DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

DEC 16 2011

The Honorable Joseph R. Pitts
Chairman, Subcommittee on Health
House Energy and Commerce Committee
U.S. House of Representatives
Washington, DC 20515-3816

Dear Mr. Chairman:

Thank you for your interest in the National Institutes of Health (NIH) Public Access policy. The policy requires the pre-publication version of peer-reviewed journal articles that directly arise from NIH funds to be made publically available on NIH's full text archive, PubMed Central, within 12 months of publication. Facilitating public access to research results is an important way to maximize the full benefit of the public's investment in NIH-supported research.

The NIH Public Access policy is a prudent and beneficial public policy for a number of reasons. First, it applies 21st century information technology to the NIH investment to promote science and health in the context of a globally networked world of scientific information. Second, it makes NIH more transparent and accountable and better able to make strategic decisions about its portfolio. Third, it ensures more rapid scientific progress and the discovery of new treatments. Moreover, it implements the statutory mandate in the Consolidated Appropriation Act of 2008 (P.L. 110-161):

"The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publically available no later than 12 months after official date of publication: Provided, that the NIH shall implement the public access policy in a manner consistent with copyright law."

Your letter raises several questions about the policy that I will answer in turn, but first, I want to address the overall issue about the potential impact of the policy on the scientific, technical, and medical (STM) publishing industry. The NIH considered the potential impact of the policy on the publishing industry during its development and has taken steps to minimize adverse impacts.

This policy applies to an estimated 90,000 journal articles per year that arise from NIH funding. These 90,000 articles represent only a fraction of the 700,000 health and biomedical journal articles indexed each year for Medline[1] and an even smaller fraction of the estimated 1.5 million STM articles that are published each year[2].

The policy does not make papers publically available on PubMed Central until up to 12 months after publication. The embargo period initially proposed when NIH was developing a voluntary public access policy was 6 months, since scientists have clearly stated that they need the information as soon as

[1] A database maintained by the National Library of Medicine indexing over 18 million biomedical research articles. See http://www.nlm.nih.gov/pubs/factsheets/medline.html for more information.

[2] Ware, M; Maebe, M. (2009) The STM Report: An overview of scientific and scholarly journal publishing. STM: International Association of Scientific, Technical and Medical Publishers. www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf.

possible for their research. Since the publishing industry asked for a longer delay, a 12-month embargo was provided as a concession.

Further, the NIH Public Access Policy applies only to final, peer-reviewed manuscripts arising from NIH funds, not the final published version. The policy asks NIH-funded authors to submit their manuscripts directly to the NIH. We have been pleased to find that hundreds of publishers have volunteered to assist their authors in supporting the policy. Thousands of journals voluntarily submit *peer-reviewed author manuscripts* to PubMed Central to assist authors in complying with the Public Access process.[3] Several hundred journal publishers also voluntarily automatically deposit *final published versions of articles* on behalf of their authors, relieving them of the need to directly submit their manuscripts. Finally, publishers representing about 1000 journals voluntarily submit the full content of their journals to PMC, regardless of whether the issue contains an article subject to the NIH Public Access policy.[4]

There is no evidence that the NIH Public Access policy has harmed the publishing industry. While the U.S. economy has suffered a significant downturn in the past several years, the STM publishing industry appears strong, with increases in both the number and price of STM journals. For example, from 2007 to 2011, the number of biological sciences and agriculture journals and medicine and health journals grew by 15 percent and 19 percent, respectively. In the same time span, the average prices of biology journals and health sciences journals have increased 26 percent and 23 percent, respectively. (See Appendix 1 for full data and citations.)

With this short background, we can briefly address your questions. However, a full discussion with your staff would be helpful to fully elaborate on these points.

Question #1: PubMed Central is accelerating scientific discovery.

NIH has asserted in congressional testimony that PMC is "accelerating scientific discovery in the biomedical sciences." What evidence does NIH have to support this claim? Does NIH have any data on the benefits derived from manuscripts deposited under the mandatory Public Access Policy, such as research productivity by scientists who otherwise would not have had access to publishers' contents?

The value of access to the peer-reviewed scientific literature that is part of the Public Access policy is evidenced by the significant usage of PubMed Central (PMC), the full text archive that disseminates the papers deposited under the Public Access policy. PMC predates the Public Access policy by many years and is currently being accessed by more than 500,000 users per weekday who retrieve approximately 1 million articles. The number of articles retrieved has doubled in the past three years.

PMC is not only a repository for journal articles in the biosciences but is part of a larger information infrastructure that is accelerating scientific discovery. PMC resides at the National Center for Biotechnology Information (NCBI), established at the National Library of Medicine (NLM) by Congress in 1988 to create a national resource for molecular biology information in response to the fragmented state of information resources in the molecular biology and biotechnology fields. Today, NCBI is the home of more than 40 free online databases, such as Genbank, the database of all publicly-available DNA sequences, which are accessed by more than 2 million people per day. It is the integration of all these resources, not merely the posting of articles that fuels scientific discoveries.

---

[3] These publishers are listed under Method D at http://publicaccess.nih.gov/select_deposit_publishers.htm.
[4] See http://publicaccess.nih.gov/submit_process_journals.htm for a full list.

The NCBI system, of which PMC is a part, provides researchers with the capability to integrate their individual discoveries with other publications and scientific data that lead to new areas for exploration. As former NIH Director Dr. Elias Zerhouni testified in 2008 before a House Judiciary Subcommittee:

> *"When viewing a report in NIH PubMed and PubMed Central databases, at the touch of a button we can link to papers that are determined to be related, as well as to papers that were actually cited. We can also link to related chemical structures, proteins, viruses, and other data, allowing us to make discoveries that advance science and even prevent deaths."*

This system is similar to what happens when a site such as Amazon.com gives you suggestions about products related to your initial search that might interest you. It does more than help you find the specific item you were looking for; it provides a context to help you make a better decision. The NCBI family of databases is doing this with a broad array of scientific data with complex connections.

Question 2: Academic institutions and drug and device companies lack access to the literature.

NIH has stated that "large academic institutions and drug and device companies can lack access" to journals. Do you have market survey data that shows who amongst these institutions are not getting access and to what they are not getting access?

The limitations of accessing full-text STM literature is simple to demonstrate. One could search any biomedical disease area or research question using PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) or other search engine and determine the percentage of papers that can be accessed without charge.

However, there are several surveys documenting access challenges in the academic setting, which reinforce the many anecdotes and complaints that we hear. One publisher survey found 15 percent of USA and Canadian scholars from all disciplines reported their level of access to the journal literature "varies" or was "poor or very poor"[5]. Another publisher survey asked if scientists have access to the articles they need; 32 percent of immunologists and microbiologists and 39 percent of other researchers responded in the negative (from "I sometimes have difficulty" to "I always have great difficulty")[6]. A 2010 survey of clinicians at an American community health clinic and independent teaching hospital found 32 percent reported feeling under-served in their current access to health information. Fifty-six percent of the sample reported having access to "most [*not all*] of the research articles I need"[7]. Finally, when the University of California surveyed its own faculty in 2007, before the NIH policy became a requirement, it found 29 percent of its life science and medical faculty felt high journal prices made it difficult for them to access the literature they need[8].

---

[5] Ware, M. & Monkman, M. (2008) Peer review in scholarly journals: Perspective of the scholarly community - an international survey. A Publishing Research Consortium report. http://www.publishingresearch.net/PeerReview.htm.

[6] Rowlands, I. and Olivieri, R. (2006). Journals and scientific productivity: a case study in immunology and microbiology. Publishing Research Consortium. http://www.publishingresearch.net/journals_scientific.htm

[7] O'Keeffe JK, Willinsky J, Maggio L (2011). Public Access and Use of Health Research: An Exploratory Study of the National Institutes of Health (NIH) Public Access Policy Using Interviews and Surveys of Health Personnel. Journal of Medical Internet Research. Vol 13, No 4.

[8] Here are the results for all disciplines from the 2007 survey. Responses to "High journal prices have made it difficult for me to access the literature I need."

	Arts	Humanities	Life & Med Sci	Physical Sci	Social Sci	Overall
Number of Respondents	45	213	259	255	342	1114
Agree	44%	38%	29%	20%	28%	29%
Disagree	47%	51%	65%	72%	65%	63%
Not Sure	9%	11%	6%	8%	7%	8%
Total	100%	100%	100%	100%	100%	100%

http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf

We also hear many disturbing stories about how lack of access to the scientific literature is problematic to the small business community, particularly our small business awardees and those in high technology fields like drug development and medical device development. A recent letter to *Nature Biotechnology* stated, "the biotech companies that employ many of our most highly trained and educated scientists can't afford to provide them access to the science journals that they need to do their jobs effectively. Innovative startup companies in many fields will always be financially constrained compared with their well-established brethren with whom they wish to compete. However, biotech companies are dependent on library access to a degree not seen in other technology areas, such as software."[9] We find this lack of access disturbing, because the NIH relies on the scholarly literature to disseminate its research results. If scientists and innovators cannot read those papers, the return on the NIH research investment is constrained.

When the literature is accessible, our investment is more effective. For example, better access to medical research makes our health care system more efficient. Over six out of 10 physicians report changing an initial diagnosis based on new information accessed via online resources/support tools. Nearly nine in 10 physicians feel that improved access to online medical information and resources has improved the quality of care at their practice[10].

Question 3: The Public Access Policy promotes American competitiveness.

Concerns have been raised that PMC is undermining American competitiveness by enabling free access to the product of U.S. research and exposing U.S. journal articles to foreign piracy. Has NIH taken steps to protect publishers' from their revenue being eroded by piracy? What steps has NIH taken to ensure that publishers' interests are respected by NIH staff and PMC users? Has NIH recently analyzed usage by non-U.S. citizens and organizations, which, by NIH's earlier report, accounts for two-thirds of total usage?

PMC makes access to NIH-supported papers, as well as almost two million other papers, many of them supported by foreign government funding and/or published in non-US journals, easier both for Americans and people in other nations. Unless publishers have themselves chosen PMC as the sole distributor of the electronic contents of their journals, all the articles that are publicly available in PMC are already accessible in electronic form outside the United States (via subscription-based Internet access and from libraries) at the time they become available in PMC.

Regarding concerns about piracy, NCBI uses processes to detect and prevent potential problems that are very similar to the processes used by publishers on their own sites when they make their content available internationally. NCBI has established sophisticated monitoring systems to protect publisher interests and prevent piracy. These systems detect and prevent bulk downloading and will immediately cut off any sites, foreign or domestic, that abuse copyrighted property.

As noted in the question, approximately two-thirds of the total PMC usage reported in an earlier report was non-U.S. Current analyses indicate that proportion remains the same. Of course, a large percentage of the content of PMC also comes from non-U.S. journals. In addition, Government and private institutions in other countries have helped to support the addition of content to PMC. In one notable case,

---

[9] Lyman S. Industry access to the literature. *Nat Biotechnol.* 2011 Jul 11;29(7):571-2. doi: 10.1038/nbt.1909

[10] www.wolterskluwerhealth.com/News/Documents/White Papers/Wolters Kluwer Health Survey Executive Summary-Media.pdf

organizations in the United Kingdom have provided substantial funding for digitization (with publisher permission) of the complete sets of pre-electronic volumes of selected journals that are now deposited in PMC.

In addition to making NIH-supported research available to American citizens and businesses, PMC offers strategic advantages to the NIH that have larger implications for American competitiveness. Science is a global endeavor, and because of increasing foreign research funding and flat funding for the NIH, the NIH is becoming an increasingly smaller funder of biomedical research relative to our competitors. Access to NIH-supported research through PubMed Central helps all scientists make scientific advances in areas that the NIH has set as a priority. The American people, businesses and healthcare system can then reap the benefits of those discoveries, and the NIH gains a larger return on its research investment.

Question 4: PubMed Central operating costs.

Questions have been raised about NIH's cost estimates to operate PMC - including staff time on PMC and on travel coordinating international mirror sites, related infrastructure spending, acquisition and input of articles and inter- or intra-agency contracts and transfers such as payments from NIH Institutes or Centers to National Library of Medicine) and National Center for Biotechnology Information (NCBI). What are these costs and what proportion of these costs are for articles that fall under the NIH Public Access Policy?

PubMed Central (PMC) has been in operation since 2000. It serves as the digital counterpart to the NLM's print journal collection and supports the NLM's legislative mandate to collect and preserve the biomedical literature. The base cost of operating PMC is independent of the cost of implementing the Public Access policy and is part of the general infrastructure costs within NCBI for supporting all of its 40-plus databases.

The NIH Public Access policy makes NIH-supported scientific papers publically available on PMC. Keeping in mind that PMC is a permanent digital archive, it has been designed so that the ongoing cost of storing and retrieving additional articles is negligible once the article is converted to PMC's archival format. The NIH Public Access policy has been responsible for only a small fraction of the papers in PMC. Therefore, the cost at NCBI for implementing the NIH Public Access Policy, $4.0 million in FY2011, is effectively what it costs to collect, process, and convert NIH-funded manuscripts to the PMC archival format.

NCBI staff has not travelled internationally in order to set up or manage the operation of PMC International sites. With the current state of technology and the PMC software that NCBI's programmers have created, they are able to do everything that is necessary via email or by a secure Internet link from NCBI to the International Center's system. In the few instances where face to face meetings have been useful, primarily at project start up, staff from the foreign centers has travelled to the NIH at their own expense.

Question 5: Publishing costs.

Although publishers do not pay peer reviewers, they add value to manuscripts through the significant cost of organizing and managing reviewers, and financial support of the peer review process. For those grants for which NIH does cover peer-review, what is the full aggregate cost for NIH, on an annual basis? Your predecessor has testified that the annual cost is $80-$ 100 million. Are these costs deducted directly from grant funds, or do they come from other operating lines?

Peer review of scholarly articles is not a paid activity; it is an uncompensated activity that researchers perform as part of their professional responsibility to the scientific community. The NIH does not, therefore, pay for peer review. It does, however, pay for the research reported in these papers, as well as directly pay the salary of scientists (through grants) while they write papers that arise from NIH funds.

The NIH also allows researchers to use grant funding to pay other charges associated with publishing papers arising from NIH funds. These charges may include "page charges" that cover layout, graphics and other costs of publication, as well as "open access fees" that some journals charge in exchange for making papers freely accessible to all readers upon publication. These fees typically run from $900 to $5000 per paper. Former NIH Director Zerhouni estimated that the NIH reimburses roughly $100 million of these fees per year out of NIH awards. Because the NIH has moved to a modular budget reporting system many years ago to reduce administrative burden on our awardees, we do not have a precise way of estimating these expenses. We do not have a separate set of funds set aside for publication expenses.

Question 6: Set aside funds for publication costs.

NIH has asserted that 'The success of the NIH model has stimulated similar efforts in other countries." Yet, the efforts of these other funding bodies are quite different. Funding agencies such as the Wellcome Trust and Medical Research Council in the UK have developed arrangements for publishers to recoup their investment by paying a fee to sponsor access to articles while other funders instruct authors to comply with publisher policies. Is the NIH considering adopting similar mechanisms and how does the NIH justify its approach in contrast to these other approaches that fund open access and do not threaten jobs in STM publishing?

The NIH has considered a separate fund for publication expenses and avoided it because of the potential for the NIH to distort prices in the publishing market. As noted in the response to question five above, the NIH does permit funded researchers to use grant funds to pay for costs associated with publication, including open access charges. The NIH does not require funded researchers to publish in open access journals; instead it allows researchers the freedom to publish in a journal of their choice.

The NIH places no conditions on the terms of use for a paper when it reimburses authors for publication costs with grant funds, beyond those of the NIH Public Access policy. The NIH requires only that the peer-reviewed, pre-publication version of the work be made accessible in PMC. Readers of this version, the final, peer-reviewed manuscript, must comply with traditional copyright restrictions. In addition, the publisher typically retains rights to the final published version of the article.

Where organizations such as the Wellcome Trust have developed arrangements to pay open access fees to publishers, they receive more in return from publishers than the NIH requires. They obtain extensive non-commercial rights to the final published version of the articles, under which any reader may freely redistribute or reproduce the content, or use it to create derivative works.

Question 7: Lay audience communication efforts of NIH and publishers.

NIH has asserted that "[w]ithout a resource like PubMed Central, the general public does not have ready access to much of the biomedical literature[.]" Why does NIH not take into consideration its own public communications efforts (e.g., through individual institute web sites), as well as publisher initiatives that make articles available to patients, their families, and the public at no or low cost; or consider working with publishers and patient organizations to support programs such as patient INFORM that make articles and interpretive material specially created for a lay audience available to the American public?

The NIH recognizes that its own Institutes and Centers and certain publishers have taken steps to make published articles and other interpretive material available to patients and their families. We view these programs as complements to, rather than substitutes for, PubMed Central and the NIH Public Access policy.

These other services do not systematically ensure that all articles resulting from NIH-funded research are available to the general public; instead they tend to make available only a select set of articles via the websites of participating institutions. In addition, they do not provide as convenient a means as PubMed Central for accessing articles in a centralized repository with sophisticated search algorithms and no advertising. Patient INFORM (http://www.patientinform.org/finding-medical-research/), for example, provides interpretive material via the websites of participating organizations, such as the American Diabetes Association, but not all material and not all participating organizations link to published journal articles. As a result, access can be disjointed and confusing for users.

The NIH has also worked closely with its own Institutes and Centers and trusted health organizations to provide interpretive information to patients and other consumers. The NLM's MedlinePlus, for example, provides consumers with access to reliable health information from NIH Institutes and Centers, Government health agencies, and other trusted health organizations, including medical societies, disease and patient advocacy groups. It had more than 40 million unique visitors in the final quarter of FY2011. The NIH's *MedlinePlus* magazine, launched in 2006, provides another channel for communicating recent research advances to lay audiences. With encouragement from Congress, the NLM has continued to expand distribution of the magazine via collaboration with a number of leading patient organizations, including the National Hispanic Medical Association, the American Diabetes Association, and the Peripheral Disease Association, among others.

These types of consumer information services, which focus on the lay public, are no substitute for the NIH Public Access policy and PubMed Central. PubMed Central makes published journal literature accessible to a much broader swath of the public, including clinicians, students and educators, and entrepreneurs in small business, thereby contributing to education, patient care, and innovation, as well as informing the lay public.

I am always happy to discuss this important issue with you at your convenience. In the meantime, my staff will arrange a briefing for your staff on this topic. Thank you again for your interest in the NIH Public Access policy.

Sincerely yours,

Francis S. Collins, M.D., Ph.D.
Director

Enclosure

cc: The Honorable Fred Upton, Chairman
The Honorable Henry A. Waxman, Ranking Member
The Honorable Frank Pallone, Jr., Ranking Member, Subcommittee on Health

Appendix 1: Publishing industry statistics

Average Journal Prices Grouped by Library of Congress Subject, from *Library Journal* Periodicals Price Surveys, 2009-2011

Subject	Average Cost Per Title 2007 [1]	Average Cost Per Title 2008 [2]	Average Cost Per Title 2009 [2]	Average Cost Per Title 2010 [2]	Average Cost Per Title 2011 [3]	Price change since 2007
Agriculture	$948	$1,038	$1,114	$1,178	$1,103	16%
Anthropology	443	400	435	450	363	-18%
Art & Architecture	223	254	273	286	276	24%
Astronomy	1,516	1,680	1,803	1,921	2,008	32%
Biology	1,720	1,806	1,944	2,035	2,167	26%
Botany	1,356	1,502	1,622	1,695	1,731	28%
Business & Economics	750	883	942	986	982	31%
Chemistry	3,241	3,420	3,647	3,792	4,044	25%
Education	492	567	619	653	556	13%
Engineering	1,753	1,961	2,129	2,242	2,035	16%
Food Science	1,180	1,327	1,442	1,530	1,564	33%
General Science	1,006	1,122	1,222	1,287	1,333	33%
General Works	154	176	185	194	241	56%
Geography	1,001	1,093	1,148	1,232	1,155	15%
Geology	1,397	1,529	1,607	1,685	1,791	28%
Health Sciences	1,193	1,304	1,413	1,486	1,470	23%
History	217	256	277	295	266	23%
Language & Literature	199	233	249	275	269	35%
Law	273	294	325	338	460	68%
Library & Information Science	454	513	557	587	616	36%
Math & Computer Science	1,305	1,397	1,480	1,541	1,593	22%
Military & Naval Science	696	639	682	713	916	32%
Music	149	180	189	197	249	67%
Philosophy & Religion	241	261	276	300	328	36%
Physics	2,928	3,094	3,248	3,368	3,499	20%
Political Science	445	527	572	612	622	40%
Psychology	525	590	649	691	737	40%
Recreation	257	324	418	442	402	56%
Sociology	517	597	647	692	659	27%
Technology	1,682	3,180	1,458	1,531	1,374	-18%
Zoology	1,236	1,362	1,479	1,532	1,647	33%

[1] 2009 Library Journal Periodicals Price Survey. http://www.libraryjournal.com/article/CA6651248.html
[2] 2010 Library Journal Periodicals Price Survey. http://www.libraryjournal.com/article/CA6725256.html
[3] 2011 Library Journal Periodicals Price Survey. http://www.libraryjournal.com/lj/ljinprintcurrentissue/890009-403/periodicals_price_survey_2011_.html.csp

Shaded rows are the topic areas most likely to contain papers that fall under the NIH Public Access Policy.

Number of active peer reviewed journals, 2007-2011, from UlrichsWeb Global Serials Directory[11]

Active Journals, Refereed/Peer reviewed	2007	2008	2009	2010	2011	Percent Change 2007 to 2011
Biological Sciences and Agriculture	3,761	3,930	4,068	4,237	4,325	15%
Chemistry	1,323	1,377	1,425	1,470	1,508	14%
Medicine and Health	7,781	8,159	8,512	8,917	9,235	19%
All science areas	17,497	18,285	19,029	19,905	20,538	17%

[11] http://ulrichsweb.serialssolutions.com/. Data accessed 11/18/11. Therefore, 2011 data may be undercounted.