

Dear National Science and Technology Council's Task Force on Public Access to Scholarly Publications,
As a medical librarian at a public institution, here are my comments on your RFI for Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research.

- 1) Agencies should make all federally funded research publications mandatorily available to the public. The amount of information published makes it almost impossible to find the information that you need in order to conduct sound research. Because of commercial ownership on published articles, computation analysis to help “get to” the most appropriate information is almost impossible because programmers do not have a corpus of literature from which to text mine effectively, thus impeding scientific discovery. The cost of storage of these digital articles may be a barrier but is minimal in comparison to storing the great amounts of raw data that may be generated from activity such as gene sequence analysis or astronomy tools. In the least, a comparable database like PubMed Central that incorporates all of the subject disciplines with full-text articles would be extremely beneficial in more efficient and rapid scientific discovery and analysis. NIH has been a good role model for the biomedical sciences, but research in incredibly interdisciplinary and for true transformational research to occur, you cannot silo information into specific subject areas. Having a single source of federally funded research publications makes the most scientific sense.
- 2) I can appreciate the commercial need to create and sustain journals, however, the research that is being published has already been funded by tax payer dollars. An embargo period can continue to benefit private publishers without hindering the progress of good science. As a librarian, I am particularly angered by the incredible profit margins of these publishers (we have seen percent increases in a small society published journal that got acquired by a commercial publisher as high as 276%). Information is the single most empowering commodity in the world, and people have a right to it.
- 3) Pros – Data mining across literature of all subject disciplines will help to progress scientific discovery from small lab environments into commercial and public good. A single individual cannot retain all of the knowledge published that would necessarily lead them to new discovery. Text mining is an obvious immediate benefit of an open source, full-text, scholarly publications database. If content is distributed, then there would in the least need to be standards for metadata, archiving, back up and disaster recovery that the government would set. If the government pursues a distributed path, then I think it would be necessary to have “accepted” third-party providers so that the standards could be monitored and enforced.
- 4) Nature Genetics, a journal from the Nature Publishing Group, has standards for distributed storage of scientific data that could be looked at. The University of Michigan has several projects that demonstrate commercial publisher cooperation with accessibility of published materials such as their Deep Blue institutional repository and the Hathi Trust, a cooperative digital works repository.
- 5) Metadata standards such as those for Medline (PubMed) are a good starting point and requiring that all future requests for federal funding be tied into publications that meet these standards (such as the PMCID requirement by NIH) are a good way of ensuring compliance.
- 6) Tying compliance into the existing award structure for federal funds would be the most efficient method and probably least disruptive method for ensuring open standards. Publishers still have many options of providing the scholarly articles in a manner that can still be profitable but still do not impede the access to the information – more quickly (during the embargo period), via mobile applications (different formats that users may want and can afford to do so)
- 7) Ideally, all information that is federally funded should be made available since the research was funded by the public. However, the published, peer-reviewed article is a good starting point.

- 8) I do not have good evidence to support data for a specific embargo period. Existing citation metrics typically need a one year period in which to gather enough citation data to develop metrics such as journal impact factor, etc. This type of information may support a one year embargo period.

Jean Song
Research and Informatics Librarian
Taubman Health Sciences Library
University of Michigan