CAL POLY | LIBRARY SERVICES
ROBERT E. KENNEDY LIBRARY

TO: Office of Science and Technology Policy

FROM: Anna Gold, University Librarian, California Polytechnic State University Library Services, San Luis Obispo, California

DATE: January 12, 2012

SUBJECT: Response to RFI: Public access to **peer-reviewed scholarly publications** resulting from federally funded research

About this response

The following information was prepared by the Kennedy Library at California Polytechnic State University in San Luis Obispo, California, in response to the request for information issued November 3, 2011, by the Office of Science and Technology Policy.

Information was primarily provided by David Beales, Associate Librarian for Engineering at Kennedy Library, in consultation with Timothy Strawn, Director of Information Resources and Archives, and University Librarian Anna Gold, with additional input provided by Marisa Ramirez, Digital Repository Librarian at Kennedy Library.

The Robert E. Kennedy Library and Cal Poly Library Services at California Polytechnic State University (San Luis Obispo) provide a comprehensive program of library services within a teaching-led comprehensive public polytechnic university, including access to scholarly and professional information, data services, and a digital institutional repository that recently passed the 1.3 million download mark.

The mission of Cal Poly Library Services is to promote open and informed inquiry, foster collaboration and innovation, support the unique needs of every student and scholar at Cal Poly, and contribute to the cultural life of our community. In common with other libraries in higher education, the Library is in a unique position in our organization to do this through access to technologies that support the creation, reuse, sharing and preservation of new knowledge.

Background

The current academic publishing model has evolved to digitally replicate the inherent protections to publishers of an increasingly obsolete print model although this model and the current infrastructure of the publishing industry are no longer the most suitable for optimizing the distribution of academics' work to their colleagues in other institutions. Libraries have participated in the process both as subscribers on behalf of their institutions, and by acting as a second tier document delivery service, making academic material available to the wider community. The limitations of the print medium meant that in the past there was little need to invest in controlling access to material, compared to the investment in making material as widely accessible as possible. This is no longer the case.

The competing arguments of publishers and the Open Access movement are well established. At Cal Poly's Kennedy Library, more widespread open access would deliver efficiencies, financial transparency and opportunity for community outreach. These would aid us in our mission through:

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

1

- The widespread adoption and implementation of interoperable data storage and retrieval systems.
- A more direct link between the cost of publication and the value of the publication product.
- A reallocation of staff and budget from compliance with protection of the publishers' products to maximizing access through compliance with open access standards.
- A more direct link between our information literacy training and the real-world information needs of small and medium-sized enterprises (SME's).

Cal Poly's Library Services offer the following further comments on the specific questions posed in the RFI:

1. *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? 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? What are the relative costs and benefits of such policies? What type of access to these publications is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise?*

   - Through their existing information literacy programs, university libraries are well placed to support the aim of "educating the next generation with 21st century skills" as set out in President Obama's "Strategy for American Innovation"[1]. However, information literacy skills are only nominally valued by SME's with limited access to peer-reviewed information; open access to peer-reviewed information would increase the value to SME's of information retrieval and evaluation skills, resulting in a direct link between students with information literacy training and university measures of student success.

   - Funded programs to provide information literacy training to SME's in local communities would benefit SME's in making best use of open access material and would provide opportunities for universities to build closer links with businesses and public libraries in their communities.

   - One step that US agencies could take to grow existing and new markets would be to provide seed funding for "Easy Access IP"[2]. In the UK, The Lambert Review recognized that public funding for university research was "intended to benefit the economy as a whole rather than to create significant new sources of revenue for the universities."[3] An outcome of the review, the "Easy Access IP" initiative is a model in which universities license a selection of their intellectual property for free to organizations that want to explore the use of the intellectual property, but for whom the risks of investing in IP would be too high. The aim of the founding partners of "Easy Access IP" (University of Glasgow, King's College London and the University of Bristol) is to "increase the engagement between universities and industry, and accelerate the transfer of university knowledge and expertise into the hands of the best commercial partner who can develop it to benefit the economy and society."

   - The required type of access to these publications should be based upon Cornell University's arXiv digital repository model of barrier-free submission and use, with standards maintained through the use of moderators and community feedback. As Cornell University recognizes, one institution cannot be relied upon to guarantee the long-term sustainability of this model.

2. *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? 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?*

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

2

- Government support for a Digital Copyright Exchange is a recommended step. One model for this is in the UK[4], where the government is bringing together interested parties to create a centralized database of intellectual property rights that can be easily bought and sold. Potential users of intellectual property (IP) would be able to exercise the required due diligence in locating the IP owner simply by searching this one central database. The listed benefits of such a database refer mainly to simplifying the use of IP in the creative industries; and this is a promising way of dealing with the issue of orphan scholarly works. It also has the potential to simplify the licensing and re-use of research-developed software, which is currently highly problematic, with multiple copyright claims frequently arise from one piece of software.

- A centralized source of information on publisher copyright policies would simplify the process of self-archiving. The SHERPA RoMEO website[5] is useful but incomplete.

- Publishers' intellectual input into the publication process is the management of the peer review process, content preparation, branding of the journal portfolio, and provision of access. The funding for this input is currently provided through control of access to articles. In the open access model, the funding for this activity is provided through the publication fee. This funding is protected by the effectiveness of the journal/publisher brand and through the efficient provision of access to the articles within the publishers' journal portfolios to the maximum number of relevant users. During the transition phase from hybrid open access publications to full open access, scientists would continue to be required by publishers to sign over copyright of their work, but as publishers become evaluated and valued based on the effectiveness with which they enable the dissemination of information, market forces would encourage them to sign back more of the authors' rights to redistribute and re-use their own work. There are no specific steps required to protect the intellectual property interests of publishers or scientists.

- During the transition from the current publishing model to an open access model, there is a lack of transparency from publishers – particularly of hybrid open access journals - as to the impact on subscription fees of the receipt of publication fees. Research funding agencies such as the Wellcome Trust are currently paying for the publication of open access material while the authors' institutions are paying for subscriptions to the same material. While this is acceptable in the short term as publishers transition to open access, it is recommended that funding bodies require that funded research only be published in journals with a clearly scheduled roadmap of how they will progress to pure open access funding.

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? Are there reasons why a Federal agency (or agencies) should maintain custody of all published content, and are there ways that the government can ensure long-term stewardship if content is distributed across multiple private sources?*

- Research communities should continue to provide the infrastructure for deposit of material in digital repositories in order to best serve the varied needs of their communities and institutions. Universities and publishers are also in the best position to provide the infrastructure for public access to scholarly publications. Competition will ensure that access services evolve to take advantage of emerging technologies. However, as publishers and universities evolve, particularly in their strategic priorities and the subject areas that they support, long-term stewardship of collections may provide little immediate competitive advantage to the host institution. It is recommended that digital and print stewardship be formalized and centrally incentivized.

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

3

- Examples of current digital preservation models include LOCKSS (Lots Of Copies Keep Stuff Safe)[6] and Portico[7].

  - LOCKSS is software that allows libraries to preserve their own copies of electronic content in an OAIS-compliant infrastructure. The content is continually updated to the most current format and the networked nature of the international LOCKSS community allows for balancing the load of providing wider access to material. The cost associated with providing access to material could be federally subsidized.

  - Portico is a centralized archive of the information content of peer-reviewed journals, stored in a standardized format. This ensures ongoing usability of content, establishes the content's authenticity, and ensures discoverability through metadata. Portico is intended to be a fail-safe for libraries in the event of catastrophic loss of access.

  The relative insecurity of Portico's centralized repository model compared with LOCKSS may be ameliorated by use of cloud storage; but it is difficult to see how the cost of providing open access to articles through this method could be proportionately shared. A centralized system of storage would also tend towards minimum common standards of storage for all users. The distributed LOCKSS model, in contrast, allows individual institutions to provide tailored storage and retrieval to their communities while providing agreed minimum levels of access to the wider community.

- Examples of print preservation models include the Western Regional Storage Trust (WEST)[8] and UK Research Reserve (UKRR)[9].

  - WEST is a distributed retrospective print journal repository program serving academic libraries in the Western Region of the US. Participating libraries consolidate and validate print journal backfiles at major library storage facilities and at selected campus locations. The resulting shared print archives ensure access to the scholarly print record and allow member institutions to optimize campus library space. The first phase of the project is due for completion in 2013. It is funded by a grant from the Mellon Foundation and by annual membership fees.

  - UKRR is a United Kingdom initiative designed to ensure completeness of the principal print research collection in the British Library, with two further copies retained within UKRR member libraries. Access to material held in the Research Reserve is provided via the British Library within 24 hours through electronic document delivery. In 2009, an initial government seed fund of £10 million ($16 million) covered the cost to individual institutions of deduplicating their print collections, allowing them to free up space in their libraries. The success of UKRR for journals has led UK Higher Education Institutions to consider taking the same approach for books.

  The model for efficient long-term preservation of print collections through cooperation is established. We recommend that the national adoption of this model be promoted through a federally funded program.

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?*

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

4

- We are not able to strongly recommend existing models for public-private partnerships at this time.

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?*

   - While metadata standards to encourage interoperable search are a worthwhile goal, they are vulnerable to being left behind by evolving technologies. OAI-PMH has been widely adopted by repositories but has been abandoned by the major search engines, as it is less useful for retrieving results than their own algorithms. It is suggested that metadata standards are most usefully considered within the limits of their user communities' standard business (for example, Categories for the Description of Works of Arts – CDWA)[10] . So long as they are XML-based, there is a useful degree of interoperability; the trend towards interoperable schema will continue while it is still useful. However, librarians are aware that any effort to maintain quality metadata standards is difficult. Metadata schema that are not directly related to the needs of the host institution/author are unlikely to be embraced wholeheartedly. Consequently, we cannot recommend particular core metadata standards.

   - Google, Yahoo and Microsoft's evolving microdata schema[11] should be investigated for its applicability to open access publishing through use of the Challenge.gov platform's prize offers for initiatives which improve public access to federally funded research. Two parallel areas to consider for prize funds are recommended; one for the most effective semantic search algorithm within the academic publishing sphere and another for effective search of universal metadata standards. The current US Patent & Trademark Office challenge[12] , in which participants compete to deliver an algorithm which will identify and locate elements within a patent document is an example of a way forward for using algorithms to work with more loosely controlled, interoperable metadata standards. NLMplus[13] is a semantic search and discovery application which improves access to the National Library of Medicine's (NLM) collection of biomedical data. It was developed in response to a challenge contest by the NLM.

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?*

   - Our chief recommendation is to minimize the cost to academic institutions of applying open access policies. To help achieve this goal, it is recommended that all federal funding agencies adopt a uniform open access policy, including an open access mandate and standardized application of copyright or creative commons protections of the authors work. The PubMed open access policy is well regarded within the library community and would be a useful initial benchmark. This will also simplify the shared services approach to backroom library operations that is of increasing importance in rationalizing shrinking budgets.

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?*

   - Ideally, all academic publication would be open access. However, peer-reviewed publications other than scholarly journal articles would be better addressed separately. For example:

o Book publication introduces issues of intellectual property protection related to author payment that are not present in journal publication. Additionally, the transition from print to electronic access is not as advanced for books as for journal articles, although this is rapidly changing. Models such as that for the National Academies Press[14] whereby free electronic access to new publications is subsidized by sales of the print version are interesting, but are not sustainable in the medium or long term as ebook readers overtake print.

o Conference proceedings are a more natural fit for open access. However, many scholarly societies are currently resistant to open access publication when it reduces the benefits of society membership by making previously exclusive journals publicly available. Conference attendance and associated proceedings are an alternative source of revenue that scholarly societies will rely on during the transition. Ideally, we would recommend that conference proceedings be included in an open access policy; but recognizing the challenges presented by scholarly society publishing, we recommend that making journal articles open access be a first priority; and that other publication types, with their associated issues, be brought into open access as a secondary undertaking.

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?*

- A mature open access publishing environment would not require embargo periods. A timetable for movement from an initially agreed embargo period, to no embargo, should be a requirement of the publishers' roadmaps toward open access (discussed in response to question 2).

- Because the interdisciplinary nature of academic publishing makes it difficult to establish embargo periods for different subject areas; and because of the transitional nature of the adjustment to open access, a standard six-month initial embargo period would protect publishers as they move towards open access.

References

1. A Strategy for American Innovation: Securing Our Economic Growth and Prosperity; The White House [Internet]. [cited 2012 Jan 9]; Available from: www.whitehouse.gov/innovation/strategy
2. Easy Access IP website: www.easyaccessip.org.uk
3. The Lambert Review of Business–University Cooperation; Her Majesty's Treasury UK. 2003 [cited 2012 Jan 9]; Available from: www.hm-treasury.gov.uk/d/lambert_review_final_450.pdf
4. Digital Copyright Exchange Feasibility Study. Available from: www.ipo.gov.uk/types/hargreaves/hargreaves-copyright/hargreaves-copyright-dce/hargreaves-copyright-dce-terms.htm
5. SHERPA RoMEO website: www.sherpa.ac.uk/romeo
6. LOCKSS (Lots Of Copies Keep Stuff Safe) website: www.lockss.org/lockss/
7. Portico website: www.portico.org/digital-preservation
8. Western Regional Storage Trust (WEST) website: www.cdlib.org/west
9. UK Research Reserve (UKRR) website: www.ukrr.ac.uk
10. Categories for the Description of Works of Arts – CDWA). Available from: www.getty.edu/research/publications/electronic_publications/cdwa

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

6

11. US Patent & Trademark Office challenge. Available from: challenge.gov/NASA/269-50-000-uspto-algorithm-challenge
12. Microdata schema. See: schema.org
13. NLMplus. See: nlmplus.com
14. National Academies Press website: www.nap.edu

Cal Poly San Luis Obispo, Kennedy Library Response to RFI on Public Access to Scholarly Publications, January 12, 2012
Contact: Anna K. Gold, akgold@calpoly.edu

7