automatically check that tagging is correct. The new guidelines and tools make it possible for vendors to do better QA of data before sending files to PMC.

Dr. Kohane noted that the open source community may be willing to generate PMC-compliant XML documents. Dr. Lipman answered that they are hoping to engage a wider community to broaden the application of these tools. Dr. Greenstein added that new journals in various formats and institutional repositories are producing opportunities as well as challenges for open access and one way to unify them would be to have standard tools. Dr. Lipman replied that some larger companies may be willing to create publishing platforms and frameworks for web communications in various disciplines. Dr. Rogawski asked about organization of articles coming in through open access agreements. The articles will have branded pages similar to regular PMC journals, but no table of contents will be available. The articles can be found via PubMed or PMC queries. Mr. Kiley suggested that an open access journal category could be formed for such articles. Mr. Sequeira replied that PMC is looking at a different model for these articles.

The Library of Congress and British Library are adopting the NLM Journal DTD as a standard for journal archiving. The Library of Congress has issued a press release to this effect. Ms. Ryan expressed that this is an exciting development and would like the library community to know more about the DTD.

As of March 2006, there were 571,000 articles available in PMC with 60% from the back issue digitization project, the oldest article dating back to 1865. PMC gets 2.35 million unique users per month. Six million articles are retrieved and 9.5 million pages are viewed each month. Ms. Humphreys noted that the numbers are higher for April, and climbing each month. Dr. Greenstein asked if NLM is seeking back files for digitization. Dr. Lipman answered that back issues are scanned for PMC participants, adding that Wellcome and JISC have helped gain journal participants.

NIH Public Access Policy

The Public Access policy has seen only a 4% participation rate since inception last May. Dr. Lipman reported that the submission system itself is working well. He reviewed the submission and quality control process and explained the improvements that have been made to benefit publishers who submit on behalf of their authors. Publishers no longer have to provide grant information when submitting an article. They can also submit

articles in bulk, bypassing the interactive NIHMS system. Although several publishers have shown an interest in submitting manuscripts on behalf of their authors, it is not clear how much this will improve overall participation rates, because authors are still required to log in to the NIHMS to verify the accuracy of a tagged manuscript and approve its release. A more effective approach that has been discussed with some publishers has been labeled “Partial PMC,” where the publishers would submit the final published version of all NIH-funded articles in XML-tagged format. In this case, there would be no need for an author to review a submitted article, because the publisher would be providing an authorized version, just as with regular PMC participants. So far, no agreements have been reached with any publisher about Partial PMC. Dr. Lipman noted that Partial PMC participation by publishers could result in researchers choosing to publish in their journals due to assistance in complying with PA policy. Researchers would like one way to deposit their data so that little time is spent on the compliance and submission process. It was stated that if the policy were mandatory, more people would comply and it would help authors if reporting and deposition could be automatically connected with their bibliography. Questions were asked about the possibility of a mandatory policy. Dr. Neil Thakur stated that NIH may or may not receive guidance from Congress and that Dr. Zerhouni is getting comments on the policy from both the public and publishers.

Dr. Kaplan reported on the most recent Public Access Working Group meeting where it was concluded that the voluntary policy is not working well. The group recommended a mandatory policy with final copy edited versions of articles available within six months of publication. The group believes that Partial PMC would work well for smaller publishers. Dr. Rogawski asked about copyright agreements and public access. Dr. Lipman answered that NLM will remove articles in PMC upon publisher request when those articles have been deposited in violation of a copyright transfer agreement. An exception is the situation where an author is a government employee, in which case the articles are in the public domain.

Lunch 12:30-1:00

VI. NLM Policy on PMC Usage Reports

Ms. Humphreys reported that some potential PMC participants would like to receive usage statistics on their material broken down by institution. They claim that libraries demand such statistics from them. However, it is general NLM policy to not keep or distribute individual statistics. It is known that such information is requested from publishers for journals under subscription but the material in PMC is already free. It was noted that PMC provides many other types of statistics, not broken down by user. Committee members agreed with, and endorsed, NLM’s stand. It was asked if library groups may have a comment on usage requests. Some members of the committee offered to poll their colleagues on this issue.

On a side note, Ms. Humphreys provided a statistic that 92% of all holdings in PMC have been viewed at least once. Ms. Ryan informed the group that some libraries are purging back issues of journals that are available in PMC due to physical space constraints.

VII. Update on PMC International

Dr. Ostell provided an update on PMC International (PMCI), the project to set up PMC mirror sites. PMCI sites eventually will be able to add local content as well. This will result in independent repositories and international article archives, with data exchange similar to that of the sequence databases (GenBank, EMBL, and DDBJ). The adoption of the NLM DTD by national libraries and publishers will help accelerate this process.

PMCI sites will use the portable PMC (pPMC) software developed by NCBI. Currently, Version 2 has been successfully tested. The British Library is scheduled to install the new version this week, with Japan and China to follow. Upgrades will be provided to Italy and South Africa next. Features of Version 2 include better bandwidth use for updates, web service updates vs. FTP, status reports, and automatic integrity checking. In Version 3, abuse filtering will be included, as well as a usage statistics package. A pPMC public demonstration site, using only the Open Access subset of PMC data that may be redistributed with very few restrictions, is also planned.

A portable version of the NIH Manuscript System (pNIHMS) is being developed. A demo version will be available in May and packaged for remote installation in July. The system supports article submission, linking to grants, exchange of data with tagging contractors, operations Q/A and author Q/A, and deposition into PMC. The system will be deployed to the Wellcome Trust site but will be available to others as well.

Dr. Kohane asked if individuals would be able to set up pPMC on their own desktops. Dr. Ostell replied yes. It uses Microsoft's IIS as a web server and works on the Microsoft SQL Server. Updates will be available to users. Mr. Wilbanks asked if the code will be open source. Dr. Ostell answered that code will be in the public domain. Base code is NCBI's C++ Toolkit which is already available as well as documentation. Dr. Greenstein asked if the grant reporting system as part of pNIHMS is available. It is not yet available to the general public but will be by year's end. Dr. Ostell reminded the group that NCBI's 'MyBibliography' (a PubMed tool for creating and updating personal bibliographies) will only link to articles in PubMed but institutions may personalize the tools for use by their own staff. Dr. Parthasarathy asked if it is possible to find only one author's articles. The user would check one author's articles and save, and the next query will show all new articles since the last query. This could also be linked to the grant reporting system. It was asked if others could view such a bibliography. Dr. Ostell replied it may be possible, but the focus thus far has been on privacy.

Dr. Kaplan asked if other federal funding agencies have been approached on the public access issue. Dr. Lipman answered that NIH was the first choice due to health impacts on the public and the large amount of funding dollars received by NIH. Dr. Kaplan noted that PMC can include a wide range of life sciences literature, not only biomedical literature.

VIII. 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.

Samuel Kaplan, Chair (Date)
PubMed Central National Advisory
Committee

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