Annotated Bibliography Series: Open Science

Last Updated 25 May 2016

Department of Geography
1984 West Mall
Vancouver, BC V6T 1Z2

Tel 1.604.822.2663
open.geog.ubc.ca

Recommended Citation:

Katz, Miriam. 2016. *Open Science*. Annotated Bibliography Series. UBC Open Geography: Vancouver, British Columbia. Retrieved from: http://open.geog.ubc.ca

Summary

- The main idea behind open science is to share information more openly and quickly, and with a larger audience.
- The Internet is the main facilitator of open science, allowing scientists to, for example, publish open notebooks and preregister their studies.
- Scientists are beginning to embrace open science, but the culture is slow to change, mainly because there are professional incentives for scientists to keep their work private, such as the imperative to publish or perish and the tenure process.
- As universities have become more intertwined with industry, the above problem has worsened.
- The question of who funds science is salient. There is an argument that if science is publicly funded, data should be available publicly.
- In order to advance knowledge, replication of scientific studies is required. Without innovative ways of licensing data to be open, replication studies cannot verify scientific findings. This has become a serious problem for experimental and quasi-experimental findings (in psychology and other disciplines).
- Preprints can serve a critical function in early dissemination of scientific findings. Journals that limit preprints should reevaluate their approach.

Annotated Bibliography

Timeline of Open Science Development

- 1500s- Open mathematical competitions and puzzles began (David, 2004)
- Early 1600s - Scientists began sending out anagrams to other scientists about their work. Once the anagrams were deciphered, the scientist had already received money for the work (Bartling and Friesike, 2014)
- 1660 - The Royal Society of London was founded and received charters from Charles II in 1662 and 1663 (David, 2004)
- 1665- The first scientific journal, *Philosophical Transaction* was founded (Bartling and Friesike, 2014)
- 1600s- Intellectual property commenced its development (Rhoten and Powell, 2007)
- 1793- Around 70 scientific organizations had been founded (David, 2004)
- 1839- Bentham argued that the state has a duty to protect inventors (Rhoten and Powell, 2007)
- 1865- Approximately 1 million professional scientists were tallied (Bartling and Friesike, 2014)
- 1980s - More than 50% of life sciences faculty in the US became consultants for industry (Rhoten and Powell, 2007)
- 1987- The US Patent and Trademark Office ruled that any biological material requiring human intervention (Rhoten and Powell, 2007)
- 2006- PLoS One Journal founded (https://en.wikipedia.org/wiki/PLOS_ONE)
- 2006- 20% of papers published every year are open access (Rhoten and Powell, 2007)
- 2006- The Dataverse Network was founded at Harvard's Institute for Quantitative Social Science (Crosas, 2011)
- 2007- The Linux web server had 59% market share, compared to 31% held by Microsoft's Internet Information Services (Rhoten and Powell, 2007)
- 2011- The *Earth System Science Data* journal was launched (Candela, 2015).

References

David, P. A. 2004. Understanding the emergence of "open science" institutions: functionalist economics in historical context. *Industrial and corporate change* 13 (4):571–589. http://dx.doi.org/10.1093/icc/dth023

This article provides a good timeline of scientific developments. David writes that from the 14th to 18th centuries, researchers received royal privileges in exchange for their patrons having monopolies on the discoveries. At that time, much of the research was conducted in order to improve patrons' reputations and it was subject to their whims. However, in the 16th century, there were many developments in the field of mathematics, including the organizing of open competitions and math puzzles that were available publicly. After that, many scientists began sharing information amongst themselves. In 1660, the Royal Society of London was founded and received charters from Charles II in 1662 and 1663. In fact, between the 1660s and 1793, approximately 70 scientific organizations were founded, which greatly aided efforts in making science more open. However, science at that time was by and large a male venture and keeping data private increased everyone's material benefit.

In this article, David cites Nelson and Arrow, who believe that open science leads to free riding. They argue that a few scientists could do most of the work, which would allow others to do very little to advance the cause. However, this does not seem to be the case at the moment. Many scientists are working to make science as a whole more open. This includes editing and using open textbooks, which could be written by one person and edited by others. If a majority of scientists make their work more available to the public, open science could explode.

Willinsky, J. 2005. The unacknowledged convergence of open source, open access, and open science. *First Monday* 10 (8). Retrieved from http://firstmonday.org/article/view/1265/1185

This article brings up a few important points about open science and open source. Willinsky writes that the Royal Society of London for Improving of Natural Knowledge, which has aided in opening up science since its founding in 1660, would not have been established without the hard work of women. This shows that women have played an important role in science becoming more open. As the years have gone by, publishers have argued that they own scientific work, despite the fact that they invest so little, as Willinsky argues. Also, during the 1960s and 70s, software was shared very openly, as it was not yet a commercial product. Thus, Willinsky illustrates that the history of science has experienced more open and more closed periods.

The last point mentioned is an important one. Because of the establishment of companies such as Microsoft, software has become property. Yet the current movement for open source and free software is strong. In the future, it is possible that software will become less proprietary, which could have many implications for open science including more sharing of technology. Also, as mentioned above, publishers get a lot of benefit from scientists' work while investing very little (although, costs of publishing do continue to rise). Perhaps scientists could publish their work independently and others would still review it. If this were to occur, open science could end up being a given in society.

Rhoten, D., and W. W. Powell. 2007. The Frontiers of Intellectual Property: Expanded Protection versus New Models of Open Science. *Annual Review of Law and Social Science* 3:345–373. DOI:10.1146/annurev.lawsocsci.3.081806.112900. Retrieved from https://web.stanford.edu/group/song/papers/ScienceandPropertyARLSS.pdf

This article is less about open science and more about the history of science and intellectual property (IP). IP was developed in the 17th century in order to deal with who benefits from original work. Locke argued that IP should last for a limited time and Bentham wrote in 1839 that the state should protect inventors. In the US, patent law evolved quite a bit over the 19th and 20th centuries. At first, anything in nature could not be patented but this changed in the 1980s when the oncomouse and a type of bacteria were patented. In 1987, the US Patent and Trademark Office ruled that any biological material requiring human intervention could be patented. At the same as these rulings, federal money for higher education began to drop. By the late 1980s, more than 50% of life sciences faculty in the US consulted for industry. Currently, the picture is mixed. Government support of higher education is still low, but Rhoten and Powell write that 20% of papers published every year are open access and as of 2007, the Linux web server had 59% market share, compared to 31% held by Microsoft's Internet Information Services.

This paper is another example of how proprietary science has been in the past. However, open access is gaining a strong foothold in North American society, as this article shows. It is interesting that the authors cite Bentham who wrote that the state should protect inventors. In the modern day, what would this look like? As mentioned earlier in this bibliography, if scientists are receiving government funding for their work, they may be required to make their work available to the general public. However, if the state protects inventors, the inventors may feel that they could keep their work hidden from others because they would consider their work to be their own intellectual property. The implications of either approach could be large.

Waldrop, M. M. 2008. Science 2.0: Great New Tool, or Great Risk? *Scientific American* (January). Retrieved from http://www.scientificamerican.com/article/science-2-point-0-great-new-tool-or-great-risk/

This magazine article brings up both advantages and disadvantages of open science. In terms of advantages, Waldrop argues that discoveries can be made more quickly with open science, especially when scientists publish their notebooks openly. Also, for people who work without mentors, open science can allow for more learning and collaboration across borders. In terms of the disadvantages, Waldrop shows that they are mainly perceived, not real. For example, scientists think that their work could be stolen if it is out in the open. However, if someone modifies a scientist's web page, for example, the page will show a time stamp; this is an incentive not to steal someone's work. Finally, one aspect of open science is blogging. At the moment, blogging is an activity that scientists do not get credit for. Thus, they may be more likely not to discuss their work online.

With regard to the last point, it is possible that in the future, as an example, scientists who are up for tenure could get credit for work such as blogging, especially if they are doing work that has international implications. Also, in relation to mentorship, open science could have huge implications. If someone in the developing world wants to work with someone in the developed world, open science and the Internet could make that possible. Finally, graduate science education could include elements of open science, so as to encourage scientists to discuss their work online.

Tacke, O. 2010. Open Science 2.0: How Research and Education can benefit from Open Innovation and Web 2.0. In *On Collective Intelligence*, 37–48. Retrieved from http://www.olivertacke.de/wp-content/uploads/2011/02/Tacke-2010-Open_Science_2.0-preview-100613-OLT.pdf

This book chapter includes further definitions of open science. Tacke writes that open science "does not only mean sharing prefabricated knowledge with others but also developing a mutual comprehension of problems and to work jointly on subjects relevant to theory and practice" (p. 39). Tacke also states that there are 3 phases of open science: generating ideas, exploring and evaluating ideas, and implementation and feedback. Given that scientific problems are becoming more and more complex by the day, Tacke argues that more people are needed to solve them, hence open science. Also, many scientists are working on issues that directly impact the public people every day, so public input should be valued and integrated into research design. Often, scientists discuss their work with the people around them; open science means that work can be discussed with anyone around the world. Finally, for scientists who are not native English speakers, they can receive help from others, if they need to write articles in English or present in English at conferences.

Communication among scientists is an important component of the open science movement. This article points to several benefits that can be accrued to research design and findings by encouraging communication between colleagues, across languages, and with the public (input at early stages of research). The final statement shows that science has become an English language venture. If science does become more and more open, will that remain the case or will other languages gain more of a foothold? Translation software has become more sophisticated over time, so it is possible that work could be translated into any language. Finally, what are the implications of the public having input into scientific work? Some have argued that laypeople should not have that power. Yet with increasing access to trends. However, with the Internet becoming so prevalent and with scientists publishing open notebooks, it is only a matter of time before the public input plays a larger role in the design and evaluation of gives more of an opinion into scientific research.

Boulton, Geoffrey, Rawlins, Michael, Vallance, Patrick, Walport, M. 2011. Science as a public enterprise: the case for open data. *The Lancet* 377:1633–1635. DOI: 10.1016/S0140-6736(11)60647-8. Retrieved from http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(11)60647-8/abstract

This commentary brings up a few important points. The authors argue that scientists see their data as their own, because they want to protect their discoveries. However, laypersons who would like to verify scientific data need access to fulfill their goals. The authors then discuss the fact that data from clinical trials is often very hard to access, despite the fact that in 1997, biomedical scientists agreed to transparency and data sharing principles at an important meeting. Clinical trial data is often protected by commercial interests, which suggests that opening up that data would probably be very difficult.

The points that the authors bring up could have several implications. Scientists feel a need to protect their data in order to protect their reputations. However, opening up data to the public could in fact improve reputations, because it would make it easier to verify the data. There are also ways of tying scientists' reputations to how much they share data; open access articles are usually cited more often than articles in subscription-only journals. Finally, transparency could improve practices in biomedicine, as they would be more open to scrutiny, giving companies incentives to strengthen research protocols.

Crosas, M. 2011. D-Lib Magazine The Dataverse Network ®: An Open-Source Application for Sharing, Discovering and Preserving Data Incentives for Data Sharing: Recognition, Visibility, Ownership. *D-Lib Magazine* :1–7. http://www.dlib.org/dlib/january11/crosas/01crosas.html.

This article describes the Dataverse Network (DVN), which was founded in 2006 at Harvard's Institute for Quantitative Social Science. Crosas writes that many researchers would prefer not to share their data and therefore, would not use a data archive. However, the DVN allows authors to have control over their data. Currently, DVN has over 37,000 studies and more than 600,000 files. Each of the data sets includes the following information: The author, distribution date, title, URL, and UNF, which is a universal numerical fingerprint. The UNF never changes and it can be assigned both to the whole data set and separate UNFs can be assigned to sections of the data. After authors upload their data, the DVN conducts the following steps:

1. The metadata is separated from the primary data
2. Statistics are generated from each variable
3. The data are reformatted so that they can be preserved
4. The UNF, or if appropriate UNFs, are generated.

Finally, Crosas mentions that there are 3 levels of access control: The first is known as a public study with terms of use, where users have to agree to the terms to access the data. The second is known as a study with restricted files, meaning that users need a password to see all the data, but the descriptive information is available publicly. Finally, a restricted study means that the metadata can be searched, but the primary data cannot be seen. For the latter two levels, users get access to the data by asking the researcher for permission, which gives the researcher control over their data.

Crosas concludes the article by stating that "An application like the DVN not only allows researchers to find and access large data sets available in known archives, but more interestingly, it helps them find and easily access small data sets from other researchers that would otherwise sit in local computers with the risk of being lost" (p. 5). This shows that data repositories such as the DVN can have very large implications. All over the world, scientific data is subject to the file drawer problem, where results that were not satisfactory are filed away and no one can see them. Some researchers may choose to make such data available to other researchers and to the public through applications such as the DVN; if they did, potentially interesting discoveries could be viewed by anyone.

Boulton, G. 2012. Science as an Open Enterprise. *Royal Society*. London. Retrieved from https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

In this report, Boulton discusses a few different aspects of open science. Firstly, he argues that open science provides more of a chance to catch scientists' mistakes as data becomes available to more and more people. One recurring issue in science is that data is only partially reported; if this were remedied through the use of the principles of open science, the culture of science would change. Second, Boulton brings up the issue of who funds science. If science is funded publicly, one could argue that the data should be available publicly. Conversely, if private interests are funding scientific research, there is a financial incentive to keep the results private.

As mentioned above, there could be a duty to make scientific data available publicly if it is funded by the public. However, the public might needs to agitate for that. This could have both positive and negative implications. Scientists would need to ensure that their research practices were reliable and valid, which could potentially take great effort. However, the public may take scientific research less seriously if there is a failure to implement the principles of open science. Thus, there is an incentive for scientists to conduct research rigorously.

Collaboration, Open Science. 2012. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science* 7 (6):657–660. http://pps.sagepub.com/content/7/6/657.full.pdf+html.

This article discusses a project on the Open Science Framework where volunteers are attempting to replicate a large number of psychological studies. The authors mention that there are many incentives for authors not to replicate existing studies. However, in many fields such as biology, only about 25 percent of studies can actually be replicated. When this article was written, 72 volunteers from 41 different institutions had participated, which entails choosing one study from an article. A successful replication means obtaining approximately the same statistical significance or a directionally similar finding. As of 2015, according to the project website, fewer than half of the studies had been replicated (https://osf.io/ezum7/wiki/home/).

The authors mention that the failure to replicate a study does not necessarily mean that the original study was flawed. It is possible that, for example, the effect size in the original experiment was lower than was actually reported or that the methodology used to replicate the study was different than the original study. However, if more than half the studies that have been replicated so far had achieved replication, confidence in the original experiments might be higher. Also, as mentioned above, there are many incentives not to replicate previous studies - if there were, and if the results were available publicly, science might take a different form.

Grand, A., C. Wilkinson, K. Bultitude, and A. F. T. Winfield. 2012. Open Science: A New "Trust Technology"? *Science Communication* 34 (5):679–689. http://dx.doi.org/10.1177/1075547012443021

The basic argument of this article is that open science would lead to the public trusting science and scientists more. The authors write that if science becomes more public, more parties would be involved, and scientific validity would be easier to determine. Open science also "supports the understanding of science as a dynamic, tentative, uncertain, and constantly revised activity" (p. 681). This means that any scientific finding is always up for debate and discussion. Finally, the authors bring up the fact that with open access journals, scientists have to pay to submit their work, which could stifle research.

The last point mentioned has many implications. Open access journals are currently proliferating. There are two ways to publish open access articles: green access, meaning that authors put their articles in a publicly accessible repository such as PubMed Central. There is also gold access, which means that authors submit their articles to an open access journal, which does not charge subscription fees. However, authors pay an article processing fee. Thus, making open access journals an affordable venture is important. Many studies have found that open access articles are cited more often, which is one incentive to maintain their use. Also, if science is viewed as tentative, as the authors suggest, open science holds more weight. Publishing results before an article is published, for example, might become more commonplace if scientists viewed their own work as up for debate.

Bartling, S., and S. Friesike. 2014. *Opening Science*. Springer Open. Retrieved from http://link.springer.com/book/10.1007%2F978-3-319-00026-8

This book discusses the history behind science and open science. Bartling and Friesike bring up the fact that science's history has not been marked by openness. In the 17th century, scientists would send out anagrams "that did not make sense without knowledge of the discovery" (p. 6). Once the anagrams were deciphered, the proponent had already received money from their patron. This changed somewhat after the first scientific journal started in 1665, *Philosophical Transactions*. Although the source of this information is not cited, they estimate that 200-hundred years later, one million professional scientists existed. They also estimate there are approximately 100-million working professional scientists as of 2014. In recent years, open science evolved from affordances of the Internet - which allows ideas to become public earlier and spread rapidly. Bartling and Friesike refer to open science as "characterized by its openness. Scientists share results almost immediately and with a very wide audience." (p. 10)

The authors discuss what they consider to be the 5 schools of open science:

1. the infrastructure school, which advocates for new platforms and tools to disseminate scientific knowledge;
2. the public school, which argues that science should be more public and more comprehensible to laypeople;
3. the measurement school, which discusses how to measure impact differently;
4. the democratic school, which argues that knowledge should be available to all who want it, and finally,
5. the pragmatic school, which argues that there should be more collaboration across different fields.

The fact that scientific practices have, for the most part, not been very open could mean that open science may not achieve all of its ambitious goals in the near future. It takes a long time for culture to change and scientific culture is no exception. This may also have an impact upon making science more comprehensible to laypeople; scientists may be reluctant or may just not know how to make their writing clearer. However, this does not mean that research cannot be made public more quickly. The Internet has made it very easy to disseminate knowledge at a fast speed. Scientists can start their own blogs to discuss their research and they can publish articles in open access journals, so that anyone can read it. If research is made publicly more quickly, it will be easier for anyone to find it and comment.

Friesike, S., B. Widenmayer, O. Gassmann, and T. Schildhauer. 2014. Opening science: towards an agenda of open science in academia and industry. *The Journal of Technology Transfer* 40:581–601. DOI: 10.1007/s10961-014-9375-6. Retrieved from http://link.springer.com/article/10.1007/s10961-014-9375-6

In this article, the authors discuss their views on four perspectives of open science. The first perspective being philanthropic where research is made more available. The second perspective is reflationary, which means that results are made available before publication. The third perspective (constructivist perspective) means that scientists from different disciplines can work together and provide different understandings of the same problem. The fourth perspective (exploitative perspective) refers to science that aims at solving everyday problems. The authors also argue that scientists who make their work more open can still claim first authorship and will reap the benefits from that. However, given that many professors operate in the publish or perish model, research is likely to be kept private.

The authors, who are from Germany and Switzerland, write in this article that, "in most EU funded projects the open distribution of knowledge is already a prerequisite to get funding" (p. 594). This could have implications for the funding of science in North America and other places around the world. If scientists had to openly distribute their data in order to get public funding, open science would be much more of a reality. However, in the North American context, there is a large amount of privately funded research. What if private groups such as foundations compelled researchers to release their data? This is important to consider.

Candela, L. 2015. Data Journals: A Survey. *Journal of the Association for Information Science and Technology* 66 (9):1747–1762. http://onlinelibrary.wiley.com/doi/10.1002/asi.23358/abstract.

This article surveys data journals, which are a very recent phenomenon. Candela writes that as of 2009, several journals were accepting data, but there were many restrictions on their volume, for example. In that same year, the *International Journal of Robotics Research* began accepting data papers and in 2011 and 2013, the *Earth System Science Data* journal and the journal *Scientific Data* were launched, respectively. For this article, Candela surveyed 116 data journals from 15 different publishers. The papers were mainly from the fields of medicine, biochemistry, genomics, molecular biology, and the agricultural and biological sciences. Candela's main finding is that there is very little consistency among the journals and what they require of authors. Some journals accept zip files of the data and some merely ask authors to make the data available. For the latter option, researchers would submit their data to an accredited repository, and they may be required to pay a fee for that. Also, if they submit a data paper to an open access journal, they would be required to pay an article processing charge as well. Finally, journals have different standards in terms of peer review. As an example, the *Earth System Science Data* journal has a two-stage peer review process: Data papers are available for comment for about 8 weeks and then the author needs to look at the comments, revise the manuscript, and then the editor can decide whether to approve the article or not.

Candela brings up an interesting point concerning peer review. He notes that a person may not be able to properly review data; it may be better for a computer to do so. Also, as he illustrates in the article, data journals have a lot of work to do in terms of standardizing the process of submission, however the data journal itself is still a very young phenomenon. Other journals have similar standards for article submission, but they have had time to establish those standards. In the future, when data journals have had time to mature, more scientists could make their data more available to those who want it.

Nosek, B. A. et al. 2015. Promoting an open research culture. *Science* 348 (6242):1422–1425. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550299/

This article mainly discusses standards for transparency and openness in science, which were developed by the Transparency and Openness Promotion Committee (TOPC) at the Center for Open Science. The authors argue that science would be more credible if it was more transparent, but that most scientists do not practice transparency principles. In terms of the standards, the authors write that scientists should be rewarded for the time they spend on, for example, conducting replication studies. Scientists would also get credit for preregistering their studies and sharing data. The above mentioned TOPC recognized that the standards would be flexible depending on the discipline and on the different journals. Already, "*Psychological Science* awards badges for "open data," "open materials," and "preregistration" (p. 1424). Thus, some disciplines are further ahead than others.

It is important that organizations such as the Center for Open Science develop committees like the TOPC develop standards, but there is still a question of whether the standards will be adopted. What would be more effective is if there were incentives for scientists to be more open with their work. If such incentives existed, scientists would be much more likely to preregister their studies and to conduct replication studies. The implications could be very large if that did happen. Science as a whole would change, though it would take time for universities and other research institutions to encourage their employees.