

OPEN SCIENCE: TOOLS, APPROACHES, AND IMPLICATIONS *

CAMERON NEYLON

*STFC Rutherford Appleton Laboratory, Harwell Science and Innovation Campus
Didcot, OX11 0QX, United Kingdom*

SHIRLEY WU

*Program in Biomedical Informatics, Stanford University
Stanford, CA 94305 USA*

Open Science is gathering pace both as a grass roots effort amongst scientists to enable them to share the outputs of their research more effectively, and as a policy initiative for research funders to gain a greater return on their investment. In this workshop, we will discuss the current state of the art in collaborative research tools, the social challenges facing those adopting and advocating more openness, and the development of standards, policies and best practices for Open Science.

1. Introduction

Openness is arguably the great strength of the scientific method. Through open examination and critical analysis, models can be refined, improved, or rejected. Conflicting data can be compared and the underlying experiments and methodology investigated to identify which, if any, is more reliable. While individuals may not always adhere to the highest standards, community mechanisms of review over time have proved effective in developing coherent and practically useful models of the physical world around us. As Lee Smolin has put it, "*we argue in good faith from shared evidence to shared conclusions*".¹

The Internet and the World Wide Web provide the technical ability to share a much wider range of the evidence, argument and conclusions driving modern research. Data, methodology, and interpretation can also be made available online at lower costs and with lower barriers to access than has traditionally been the case. Wikis and blogs enable geographically and temporally widespread collaborations, the traditional journal club can now span continents, and the smallest details of what is happening in a laboratory can be shared.

* This workshop is supported by Burroughs Wellcome Fund.

The potential of online tools to revolutionize scientific communication and their ability to open up the details of the scientific enterprise so that a wider range of people can participate is clear. In practice, however, the reality has fallen far behind the potential. Although this is in part due to a need for tools that are specifically designed with scientific workflows in mind and to the inertia of infrastructure providers with pre-Internet business models, it is predominantly due to cultural and social barriers within the scientific community.

There will always be places where complete openness is not appropriate – for example, where personal patient records may be identifiable or where research is likely to lead to patentable results. These, however are special instances for which exceptional cases can be made, not the general case across the whole of the global research effort. Along with funders, government, and special interest groups there is a growing community of scientists interested in adopting more open practices in their research, and this community is developing as a strong voice in discussions of science policy, funding, and publication.

2. The case for open science

The case for taxpayer access to the taxpayer funded peer reviewed literature was made personally and directly in Jonathan Eisen's first editorial for PLoS Biology.² As a scientist in a small institution, he was unable to access the general medical literature for the information he desperately sought for an urgent family medical situation. More generally, as a US taxpayer he was unable to access the outputs of US government funded research or indeed of research funded by the governments of other countries.

A similar case can also be made for research data. Andrew Vickers, in a New York Times essay³, dissected the reasons that scientists gave for not making cancer treatment data available – data that could enhance patient survival times and quality of life. In other fields, the case for data sharing may seem less clear. There is little obvious damage done to the general public by not making the details of research available; however, in the non-clinical sciences, aggregation and re-analysis can also lead to new insights, more effective analysis, and even new types of analysis. Sharing enables more effective and

more efficient science. And this really is the crux of the matter: making the data, process, and conclusions available is nothing more than doing good science.

3. Tools for Open Science

The rapid expansion and development of tools that are loosely categorised under the banner of 'Web2.0' is what makes the sharing of research material practical. Many of the generic tools available have been adopted and used by a wide range of researchers but often these tools do not fit into the existing workflows of researchers.

Tools, whether they be social networking sites, electronic laboratory notebooks, or controlled vocabularies, must be built to help scientists do what they are already doing, not what the tool designer feels they should be doing. In the current environment, combined with the requirements and desire to provide access to laboratory data, the most obvious target is tools that make it easier to capture the research record so that it can be incorporated into and linked from papers. Once it is electronic, one can choose to make the record public at any stage.

4. Social issues and measures of success

Scientists are inherently rather conservative in their adoption of new approaches and tools. New methodologies often struggle to be accepted until the evidence of their superiority is overwhelming. The wider community is waiting for evidence of benefits before adopting either open access publication or open data policies. This actually provides the opportunity for individuals to take a first mover advantage.

Quantitatively measuring success in the application of open approaches relative to traditional approaches is a challenge, as demonstrated by the continuing controversy over the citation advantage of open access articles. However, Open Science has a clear public relations advantage as the examples are out in the open for anyone to see.

There are both benefits and risks associated with open practice in research. Often the discussion with researchers is focused on the disadvantages and risks. These concerns should not be dismissed, but taken seriously and considered. Radical change never comes without casualties, and while some concerns may be misplaced there are many real risks. What is important is providing

information to enable people to balance the risks and benefits of any approach they take.

5. Standards for Open Science

Two approaches to standards for Open Science are currently being discussed. The first of these is 'the fully supported paper'. In essence this is the idea that the claims made in a peer reviewed paper in the conventional literature should be fully supported by a publicly accessible record of all the material that contributes to those claims. The technical challenges of delivering such a record are substantial; however, it is difficult to argue that this shouldn't be available. While the target is challenging, it is simply a proposal to do good science, properly communicated.

While the fully supported paper would be a massive social and technical step forward it in many ways is no more open than the current system. It does not deal with the problem of unpublished or unsuccessful studies that may never find a home in a traditional peer reviewed paper. What then are the standards that need to be met before an organisation can claim they are doing Open Science? Science Commons have published four 'Principles for Open Science'⁴ which focus on the accessibility of published literature, research tools and data, and the development of cyberinfrastructure to make this possible. These principles provide a set of criteria that could form the basis of standards. The details are important, and will take time to work out. In the short term it is, therefore, arguably more effective to identify and celebrate examples of best practice and observe how it works in the real world. This will raise the profile of Open Science without making it the exclusive preserve of those who can immediately change their practice.

6. Summary

The development of deposition and data sharing mandates by a range of research funders show that real progress is being made in increasing access to both the finished products of research and the materials that support them. There is, however, a risk of over-enthusiasm driving expectations that cannot be delivered and of alienating the mainstream community that we wish to draw in. The fears and concerns of researchers in widening access to their work need to be

addressed sensitively and seriously, pointing out the benefits but also acknowledging the risks involved in adopting these practices.

Now is the right time to find examples of best practice; to celebrate these and to see what can be learnt from them. Now is the right time to be clearly articulating specific aspirations, and to provide targets that we can work towards. Now is the right time to organize as a community. The fully supported paper and the Science Commons principles are useful starting points, but the community will benefit from a concerted effort to develop and actualize additional goals, standards, and resources. Open Science is gathering momentum, and that is a good thing. But equally it is a good time to take stock, identify the best course forward, and effect change as widely as possible.

References

1. L Smolin, *Perimeter Institute Recorded Seminar Archive*. **PIRSA#08090035**, <http://pirsa.org/08090035/> (2008).
2. J.A. Eisen, *PLoS Biology*. **6(2)**, e48 (2008)
3. A. Vickers, *New York Times*. **January 22** (2008).
4. Science Commons, <http://sciencecommons.org/resources/readingroom/principles-for-open-science/> (2008)