Subject: **Request for Information: Public Access to Peer-Reviewed Scholarly ,Publications**

Date: January 2, 2012 9:23:22 PM EST

To whom it may concern

I am a professor of computer science at MIT who has published over 150 articles in peer-reviewed publications. For several years I served as "electronic publications chair" for SIGACT, the theory group of the Association for Computing Machinery.   I do research on information retrieval and management including user interfaces and the Semantic Web.
  As co-PI of the SIMILE project (http://simile.mit.edu/) I worked with MIT libraries developing tools for long-term archiving and management of digital artifacts.   The Library of Congress is currently funding us to develop these tools further
(http://www.simile-widgets.org/exhibit3/Exhibit3%20press%20release.pdf). I work hard to provide open access to (draft versions of) all my publications (http://people.csail.mit.edu/karger/papers.html) .

I strongly support legislation requiring open access publication of federally funded research. As a scientist who often receives federal funding and who publishes frequently, I have a significant stake in this question and am eager for the opportunity to give my work away. The major obstacle to doing so is the publishers, both commercial and nonprofit, who wish to profit by gate-keeping access to our work.

The key point I wish to make is that the most effective action for the government to take is small and targeted: it should simply require that all recipients of federally funded research provide open access to (the camera ready version of) all of their publications, using whatever common publication format is used in their own domain (I expect that for many domains this would be pdf format).   The government should _not_ be diverted from this step by wrestling over relatively unimportant question like standardizing around specific machine-readable citation standards or designing and building a central repository of published articles. The forces that will accomplish these refinements already exist, barred from acting only by the refusal of publishers to make their peer-reviewed content available for access.

Addressing question 1: I do not believe that government can or should grow existing or new markets. If the government simply forces open access, these markets will emerge on their own without encouragement.
It is worth noticing the number of tools, such as Citeseer, Google Scholar, and Microsoft Academic search that have emerged _despite_ the lack of open access. These tools currently make do with material that authors publish

voluntarily online. If peer-reviewed publications were forced open, these tools would get substantial better and more useful, and other tools would emerge to complete with them.

Addressing question 2: I do not believe that publishers should be considered to hold any "intellectual property interests". They are publishing the work of scientists, which has been reviewed by other scientists. These days, publishers rarely even provide editing services, instead publishing whatever "camera ready copy" is provided by the authors.   The other stakeholders all accrue benefit through open access publication. The federal government, which funds the research,_does_ have intellectual property rights, and that they should be able to exercise those right to demand open publication of the work they have funded.

Addressing question 3: I believe that the weight of history favors decentralized approaches combined with shared standards. I believe that the ideal legislation would simply require every federally funded scientist to publish their work (i) at a web accessible location and (ii) with bibliographic information in some machine readable format. If this is implemented, a variety of existing and future services such as citeseer, google scholar, microsoft academic search, and DBLP can compete to offer the best aggregation services. As for archiving, this is something which our numerous (decentralized) university libraries have been doing very well, and are well suited to continue doing.

Addressing question 4: I do not believe that new public-private models are necessary, as an existing public-private model, the university, has substantial experience in archiving and dissemination of material. Just as they have served as archives for paper publication, they have been developing substantial expertise in archiving of digital publication. If they can get access to that digital content, they will do a great job archiving and disseminating it.

Addressing question 5: I encourage providing the research community sufficient flexibility to innovate on this question over time, instead of enforcing a standard at the beginning.   It is unlikely that every discipline will benefit from the same standards, so undesirable to impose one.   Simply requiring that the publications themselves be openly accessible will be enough to break the publishers locked gate; innovation will follow naturally. For example, the Dublin Core has emerged as a de-factor standard for describing basic publication data; numerous projects make use of it and numerous tools produce it.
Requiring use of Dublin Core metadata would not be too onerous, but I also feel it would not be necessary. Aggregation tools like those I described above have worked quite successfully with the raw documents.

Addressing question 6: I work in a field that is eager to disseminate information.   A requirement to provide open access would be no burden at

all. Instead it would remove the current burden, of simultaneously managing a "locked" copy on behalf of the publisher and a "draft" that we try to make publicly available.   The simplest possible requirement would be to provide, with any application for federal funding, a complete list of (publicly accessible) URLs of articles that were the product of federal funding.   The federal government could turn around and publish such lists, unaltered, where they become the raw material for the aggregation services mentioned above that need to find the articles. In general, my list will simply grow over time as I publish additional work; occasionally, I will need to revise the list to reflect moved documents. An alternative would be to require the submission of just a single link, where a document could be found listing the given publications.

Addressing question 7: In my field, the vast majority of cutting edge research appears as conference publications, with journals trailing years behind if at all. Thus, it is essentially to include conference proceedings in the publication requirement---they are the de facto journals for computer science.

Addressing question 8: I believe any postpublication embargo period is damaging.   There may be value to researchers in delaying publication of their results while they exploit them for private gain. But researchers already have the opportunity to "embargo" their research for as long as they wish simply by delaying publication. I can think of no case where researchers need to publish their work in order to accrue the private benefit. Note that when they publish in a peer-reviewed journal, they are immediately making the work available to a large population of scientists with access to the publication. Given that the work is federally funded, I see no justification for the public at large to be prevented from accessing the work at the same time.   Work in my field also moves exceedingly fast: work published in a conference is often obsoleted by follow-on work published in a conference 3 months later. Thus, an embargo of even a month could noticeably impact the pace of research.

Thanks for you attention
David Karger
Professor, EECS
MIT Computer Science and AI Laboratory