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 – November 22, 2004

The meeting of the PubMed Central National Advisory Committee was convened on November 22, 2004 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 1:35 p.m. Mr. James Williams presided as Chair.

Members Present
Michael Eisen, Ph.D., University of California at Berkeley
Richard Johnson, The Scholarly Publishing & Academic Resources Coalition (SPARC)
Heather Joseph, M.A., BioOne
Samuel Kaplan, Ph.D., Houston Medical School
Paula Kaufman, M.B.A., University of Illinois at Urbana-Champaign
Debra Lappin, J.D., Princeton Partners Ltd.
Sarah Thomas, Ph.D., Cornell University
Linda A. Watson, M.L.S., University of Virginia
James Williams, M.S., University of Colorado at Boulder
David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM, NIH, and PubMed Central National Advisory Committee Executive Secretary

NLM Staff Present
Dennis Benson, Ph.D., Branch Chief, NCBI
Martha Fishel, NLM
Betsy Humphreys, Associate Director for Library Operations, NLM

Donald King, M.D., Deputy Director for Research and Education, NLM
James Ostell, Ph.D., Branch Chief, NCBI
Ed Sequeira, PubMed Project, NCBI

Visitors Present
Laura Brockway, Federation of American Societies for Experimental Biology
Karen Schools Colson, Association for Research in Vision and Ophthalmology
Morna Conway, The Conway Group
Bridget Coughlin, Proceedings of the National Academy of Sciences of the United States of America
Ray Everngam, American Society for Cell Biology
Linda Illig, American Society for Microbiology
John Inglis, Cold Spring Harbor Laboratory
Krista Koziel, American Society of Human Genetics
Jana Lieberman, HHS
David Martinsen, American Chemical Society
Nancy Rodman, American Society for Biochemistry and Molecular Biology
Michael Rogawski, National Institute of Neurological Disorders and Stroke
Beth Rosner, American Association for the Advancement of Science
Crispin Taylor, American Society of Plant Biologists
Nancy Winchester, American Society of Plant Biologists

I. Call to Order and Opening Remarks
The meeting was called to order at 9:33 a.m. Mr. Williams welcomed members of the PubMed Central National Advisory Committee. Committee members were then introduced. The Committee officially adopted the minutes from the May 2004 meeting. Four members are ending their membership on the committee, Rick Johnson, Chaitan Khosla, Sarah Thomas, and Linda Watson. The date of April 28, 2005 was confirmed for the next meeting and a tentative date of October 20, 2005 was set for the following meeting. An addition was made to the agenda for an update by Rick Johnson on the authors' addendum to publication agreements.

Mr. Williams requested an update on the Fairness in Clinical Trials Act. Betsy Humpreys explained that this act is trying to get clinical trials registered at the time of set up and recruitment. The FDA Modernization Act in 1997 required that clinical trials sponsored or governed by NIH or FDA had to be registered in a public database. This led to the development of the NLM database ClinicalTrials.gov. It covers only certain categories of trials funded and managed under FDA's aegis. There is currently an interest to have privately funded trials and their summaries available to the pubic as well.

At this time ClinicalTrials.gov holds approximately 12,000 registered trials. Dr. King added that the database also acts as a place for recruitment. A problem has been low compliance, only about 50%, by companies because there are no penalties for noncompliance. Another foreseeable problem could be a lack of results reporting. The Clinical Trials database has been opened to the international community and five international trials are registered at this time. However, an authenticating body for these

trials is not always known. A decision will be made by Congress next year on whether it will be mandatory and how many results of trials must be reported. The International Committee of Journal Editors published a joint statement this fall saying that at some point next year, articles with unregistered trials would not be accepted.

Mr. Williams asked about current results reporting and Dr. King responded that results must be reported to the FDA but NLM does not have access to them because it is proprietary information. Some drug companies are putting trials and a summary of results on their own websites.

II. Public Access to NIH-Supported Research

Dr. Lipman noted that PMC Advisory Committee does not advise NIH on the NIH public access policy decision. November 16 was the last day for public comment on the proposed NIH policy. At this time the comments are being taken into account in the report Dr. Zerhouni will provide to the House. In light of the many publishers present, Dr. Lipman decided to present an overview of PMC capabilities and the possible impact the NIH public access policy could have.

PMC Overview and Usage

Dr. Lipman provided an overview of PMC, usage statistics, functional capabilities, and archiving.

In October, there were 160 journals in PMC and 350,000 items, with more than 50% of these from the back issue digitization project. 920,000 unique IP addresses, representing an estimated 2 million-plus users, used PMC during the month. Growth in 2005 is expected to include 20-40 new journals with 40,000-60,000 new articles supplied by publishers and about 300,000 from the back issue digitization project. A report on site performance and availability by a commercial monitoring company indicates that PMC is running well compared to the benchmark level for corporate Web sites.

Functional Capabilities

Using examples, Dr. Lipman then illustrated the variety of links available between PMC articles and related NCBI resources. These links allow users to find information that would not generally be found with a direct search. He demonstrated links to PubChem, a new small molecules database which holds chemical information and structures, PubMed, and sequence databases.

Dr. Lipman next provided an example of digitized back issue content as it appears in PMC. Computational analysis of the OCR text has made citation matching on the order of 90% possible for many journals.

Automated analysis is also performed to detect and link accession numbers, structures, and organisms to other NCBI databases. Dr. Lipman mentioned some issues encountered in digital journal archiving including the quality of source materials. He said that active use of the materials, such as in a searchable database, helps to confirm effective

preservation and quality. Distribution of content to collaborating archives adds validation and security.

Tentative Public Access Workflow
Dr. Lipman provided information on a tentative workflow should the NIH public access policy be enacted. He discussed the topics of author submission, conversion to standard XML format, allowance for author review, and access to the public. All author manuscripts would be converted to a common NLM journal article XML format by a vendor. Having all content in one format, versus the variety of data formats in which PMC currently gets material from participants, would make intake of author manuscripts into PMC simpler in some ways.

Q: Beth Rosner, AAAS, asked about the look of manuscripts, branding, and figures and legends.
A: Dr. Lipman answered that PMC participants are given a menu of formatting and article presentation options to choose from, in addition to the journal banner on each page page which can include links back to the publisher's site. For public access articles there would be a common presentation style for all journals that are not PMC participants.

Q: Ms. Rosner asked if PMC would ever allow advertising on the site.
A: Dr. Lipman said that advertising presents an issue for a government website and is not available at this time. However, PMC makes it easy for users to go to the publisher site via links in the journal banner. In addition, the LinkOut service provides links to commercial and vendor sites.

Q: Mike Rogawski, NINDS, asked if NIH has considered no six month moratorium for papers published by intramural scientists because they are prohibited from copyright.
A: Dr. Lipman does not know if this will be a policy decision. PMC is equipped to take such articles with permission from publishers. Members expressed encouragement for early submission of articles.

Q: Heather Joseph asked about cost estimates for tagging of author manuscripts.
A: Dr. Lipman said that estimates are being received but no decision on the provider for tagging has been finalized.

Q: Debra Lappin stated for the record that she has worked as a policy consultant to an advocacy group on the public access issue and had been cleared for attendance at today's meeting by the NIH Ethics office. She asked about a cost element in the policy, and if any other data is available such as the average grant costs represented by a single paper.
A: Dr. Lipman said that there are no official costs, but if an analysis were to be done the number probably would be at least $150,000 per paper.

Q: Nancy Rodman, ASBMB, asked about the possibility of linking to the journal site for an article rather than having it reside in PMC.
A: Dr. Lipman responded that having articles in one place results in a higher quality archive and provides an opportunity for new uses of information as computational

methods advance. Dr. Eisen added that functionality is limited if content is not housed in one place in one format. In addition, user feedback is essential to database improvement and would be difficult to gather with articles appearing on various sites. Ms. Watson noted that data mining is improved when, specifically, all full text information appears in one place.

Q: Laura Brockway, FASEB, asked if the author would be able to correct content errors after the article has gone through copy editing.
A: Dr. Lipman answered that a communicating submitter will be asked to contact PMC as soon as the article has been published, if any significant content errors have been detected. The goal is to capture major errors and correct them before the article is released in PMC six months after publication. It was restated that if the journal publisher provides a copy edited version of the article, that version will become the default view.

Q: It was asked if fixes were factored into the cost estimates.
A: Dr. Lipman replied that they were.

Q: Morna Conway, The Conway Group, asked if the public access system would prefer articles marked up by a journal's print composition system.
A: Dr. Lipman said that tagging done purely for print composition does not make the process easier, so a Word document from the author may be simpler. However, if a journal wants to provide fully tagged XML articles, the system will accept them.

Q: Dr. Rogawski asked about the possibility of accepting as PMC participants journals that only have articles in PDF format, as some small publishers do not have the funds to have the articles tagged in XML.
A: Dr. Lipman replied that back issue content has been accepted in PDF but at this time participating journals must have all new content tagged.

Lunch 11:45 to 12:15

III. Upcoming membership changes
Mr. Williams and Dr. Lipman recognized departing members, Rick Johnson, Sarah Thomas, and Linda Watson. Members expressed that they have enjoyed their tenure on the committee. They also stated that the rich opportunity for new discovery by utilizing the NCBI system needs to be communicated to the public. This is an example of how the proposed NIH public access policy can benefit the public in scientific information.

IV. PubMed Central Update
Dr. Lipman reported that participation agreements have been signed by the Annals of Surgery, Journal of Anatomy, Journal of Physiology and Medical History as part of the Wellcome Trust agreement. Five other journals are tentative.

PNAS has moved to author's choice open access option and Genome Research will also offer this option in January 2005. Nucleic Acids Research will move to full open access in January 2005.

PMC has continued to hold to a two-year maximum embargo period for research articles, as recommended by the advisory committee.

Automated text analysis of digitized articles has been very successful in identifying the individual elements of reference citations in the OCR text. The elements are then matched against the PubMed and PMC databases to create well-formed citations and provide links from the article to the abstracts or full-text of the references.

The underlying data structure of the PMC system is undergoing changes to conform to the NLM Journal Archiving DTD. This includes a rewrite of all PMC software. Journals are being moved to the new system in batches due to the large volume of articles. The migration to the new system will be complete early next year.

Portable PMC (pPMC) is a package of software tools that will allow a collaborating archive to maintain a working copy of the PMC archive. The software will perform a search on the remote site, send the query to the NCBI site, then render it on their site. The toolset is being built in response to interest by other countries wanting to build their own archive. The Japanese government and the Wellcome Trust in the UK have shown interest in maintaining a local copy of the PMC archive. Only content that participating PMC publishers approve for redistribution will be included in pPMC. Publishers will soon be asked to review the software and for permission to include their information for testing and public use.

A software company is interested in providing installation and outside support for pPMC. Mr. Williams asked if there is a conflict of interest with the supporting company. Dr. Lipman said that there is no exclusive agreement with them and they have been helpful in other areas such as authoring tools. The software will be available to other groups.

Ms. Watson asked if the European archive will include new content from non-PMC participants as a result of the proposed NIH policy. The public access policy envisions that there could be multiple archives for the NIH-supported manuscripts, but the details for distribution have not been determined. Mr. Williams asked which site will be up first. The UK site will most likely be up first, depending on approval by publishers. Dr. Eisen asked if third parties would be able to use the pPMC software as well. Dr. Lipman said that at a later stage with further development other groups could get the software. Mr. Johnson asked if the NIH public access decision would influence functionality. Dr. Lipman said that it would not. Ms. Joseph asked about the search function of pPMC as it is not a mirror site. Dr. Lipman replied that the PMC infrastructure is used for retrieval but capability can be improved by other groups.

V. Update on Author's Addendum

Rick Johnson provided an update on an author's addendum drawn up by SPARC. He began by explaining that SPARC has encouraged universities to create digital repositories of intellectual capital from individual institutions. A way for authors to retain rights to their articles was needed which allows them to deposit and reuse their materials in these

repositories. SPARC attempted to address this issue by creating a legal author's addendum to copyright transfer agreements when an article is accepted for publication.

A draft has been decided upon from two versions available for public comment. This version contains three key rights; (1) authors have the right to reproduce and distribute the article in any medium for noncommercial purposes, (2) the right to prepare derivative works from the article, and (3) the right to authorize others to make noncommercial use of the article so long as the author receives credit. An educational package will be built around the agreement to provide authors information on how and why to use it. Ms. Lappin added that the addendum provides a clear message of authorized uses of the material beyond agreements in other repositories.

Dr. Eisen asked about the possibility of links from PubMed to the university archives. Dr. Ostell replied that there is currently a LinkOut category for author's website, which either authors can provide or the library creating the archive. Another possibility is a new category for institutional repository.

VI. Adjournment

The PubMed Central National Advisory Committee adjourned the public meeting at 1:35 p.m.

CERTIFICATION

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

(Date)	(Date)
James Williams, Chair PubMed Central National Advisory Committee	David J. Lipman, M.D., Director, National Center for Biotechnology Information, NLM