

Dr Cameron Neylon – U.K. based research scientist writing in a personal capacity

Introduction

Thank you for the opportunity to respond to this request for information. As a researcher based in the United Kingdom and Europe, it might be argued that I have a conflict of interest. In some ways it is in my interest for U.S. federally funded research to be uncompetitive. There are many opportunities that have been brought through evolving technology that have the potential to increase the efficiency of research itself, as well as its exploitation, and conversion into improved health outcomes, economic activity, a highly trained workforce, and technical innovation. Globally this potential has not been fully realised. In arguing for steps that work towards realising that potential in the U.S. it might be expected that I am risking aiding a competitor and perhaps in the longer term reducing the opportunity for Europe to overtake the U.S. as a global research contributor.

However I do not believe this to be the case. The potential efficiency gains and the extent to which they would increase the rate of innovation and economic development are so great, that their adoption in any part of the world will increase the effectiveness and capacity of research globally. Secondly the competition provided by a resurgent U.S. research base will galvanise action in Europe and more widely, leading to a “race to the top” in which, while those at the lead will benefit the most, there will be significant opportunities for the entire research base. My contribution is made in that light.

Preamble

The RFI and the America Competes Act are welcome developments in the area of public information, as they take forward the discussion about how best to improve the effectiveness of publicly funded research. Nonetheless I must respectfully state that I believe the framing of the RFI is flawed. The concentration on the disposition of intellectual property risks obscuring the real issues and preventing the resolution of current tensions between researchers, the public that funds research, federal agencies, and service providers, including scholarly publishers.

The intellectual property that is generated through publicly funded research takes many forms. It includes patents, the scholarly communications of researchers (including peer reviewed papers), as well as trade secrets, and expertise. The funder of this IP is the taxpayer, through the action of government. Federal funders pay for the direct costs of research, as well as the indirect costs including, but not limited to, investigator salaries, subscription to scholarly journals, and the provision of infrastructure. That the original ownership of this IP is vested in the government is recognised in the Bayh-Doyle act which explicitly transfers those rights to the research institutions and in response places an obligation on the institutions to maximise the benefits arising from that research.

The government chooses to invest in the generation of this intellectual property for a variety of reasons, including wealth generation, the support of innovation, the creation of a skilled workforce, evidence to support policy making, and improved health outcomes. That is, the government invests in research to support *outcomes*, not to generate IP *per se*. Thus the *appropriate* debate is not to argue about the final disposition of the IP itself, but how best support the *services* that take that IP and generate the outcomes desired by government and the wider community.

A focus on services greatly clarifies the debate and offers a promise of resolution that can support the interests of all stakeholders. It will allow us to identify what the required services are, as well as how they differ across different disciplines and for different forms of IP. It will provide a framework in which we can discuss how to provide a sustainable market in which service providers are paid a fair price for their contribution.

If we focus on the final disposition of IP it will be easy to create a situation in which we argue about who made what contribution and the IP is either divided to the point where it is useless, or concentrated in places where it never actually gets exploited. If instead we focus on the deliver of *services* that support the generation of *outcomes* we will have a framework that recognises the full range of contributions to the scholarly communications process, allows us to optimise that process on a case by case basis, and ultimately forces us to focus on ensuring that the public investment in research is optimally directed to what is intended to achieve: making the U.S. more economically successful and a better place to live.

Response

(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?

1 a) New markets for traditional peer reviewed publications

There are two broad forms of new market that can be identified for peer reviewed publications resulting from federally funded scientific research. The first of these is “new” markets for the traditionally published paper. There is massive and demonstrated demand from the general public for access to peer reviewed papers, particularly for access to medical research. A second crucial market for traditional papers is small and medium enterprise. The U.S. has a grand tradition of the small scale technical entrepreneur. In the modern world these entrepreneurs require up to date information on the latest research to be competitive. Estimates of the loss to the U.S. economy from the current lack of comprehensive access to peer reviewed papers by SMEs are around US\$16 B (<http://osc.hul.harvard.edu/stp-rfi-response-january-2012>).

Education at levels from primary through the postgraduate can also benefit from access to current research, and effective training of a modern skilled workforce is dependent on training being up to date. I am not aware of any estimates of the potential national costs due to deficiencies in education that result from a lack of access to current research but an investigation of these costs would be worthwhile.

The incremental cost of providing immediate access upon publication to peer reviewed research communications is *at worst zero*. The incremental cost of making a publication more widely available once the sunk costs involved in its preparation and peer review have been covered is zero. The infrastructure exists, both in the form of journal websites, and other repositories to serve this content. The question is how to create a sustainable market in which the services required to produce peer reviewed papers can be supported.

Open Access publishers, such as the Public Library of Science and BioMedCentral have demonstrated that it is financially viable to make peer reviewed research freely available via charging for the *service* of publication up front. The charges levied by PLoS and BMC are in fact less than those charged by subscription based publishers for vastly inferior “public access” services. For instance, the American Chemical Society charges up to \$3500 for authors to obtain the right to place a copy of the paper in an institutional or disciplinary repository but limits the rights to commercial use (including for instance use in research by a biotechnology startup or for teaching in an institution which charges fees). By contrast the charge made by PLoS for publication in PLoS ONE is \$1350. This provides the service of peer review, publication, archival, and places the final, peer reviewed and typeset, version of the paper on the web for the use of any person or organisation for any purposes, thus maximising the potential for that research to reach the people who can use it to generate specific outcomes.

Again, the debate over where the IP is finally located, in which a publicly funded author has to purchase a limited right to use their own work, having donated their copyright to the publisher, is ultimately sterile. The debate should be focussed on the provision of publication services, the best mechanisms for paying for those services and ensuring a competitive market, and the value for money that is provided for the public investment. It is noteworthy in this context that a number of new entrants to this market, who have essentially copied the PLoS ONE model, are charging *exactly the same fee*, suggesting that there is still not a fully functional market and that there is a significant margin for costs to be reduced further.

1b) New service based markets for the generation of new forms of research outputs

A second set of markets are opened up when the focus is shifted from IP to services. The current debate has been largely limited to discussion of a single form of output the peer reviewed paper. However when we consider the problem from the angle of what services are required to ensure that the public investment in research generates the maximum possible outcomes, we can see that there will be new forms of services required. This include, but are not limited to, data publication and archival, summarization and current awareness services, integration and aggregation services, translation and secondary publication services.

The current focus on the ownership of IP for a narrow subset of possible forms of research communication is actively preventing experimentation and development of entirely new services and markets. Given the technical expertise contained within the U.S. these are markets where U.S. companies could be expected to take a lead. However the cost of entry to these markets, and the cost of development and experimentation, are made artificially high by uncertainty around the rights to re-use scholarly material. It is instructive that almost all innovation in this space is based on publicly accessible and re-usable resources such as PubMed, articles from Open Access journals, and freely available research data archives online. The federal government could support a flowering of commercial innovation in this space by signalling that it was concerned with creating markets for services that would support the *effective, appropriate, and cost effective* dissemination and accessibility of the full range of research outputs.

(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?

Again, I wish to emphasise that the focus on intellectual property is not helpful here. It is crucial that all service providers, including publishers, research institutions, and researchers themselves receive appropriate recompense for their contributions, intellectual and otherwise, and that we create markets that support sustainable business models for the provision of these services as well as providing competition that ensures a fair price is being paid by the taxpayer for these services and encourages innovation. This is actually entirely separate to the issue of intellectual property as many of the critical contributions to the process *do not generate any intellectual property in the legal sense*. Let me illustrate this with an example.

I have gone through the final submitted version, after peer review, of the ten most recent peer reviewed papers on which I was an author. I have examined the text and diagrams of these, which were subsequently accepted for publication in this form, for any intellectual property that was contributed by the publishers during the peer review process. *I have found none*.

I am not a lawyer, so this does not constitute a legal opinion but in my view the only relevant intellectual property here is copyright. No single word of text, or any element of a diagram was contributed to these documents by the publishers. In some cases small amounts of text were suggested by external peer reviewers and incorporated. However in the fifteen years I have been carrying out peer review I have never signed over the copyright in my comments to a publisher, nor have I been paid for the review of papers, so there is no sense in which the publisher has any rights to text or comments provided by external peer reviewers. The final published versions of these papers do have a small contribution of intellectual property from the publishers, the typesetting and layout in some cases, but these are not relevant to the substance of the research itself.

But my main point is that this argument is ultimately not helpful. The publishers for each of these papers *have* provided a range of *critical* services, without which the paper would not have been published, including the infrastructure, management of the peer review process, archival, and deposition with appropriate indexing services. These important services are clearly ones for which a fair price should be paid to the service provider. It is therefore the services that we require to purchase and the most effective and appropriate mechanism by which to purchase them, that should be the point of discussion, not the disposition of intellectual property.

Our focus should therefore be on identifying for the full range of research outputs:

- i) How to ensure that they are accessible to the widest possible range of potential users. This might include maximising rights of re-use, ensuring that the outputs are discoverable by appropriate means, translation, interpretation, and publication in alternative media.
- ii) Identify the services available, or if not available the services required, to achieve the maximum level of accessibility
- iii) Work with service providers to identify appropriate business models that will support the provision of the required services and the development of markets that will ensure a fair price is received for those services.
- iv) Tension the desired accessibility against the resources available to purchase services to provide that access. With limited resources it may be necessary and appropriate to choose, for

instance, between paying for peer reviewed publication and generating material targeted at a specific audience most likely to be benefit from the research output.

The optimal solution for most of these issues is currently unclear. There is one exception to this rule. Once the costs of preparing and reviewing a research output and making that output available online have been met there is no economic benefit or reduced cost achieved by reducing access to that output. There is no gain in paying the full costs for a service that places an output online but then limits access to that output.

(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?

Again, I feel this frames the question the wrong way, focusing on control and ownership of resources rather than the provision of services that enable discovery and use of research outputs. The question is not one of whether a distributed or a centralized approach is globally the best. This is likely to differ between disciplines, types of research output, and indeed across national borders. The question is how best to ensure that the outputs of federally funded research outputs are accessible and re-usable for those who could effectively exploit them. This will require a wide range of services focusing on different disciplines, different forms of research, but also crucially on different *user groups*.

The question for government and federal agencies is how best to provide the infrastructure that can support the fullest range of publication, discovery, archival, and integration services. This will inevitably be mix of services, and technical and human infrastructure, provided by government, commercial entities, and not-for-profits, some of which are centralised, some of which are distributed. Economies of scale mean that it will be more cost effective for some elements of this to be centralised and done up-front by federal agencies (e.g. long term preservation and archival as undertaken by the Library of Congress), whereas in other cases a patchwork of private service providers will be appropriate (specialist discovery services for specific communities or interest groups).

Once again, if a service based model is adopted in which a fair price for the costs of providing review and publication services is paid up front, guaranteeing that any interested party can access and re-use the published research output, then government will be free to archive and manage such outputs where appropriate while not interfering with the freedom to act of any other interested public or private stakeholder. This model can provide the greatest flexibility for all stakeholders in the system.

(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?

There are a range of such models ranging from ArXiv through relatively traditional publishers like PLoS and BMC to new and emerging forms of low cost publication that disaggregate the traditional role of the scholarly publisher into a menu of services which can be selected from as desired. It is not the place of government, federal agencies, or even scholarly communities to

attempt to pick winners at this very early stage of development. Rather the role of government and federal funding agencies is to make a clear statement of expectations as to the service level expected of the researcher and their institution as a condition of funding and an appropriate level of resourcing the support the purchase of such services as required for effective communication of research outputs.

The role of the researcher is to select, on a best efforts basis, the appropriate services required for the effective communication of their research, consistent with the resources available. The role of the funder is to help provide a stable and viable market in the provision of such services that encourages competition, innovation, and the development of new services in response to the needs of an evolving research agenda.

(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?

Standardisation and interoperability remain challenging problems both technically and politically. Federal agencies should take advice on the adoption of standards when and where they have widespread adoption and traction. However it is in general unwise for government to select or impose standards where there is not already widespread adoption. Federal agencies are well placed to provide an overview and where appropriate help to create “mid-course corrections” that will help to align the development of otherwise disconnected communities. The funding of specific targeted developments to support standards and interoperability development is appropriate. Consideration should be given at all times to aligning research standards with standards of wider relevance (e.g. consumer web standards) where appropriate and possible as these are likely to be better funded. There are however risks that the development of such standards can take directions not well suited to the research community.

Standards adopted by federal agencies should be open in the sense of having:

- a) Clear documentation that enables third parties to adhere to and interoperate with the standard.
- b) Working implementations of the standard that can be examined and reverse engineered by interested parties.
- c) Defined and accessible processes for the development and ongoing support of the standard.

(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?

Federal agencies, consistent with the Paperwork Reduction Act and guidance from the Office of Management and Budgets should adopt a “write once - use many” approach. That is that where possible the reporting burden for federally funded research should be discharged once by researchers for the communication of each research output. This means in turn that services

purchased in the communication of that research should be sufficient to provide for any downstream use of that communication that does not involve a marginal cost.

Thus, for instance, researchers should not be expected to write two independent documents, the peer reviewed paper, and a further public report, to support public access policies. Reporting on the outcomes of federally funded research should depend, as far as possible, on existing previous communications. The providers of publication services should be encouraged to remove or modify existing restrictions that limit the accessibility of published research outputs including for instance, length limitations, limitations on the use of links to background information and unnecessary use of highly technical language. Service providers should be explicitly judged on the accessibility of the products generated through their services to a wide range of potential audiences and users.

(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?

Yes. All research outputs should be covered by coherent federal policies that focus on ensuring that global outcomes of the public investment in research are maximised. The focus purely on research articles is damaging and limiting to the development of effective communication and thus exploitation.

(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?

Once the misleading focus on intellectual property is discarded in favour of a service based analysis it is clear that there is no justification for any length of embargo. Embargoes seek to ensure a private gain through creating an artificial scarcity by reducing access for a limited period of time. If a fair price is paid for the service of publication then the publisher has received full recompense in advance of publication and no further artificial monopoly rights are required. As noted above the costs of providing such services are at most no higher than is currently paid through subscription costs. With appropriate competition the costs might indeed become lower.

From the perspective of exploiting the public investment in research embargoes are also not justifiable. Technical exploitation, commercial development, and the saving of lives all depend on having the best and most up to date information to hand. Once a decision has been taken to publish a specific research result it is crucial that all of those who could benefit have access, whether they are private citizens with sick family members, small business owners and entrepreneurs, not-for-profit community support organisations, or major businesses.

Given the current environment of intellectual property law it *may* be appropriate under some circumstances for *the researcher or their institution* to delay publication to ensure that the research will be fully exploited. However there is no benefit to either the researcher, their institution, or the federal funding agency in reducing access once the research is published. Further it is clear that reducing access, whether to specific domains, communities, or for specific times, *cannot* improve the opportunities for exploitation of the research. It can only reduce them.

Conclusion

To conclude, to focus on the final disposition of intellectual property arising from the authoring of research outputs relating to federally funded research is to continue a sterile and non-productive discussion. Given that the federal government funds research, and provides its agencies with a mandate to support research through direct funding to research institutions, it is incumbent upon government, federal agencies, and the recipients of that funding to ensure that research communication is carried out in such a way that it optimally supports the exploitation and the generation of outcomes from that research.

To achieve this it is necessary to purchase services that support effective communication. These services have traditionally been provided by scholarly publishers and it is right and proper that they continue to receive a fair price for those services. The productive discussion is therefore how to develop the markets in these services that means service providers are viable and sustainable, and that there is sufficient competition to prevent price inflation and encourage innovation. That such services can be economically provided through a direct publication service model where the full costs of review and publication are charged at the point of publication has been demonstrated by the success of PLoS and BioMedCentral.

However this is just a starting point. A fully functional market will encourage the development of a wide range of competitive services that will enable researchers to select the most cost effective way of communicating and disseminating *their* research and ensuring that it reaches the widest possible audience and in turn is exploited fully. This in turn will enable federal agencies to support research, and its communication, in a way that ensures that the public investment is exploited fully for the benefit of the U.S., its citizens, and its economy.