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 advises the Director of NIH, the Director of NLM, and the Director of 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 – October 26, 2006

The meeting of the PubMed Central National Advisory Committee was convened on October 26, 2006 in the Board Room of the National Library of Medicine (NLM), Bethesda, Maryland. The meeting was open to the public. Dr. Samuel Kaplan presided as Chair.

Members Present

John Hawley, B.A., American Society for Clinical Investigation
Heather Joseph, M.A., SPARC
Samuel Kaplan, Ph.D., Houston Medical School
Robert Kiley, M.S., Wellcome Trust
Debra Lappin, J.D., B&D Consulting
Bob Roehr, B.A., Self-Employed
Mary Ryan, MLS, University of Arkansas Medical Sciences
Anthony So, M.D., Duke University
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

Janet Coleman, NCBI
Jane Davenport, IEB, NCBI
Mark Desierto, IEB, NCBI
Martha Fishel, PSD, LO, NLM
Marla Fogelman, IEB, NCBI
Al Graeff, Deputy Director, NCBI
Betsy Humphreys, Deputy Director, NLM
Laura Kelly, IEB, NCBI
Donald King, Deputy Director, Research and Education, NLM
Sheldon Kotzin, Chief, BSD, NLM
Sergey Krasnov, IEB, NCBI
David Landsman, Branch Chief, CBB, NCBI
Adeline Manohar, IEB, NCBI
Jim Ostell, Branch Chief, IEB, NCBI
Jerry Sheehan, Assistant Director Policy Development, NLM
Edwin Sequeira, IEB, NCBI
Kent Smith, NCBI
Deborah Zarin, NLM

Visitors Present
Shelley Andrews, Oxford University Press
Collette Bean, John Wiley & Sons
Laura Brockway, FASEB
Martin Frank, American Physiological Society
Ray Garant, American Chemical Society
Kristin Harrison, Elsevier
Alice Ra'anan, American Physiological Society
Mark Sobel, American Society for Investigative Pathology
Hollie Stephenson, Association of Independent Research Institutes
Crispin Taylor, American Society of Plant Biologists
Neil Thakur, NIH, OD

I. Call to Order and Opening Remarks

The meeting was called to order at 9:45 a.m. Dr. David Lipman welcomed members of the PMC Advisory Committee, and members and visitors introduced themselves. Minutes from the April 26, 2006 meeting were approved.

II. Upcoming Committee Membership Changes

Dr Lipman announced membership changes that will take effect at the next meeting, tentatively scheduled for April 19, 2007. New members are: Prudence S. Adler, Associate Executive Director, Association of Research Libraries; Sarah Michalak, University Librarian and Associate Provost for University Libraries, University of North

Carolina at Chapel Hill; Mark E. Sobel, M.D., Ph.D., Executive Officer, American Society for Investigative Pathology; Jan Velterop, Ph.D., Director of Open Access, Springer Publishing; and Gary E. Ward, Ph.D., Associate Professor of Microbiology and Molecular Genetics, University of Vermont. Dr. Lipman gave farewell plaques to four members who are leaving the committee: Heather Joseph, Debra Lappin, Bob Roehr, and Dr. Sam Kaplan. Dr. Marc Kirschner, who was not present, is also leaving the committee.

III. Current PMC Statistics

Dr. Lipman presented recent statistics on the PMC database. As of October 23, 2006, 800,000 articles were available, 60% of which were from digitized back issues. Dr. Lipman noted the importance to researchers of having these back issues available. For the month of September 2006, PMC was accessed by 1.95 million unique IP addresses. Dr. Lipman discussed the difficulty in estimating users over a month-long period based on unique IPs, explaining that it is much easier to come up with an accurate multiplier to account for multiple users of a single IP address during only a day's usage. Regardless of what measurement is employed, he said, usage is continually going up. About 7 million articles were retrieved during September, with 10 million total page views.

Mr. Roehr asked how the PMC usage figures compare to PubMed, and Dr. Dennis Benson replied that PubMed has about 5 to 10 times as many users as PMC.

IV. Update on Public Access Policy

Dr. Lipman said the number of submissions to PMC under NIH's public access policy has been stable for quite some time, with less than 5% of NIH-funded researchers submitting their articles. There have been discussions with some publishers about bulk uploading of NIH-supported articles, and those talks have progressed furthest with Elsevier. Under the arrangement with Elsevier, authors will be saved the 7-10 minute uploading process, Dr. Lipman said; however, authors will still need to interact with NCBI to confirm that the manuscript is theirs and to review the final tagged version.

Elsevier has an opt-out policy, whereby it will submit all papers resulting from NIH-funded research unless the author opts not to participate. NCBI has worried that this policy could be a nuisance for authors because they may not realize their articles are being submitted to PMC on their behalf, Dr. Lipman explained. While it is too early to make accurate projections, Dr. Lipman said, so far it appears that about 25% of those Elsevier manuscripts that have been submitted are receiving the necessary author reviews to have them progress through the system.

Dr. Lipman noted that the House of Representatives passed bill language making the public access policy mandatory with a 12-month embargo between publication and availability to the public on PMC. However, he said, it is still unclear whether there ultimately will be legislation.

Dr. Kaplan asked how NIH could make the public access policy mandatory in the absence of legislation. Dr. Lipman replied that the Offices of General Counsel for many of the agencies feel there is a strong legal basis for that type of a requirement under existing law relating to government use license. However, he noted that publishers have gotten opposing legal opinions, making the situation complicated.

V. PMC Selective Deposit Agreements

NCBI has worked out a selective deposit agreement with the American Society of Hematology for their journal, *Blood*, Dr. Lipman said. On behalf of its authors, ASH will submit to PMC those *Blood* articles that are the result of NIH-funded research. ASH will be submitting the copyedited, tagged version of articles. Dr. Lipman explained that because this is an agreement between the publisher and NLM to deposit the published article, similar to regular PMC participation, the publisher can determine the banner and other elements relating to the look of *Blood* articles in PMC.

Dr. Lipman noted that some publishers are not comfortable with having their typeset PDF files shown on PMC, and instead want the articles on their own websites. In those cases, NCBI is willing to have PMC point to the publisher's site for the article so long as it is freely available on that site. Dr. Lipman explained that because the main purpose of PMC is for archiving and to have a digital format that supports linking and other features, in the interest of moving the project forward, it is acceptable to have the PDFs on publishers' websites.

VI. PMC International Agreements

Dr. Lipman reported that the standard participation agreement for PMC has been revised. Publishers joining PMC from July 2006 onward must agree to allow their content to be made available to PMC international sites.

NCBI believes there are a variety of reasons for having PMC be part of an international network instead of being a U.S.-only archive, Dr. Lipman told the committee. One reason is that multiple live versions of the archives offer the most safety in terms of protecting the content, which represents billions of dollars of research investment. Dr. Lipman noted that NCBI's 20-plus-year collaboration with the U.K. and Japan for their sequencing databases has had tremendous benefits, such as better curation of structural content. In addition, the slightly different approaches to making the sequencing data available have stimulated new ways of using the information.

Another issue is the risk for balkanization of data as more countries begin requiring submission of government-funded research to a public archive. So far, the U.K. is the only country with a mandatory submission policy, but Canada has a draft policy, and a number of other countries, including France and China, are considering policies, Dr. Lipman said. Because countries spend so much money on research, they may feel

conflicted about sharing their results, so having collaborating international archives provides a better chance for robust, consistent archives. The best insurance for a comprehensive archive, he maintained, is to have agreements in place for sharing the data before mandated submission takes effect in more countries.

Additional advantages accrue if there is a shared format and software foundation. This reciprocity would minimize cost because the data would already be in form suitable for PMC, and it would reduce the complexity for publishers, who would not have to worry about different policies at different sites or different rendering on the sites.

Most existing PMC publishers have agreed, or are nearing agreement, about placing their data in collaborating international archives, with only a few saying they do not want to participate.

Concerns raised by those publishers opposed to PMCI center on two issues: loss of control over their content on foreign sites, and reduced visits to their websites as a result of the content being on multiple servers. Addressing the first issue, Dr. Lipman said that loss of control already exists because once an article is free on a publisher's site, anybody can access it. The publisher's protection is copyright, which would allow it to take legal action against anyone selling copies of its articles. Regarding the concern about reduced website visits, Dr. Lipman pointed out that the American Society for Cell Biology has seen use of its journals and subscriptions increase since making the content available on PMC. Also, having content on servers in foreign countries will bring in other users who might not have used a retrieval system that was not in their native language, Dr. Lipman said.

Ms. Lappin asked whether PMC needs to negotiate with each publication separately or if there is an overarching policy. Dr. Lipman replied that it has already been decided that all new PMC agreements will specify that articles can be shared with partner archives. For publishers already in PMC, however, NCBI needs to honor the existing agreements.

Mr. Roehr asked whether articles from other countries would be translated into English and whether use of translation software was possible. Dr. Lipman said that the vast majority of articles coming from France and Canada are in English. NCBI does not translate articles that are not in English.

Dr. So noted that the existing reciprocity agreements are with other industrialized countries and asked what might develop with emerging countries. Dr. Lipman reported that there are working, but nonpublic, PMC partner sites in Italy, South Africa, Japan and China. Dr. Kaplan then asked about Central and South America and India, where he said there is vibrant biology and chemistry work. Dr. Lipman replied that once the U.K. system is running smoothly and there is a track record of success, NCBI will look to expand to other locations.

VII. Reports from the Field

Ms. Mary Ryan discussed results of a two-question survey she did of 125 academic health sciences library directors in the U.S. She reported 33 responses to the first question: whether the libraries had changed policies for retention of print journals because of the availability of back issues in PMC. Fifteen libraries responded that they had changed policy to get rid of back issues once they become available on PMC. Another 15 said they had not changed their policy, primarily because they did not currently have a space problem. Three libraries said they are considering changing their policy because they anticipate a space problem in the near future.

The second question was whether the libraries want NLM to give PMC usage statistics by institution to the publishers. Of the respondents, 24, or 75%, said "no," and one said "no, no, no." Some respondents were adamant in their comments that they want the usage statistics, but that they want them from NLM and not second-hand from the publishers, Ms. Ryan said. Eight library directors responded "yes" to the question. Ms. Betsy Humphreys noted that NLM does not provide statistics by institution because it does not keep track of usage that way.

Dr. Lipman said that the question of interest is whether usage statistics by institution is critical information for libraries. Some publishers, Dr. Lipman noted, have requested this information, saying that librarians use these statistics to determine whether to renew subscriptions. Publishers have expressed concern that making articles freely available in PMC some time after publication could result in usage data that makes librarians cancel subscriptions. Dr. Lipman said that in talking to librarians, they reported that they were interested in usage of the journals during the paid period (i.e., before free availability on PMC or the journal's website) and not the later usage. That is the critical question, he said. Ms. Ryan noted that she worked with Ms. Humphreys in formulating the question, but that it was "tricky" to ask. Most of the respondents said they only look at usage statistics for things they pay for because that is how they make their collection development decisions, Ms. Ryan said. One respondent, she added, said they would like to see statistics for usage of back issues because they might decide to get a current subscription if there was sufficient interest in the older issues.

Break 10:45 – 11:00

VIII. Status of UKPMC Project

Mr. Robert Kiley gave a presentation on the Wellcome Trust's PMC deposition policy and the U.K.'s PMC project. Wellcome Trust (WT) requires researchers it funds to deposit a copy of any resulting research papers in PMC no later than 6 months after publication. Mr. Kiley noted that there are four reasons WT supports open access, the first being improved availability of the research. To evaluate access to WT-funded research, the group took a list of WT-funded articles in PubMed to two leading U.K. libraries to see how many of the papers could be obtained. One in 5 papers was in a

journal not available through the libraries, indicating that access is still an issue, Mr. Kiley said.

Another reason for open access is to integrate the research articles with genome sequences and other data. Thirdly, WT wants to know if the 450 million pounds it spends on research each year is making a difference. The fourth reason is to ensure long-term preservation of the articles.

Compliance with WT's open-access policy can be achieved in several ways. Articles can be published in open-access journals or hybrid open-access models. If the journal doesn't offer these options, the publisher must allow the author to self archive at least the peer reviewed author manuscript and make that freely available in PMC and PMC international archives within 6 months of publication. If the publisher cannot comply with any of these options, WT recommends the grantee look for another publisher.

WT offers publishers a fee for making the articles freely available in PMC upon the date of publication. WT requires that the deposit be in XML, although if this is unavailable, for a transition period, PDF's will be accepted. . The problem with the PDF, Mr. Kiley explained, is that the author will need to review the final article, and failure to do so will result in the article not being included in PMC. In brief, the WT fear that they may end up paying an OA fee, paying for the article to be converted from PDF to XML and, if the author fails to sign-off the XML conversion, the article will still not be available via PMC and UKPMC.

WT is working with the University of Nottingham on a database (see: http://www.sherpa.ac.uk/romeo.php) that will provide information on those publishers that are compliant with WT's open-access policies, Mr. Kiley reported, noting that 54% of publishers have an open-access program that is compliant. About a third of publishers, mostly societies, and many of which are American, do not currently have a policy that meets WT standards. The remaining 15% of publishers are still in negotiations with WT.

WT will require that articles be mirrored to the other PMC sites, such as U.K. PMC. The U.K. system is intended to serve the needs of the U.K. research community and therefore may eventually include some information that would not be mirrored in other PMC sites, such as U.K. clinical guidelines, Mr. Kiley said. In addition, WT wants to integrate U.K. PMC with the 7 other major U.K. grant systems. This group of 8 funders together makes up 90% of research funding and is contributing to the costs of setting up the infrastructure for U.K. PMC. Five of these groups, which include government funders such as the Medical Research Council as well as charities, already have open-access policies, and the rest are developing them.

A contract to run U.K. PMC has been awarded to the British Library, which has subcontracted the task to the University Manchester. The system is expected to "go live" in January 2007.

Dr. Kaplan asked what the policy would be if research was funded by an EU government grant as well as WT. Mr. Kiley explained that projects receiving *any* WT funding are covered by its policy, and that it would provide open-access publishing fees for the research.

WT estimates that it funds research resulting in about 4,000 papers per year, and the costs for hybrid open-access publishing fees, at about $3,000 per paper, represent about 1%-2% of WT's research budget. Mr. Kiley noted that WT is rarely the only funder of a research project and that it is hoping other funders will eventually contribute to the publishing costs.

Mr. Roehr asked what other European funders are doing. Mr. Kiley responded that the French are working with NCBI to see if they can get a system up and running. A number of funders have contacted WT, and there is discussion about U.K. PMC becoming EU-wide, he said.

Dr. Lipman reported that NCBI has met with the French public research body, INSERM, and the non-profit, private research group Institut Pasteur; the groups have started work on an archive for research they fund. There appear to be efforts toward a mandatory open-access policy in France, but the situation there is more complicated, Dr. Lipman said.

Mr. Roehr asked about the situation in the U.S. with private funding groups such as Howard Hughes. Dr. Lipman replied that NCBI has been contacted by a number of philanthropies and charitable foundations, all of whom are working on policies for requiring access to their published research.

IX. Discovery Initiative

Dr. Lipman described the Discovery Initiative to the committee, saying it is one of the most important projects NLM is working on. Explaining the rationale for the project, Dr. Lipman said that while NLM and NCBI have done a powerful job in organizing their databases so that related information is connected, many users are not taking full advantage of the available tools. Instead, many users come to NCBI's website, do a search, retrieve what they were looking for, and then leave. The aim of the Discovery Initiative is to make the most pertinent related information easily visible, using a similar approach to the techniques used by websites such as Amazon.com, where potentially desired items are consistently and conspicuously presented on the page.

As a first step, NCBI has redesigned the webpage shown when a user clicks on a hit from a search in PubMed. The new view provides a list, on the right-hand side of the page, of the five articles that are most highly related to the retrieved article. This has resulted in 25% to 30% of users retrieving related articles, compared to 3%-4% before. Dr. Lipman explained that NCBI plans to introduce various new views intended to maximize the possibility that users find relevant information, then monitor usage and fine-tune the system accordingly.

Ms. Debra Lappin commented that this concept is very important not just for NLM databases but for information science, and that there are many applications for other kinds of research.

Ms. Betsy Humphreys said that the Discovery Initiative fits well with NLM's interest in improving effective use of NLM resources and in integrating information within systems such as electronic health records.

Mr. Roehr asked whether NCBI could create a variety of algorithms for identifying relevant information that is geared to different users, such as basic researchers, clinicians, or educators/students. Dr. Lipman replied that NCBI is looking at how to identify subsets of users and to provide information that's more appropriate to them.

Ms. Mary Ryan asked if there will be links to users' libraries if the full text of an article is not available for free on PMC or the publisher's site. Dr. Lipman noted that one of the problems with the recent redesign of the PubMed search retrieval page is that the library icons went to the bottom of the page, but that NCBI will be moving the icons back to the top of the page.

Mr. John Hawley commented that he would be interested in seeing his society's journals appearing frequently in the list of five related articles on a PubMed page. Dr. Lipman replied that the odds of being listed would be improved if a journal provided its citations to NCBI, because citations are one of the factors used in determining related articles. Currently, NCBI only has access to citations for those journals participating in PubMed Central.

Mr. Hawley asked about the percentage of users who follow PubMed links to get free text from PubMed Central. Dr. Lipman said he did not have the data with him, but that the percentage was high. Dr. Kaplan noted that in the case of his journal, the number of users accessing text through PMC has been increasing three- or four-fold each year.

Dr. Anthony So asked whether Discovery Initiative enhancements would provide a way for users to follow what other users had done in a search, either in related links or bookmarks, or perhaps with comments made by experts. Dr. Lipman noted that there are a lot of scientists who would like to put notes on inquiries for themselves. The question, he said, is if there is way they could share that, maybe by emailing. NCBI will be exploring these and other options, Dr. Lipman said.

Dr. Kaplan commented that marking up text would be a superb teaching tool for students and instructors. Dr. Ostell responded that there would be no problem with having people annotate their own materials, but that annotating others' material would require some type of editorial control. Dr. Lipman added that having users email the annotated document might be a possibility or working with a scientific society that would exercise editorial oversight.

Dr. Lappin commented using federal dollars to see how people learn, how they research, and how that research can be facilitated, is very exciting and of interest to the public.

Lunch break 12:30-1:15

X. Remarks by NLM Deputy Director

Ms. Humphreys reported on the budget situation. She noted that there is a continuing resolution for FY 2007 and that a variety of outcomes are possible, but that the best that probably can be expected is a flat budget for NIH and NLM.

The committee was provided with copies of NLM's recently completed Long Range Plan. NLM and its Board of Regents engaged more than 100 experts to develop goals and recommendations for the plan. She highlighted several areas of focus in the plan: the Discovery Initiative, disaster preparedness, whole genome association studies, research and development with electronic health records, and training and recruitment.

XI. Discovery Initiative Portals

Dr. Jim Ostell reviewed changes to NCBI's retrieval software that are intended to provide more flexibility and stability. Under the current design, when a user does a search, the resulting web page is generated by one piece of code. This has the advantage of being fast. The disadvantage is that it is hard to make changes intended for one database without possibly affecting other functions.

The new system, which is expected to be operational in December, involves use of a portal server. There will be separate programs for Boolean searches, document summaries, results, and other functions. Style sheets will be used to generate page displays. The programs can all run in parallel, allowing the system to handle more work than a single program, Dr. Ostell explained.

XII. Whole Genome Association

Dr. Ostell briefed the committee on NLM's and NIH's involvement with whole genome association studies. He explained that several institutes at NIH with large studies have been getting permission to collect subjects' DNA and genotype it, with the goal of identifying genetic associations with disease. Because the data is so expensive to gather, NIH wants to make it available to as many researchers as possible, while at the same time ensuring the privacy of research subjects. NIH asked NLM to build a database to house this genotype and phenotype information.

Because the phenotype data is so diverse, NCBI's strategy in building the database has been to focus on study documentation, protocols, and questionnaires. NCBI is structuring the data into tables, creating data dictionaries, and then providing the summarized results to the submitter to review for accuracy. Questions and other document sections are linked

to the variables measured. The data are then tagged into XML, which can be viewed as HTML and indexed.

The open-access portion of the database will allow users to browse and search projects and studies, protocols, questionnaires, summary data for measured phenotypes, and genotype summaries. Users can also view pre-computed associations between phenotypes and genotypes. This part of the database will go public in December with data from a study on age-related macular degeneration and a study on Parkinson's disease. A controlled-access portion of the database, available in the spring, will provide individual-level phenotype and genotype data for those users who have obtained permission from the sponsoring institute

Dr. So asked about the implications for the "inventive threshold" of what is considered novel. Dr. Ostell replied that there has been a conscious effort to raise the bar on patentability in order to prevent blocking patents. In response to a question from Dr. So about whether there would be new whole genome association studies, Dr. Ostell said that NIH is continually developing new studies. He noted that there is a group developing a trans-NIH genome-wide policy that could require that new studies obtain patient consent for use of the resulting data in a database such as dbGaP.

XIII. Adjournment

Dr. Kaplan said there would be follow-up on a date for the next meeting of the committee as several members were not in attendance. The meeting adjourned at 2:15 p.m.

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

_____ _____

Samuel Kaplan, Ph.D., Chair (Date) David J. Lipman, M.D., Director Date