Professor Victoria Stodden
Department of Statistics
Columbia University
New York, NY

http://stodden.net

(1a) Are there steps that agencies could take to grow existing and new markets related to the access and analysis of peer-reviewed publications that result from federally funded scientific research?

Increased funding could be made available for grants for text-oriented scientific research. Even things like research metrics could be developed. We don't have open ways to evaluate citation and the importance of published paper, and we could also do this for scientific data and code reuse. But the key point is to make the scientific literature available to the public for reuse, by making the publications openly available in a centralized location on the web.

Just permitting the open access to the scientific literature cannot but increase the use and potentially commercialization applications that can be generated from the newly available knowledge. The National Institutes for Health, for example, is making its articles published since 2009, and one can only image the number of potentially commercialization ideas and discoveries in the engineering literature just to take one example.

(1b) How can policies for archiving publications and making them publically accessible be used to grow the economy and improve the productivity of the scientific enterprise?

It is clear that making more information about scientific discoveries available cannot reduce economic growth. But we can be sure of more than that. Embedded in the literature are many scientific discoveries that could be commercialized and developed by industry, that are either currently unavailable or available via patent licensing fees paid through the institution's Technology Transfer Office. With the onset of the Internet there is no need for such a convoluted way of making scientific knowledge available as the public good it is.

I believe the productivity of the scientific enterprise will increase, not only because of the increased ability to carry out metrics using a corpus of scientific publications, but because all the results will be open to wider scrutiny. With the concurrent movement toward reproducible computational science – making the data and code that created the published result conveniently available – there is a very power change in the ability of the broader public to understand the scientific discoveries being made.

(1c) What are the relative costs and benefits of such policies?

There are no costs, except publisher fees for some subscription journals, which could impact publisher profits. The publication model we have today grew from a pre-digital age when it was costly to print paper journals. Today, the writing, typesetting, and review – nearly all that is necessary for publication – is done by scientists on a volunteer basis. Furthermore, the support for such a publishing system is also archaic, heavily subsidized and influenced by federal funds.

We don't have a free market in academic publishing. Because of the integrated nature of federal funding for the research that is published, as well as the time for review, and for page fees and other publishing costs comes typically from federal funds. The argument is similar for scientific societies that publish subscription journals.
It is not clear what benefit they add for the millions and billions in subscription costs, outside of sorting discoveries by journal prestige.

Because the journal publishing is in such a federally controlled market it is appropriate to rectify market failures due to the advance of technology. Unlike in a free market where firms would have an incentive to adapt and reduce costs, the publishers have none. Because of the difficulty for scientists to publish outside the most prestigious journals, federal action is required to move to a more optimal place, from society's perspective: open access to scientific knowledge.

(1d) What type of access to these publications is required to maximize
U.S. economic growth and improve the productivity of the American scientific enterprise?

Without constraint, publicly availability of scientific publications, from their publication date. I believe the publishing industry must make the argument that sequestering scientific knowledge increases economic growth or scientific productivity – as a scientist I do not see how any argument exists other than protecting the publishing industry's bottom line, and that cannot be an interest of public policy, but rather creating the environment for open scientific knowledge.

(2a) What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders involved with the publication and dissemination of peer-reviewed scholarly publications resulting from federally funded scientific research?

We must remember scientific discoveries, a public good, are at the core of this discussion. How they are best transmitted to the public and built upon is with as much exposure and access as possible. To this end the federal agencies should enact a PubMed Central similar to the NIH's for published documents. Make these documents available under an attribution only license, for example CC-BY. Enforce deposit by making grant money contingent and have the deposit occur within 6 months of publication.

Establish federal guidelines recommending scientists keep copyright rather than sign it over the publishing house, who did almost zero in the production of the document and only hold the prestige of the journal as an inducement to publish.

(2b) Conversely, are there policies that should not be adopted with respect to public access to peer-reviewed scholarly publications so as not to undermine any intellectual property rights of publishers, scientists, Federal agencies, and other stakeholders?

Intellectual property rights should not be established with publishing houses. This is the root of the open access problems today. If publishing houses and academic societies could be trusted to act in the best interests of society and not in their own best interest, we would have open access already. Federal policies must rectify this accident left over from the pre-digital age.

(3) What are the pros and cons of centralized and decentralized approaches to managing public access to peer-reviewed scholarly publications that result from federally funded research in terms of

interoperability, search, development of analytic tools, and other scientific and commercial opportunities?

The federal funding agencies have the most power to break the collection action lock preventing open access and should also be the central archivist for the published articles. Any decentralized archives are productive in tandem in that they provide backup and reduce the load on the main repository (provided the same versions of the papers are deposited in each). A central repository facilitates research on the corpus of text, encourages interdisciplinary research and breaks barriers to cross disciplinary communication (you don't have to know in advance which repository will house the information you are interested in) and will last. Links to papers will persist over time. This is the main reason why it is important for our scientific knowledge to be housed by a federal service, like the Library of Congress does for other material. Having private or even institutional repositories as the sole guardian of our scientific culture subjects it to potential loss when they go out of business or decide that such an effort is no longer aligned with their interests.
For example, researchdatasets.google.com was dissolved with about a month's notice after Google had established the site to warehouse scientific data there. Preserving our scientific culture is a federal government task.

(4) Are there models or new ideas for public-private partnerships that take advantage of existing publisher archives and encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research?

I don't believe the publishers are interested in budging one inch toward open access and I believe the federal government will have to take steps to ensure these corpora are made open. It will be easies to do this for paper not yet published, but as the economic value (through access fees) declines over time publishers may be able to give up older papers, say older than a year, with no appreciable loss to profits.

(5) What steps can be taken by Federal agencies, publishers, and/or scholarly and professional societies to encourage interoperable search, discovery, and analysis capacity across disciplines and archives? What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities? How should Federal agencies make certain that such minimum core metadata associated with peer-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to Federal science funding?

I believe the most important first step is making the published scientific articles available in a repository. After that it will be easier to see what is missing in search and how to best rectify these problems. I would not try to establish standards before open access is established. It will create unnecessary barriers to establishing open access and the inevitable fact is that technologies will change and so will our search, tagging, location, classification, and retrieval abilities. The only thing that must be established is a DOI must be placed on each deposited paper for version control and citation reasons.

(6) How can Federal agencies that fund science maximize the benefit of public access policies to U.S. taxpayers, and their investment in the peer-reviewed literature, while minimizing burden and costs for stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?

There is almost no burden to submitted a published article to an open access repository for scientists. There will be an expense in establishing such federal repositories which is unavoidable, but these textual corpora are small compared to, say, scientific data.
Publishers must also act in the public interest, since this is scientific publication and should not be considered a money making venture. Making profits by controlling scientific communication is bordering on unethical and certainly repugnant.

(7) Besides scholarly journal articles, should other types of peer-reviewed publications resulting from federally funded research, such as book chapters and conference proceedings, be covered by these public access policies?

I believe the more of our scientific culture that is made publicly available the better, with exceptions for national security or confidentiality interests (which should be taken on a case by case basis).

(8) What is the appropriate embargo period after publication before the public is granted free access to the full content of peer-reviewed scholarly publications resulting from federally funded research? Please describe the empirical basis for the recommended embargo period. Analyses that weigh public and private benefits and account for external market factors, such as competition, price changes, library budgets, and other factors, will be particularly useful. Are there evidence-based arguments that can be made that the delay period should be different for specific disciplines or types of publications?

Zero days is the appropriate embargo period. Scientific knowledge is a public good. This is why there are massive federal subsidies for scientific research in the first place.

I believe scientists will not suffer without an embargo period, since it is exceedingly difficult to scoop an authors next publication on his or her data. Evidence shows that the typical result is that research from a different field use the data in ways the original author did not anticipate and the original author gains citations with no loss of publications.

(9) Please identify any other items the Task Force might consider for Federal policies related to public access to peer-reviewed scholarly publications resulting from federally supported research.

With open access to journal articles, the task force may do well to coordinate with the data policies task force. The reason is because shared data must be versioned and linked to its published articles, which presumably is in the open access federal repository.

Thank you for creating this RFI and listening to our input.