CHAPTER 8

Open Science: A Global Enterprise

Luc Henry and Martin Vetterli

INTRODUCTION

Over the past decade, "open" has become a mantra. "Open data", "open innovation", "open government" movements all call for a more fluid exchange of information between supply and demand. Similarly, initiatives aiming at making academic research more accessible have emerged in the early 2000s and claim their place under the umbrella of "open science". Due to a perceived reproducibility crisis and the explosion of digital technologies, the pace of the open science movement has recently accelerated. The adjustments it calls for have become a necessity to improve the access to and the diffusion of high-quality research results.

However, in order to establish a culture of robustness in academia, as well as find a new equilibrium between quantity and quality, the open science discourse should shift priority from mere access to careful curation. We believe that there is a need for the adoption of a new set of best practice in (digital) scholarship, and, as a consequence, the evaluation methods for both individual researchers and the results they publish should be revised. Considering the complexity of the task, we argue that individual countries, funding organizations or institutions alone cannot be responsible for the systemic change needed in order to make a swift transition to a sustainable digital scholarship.

WHAT IS OPEN SCIENCE?

"Open Science is at a stage where no-one is quite sure what it is, but they think it's a good idea." — Martyn Rittman, Publishing Services Manager MDPI

The diversity of initiatives that compose open science makes it difficult to agree on a unified definition. Although some have hijacked the term for commercial purposes, most of the open science practices — around open access and open research data for example — are aiming at making the output of (publicly funded) research freely accessible on the Internet, and reusable by anyone without restriction. Others want to transform the scientific endeavour and make it more fluid, more collaborative and participative, more fair and transparent in general.

A multitude of projects challenging the status quo of how knowledge is produced, disseminated and reused have adopted the terminology. These include non-traditional and dynamic publication formats, collaborative authoring tools, post publication peer-review, the widespread adoption of preprints (e.g. arXiv, bioRxiv, etc.) and other repositories, but also some forms of citizen science, the use of social media, etc. Taking this diversity into consideration, how can one separate the wheat from the chaff and decide which initiatives should be taken seriously and adopted by researchers and their host institutions?

The motivations of open science advocates are rooted as much in recent developments of the scientific method as they are in a set of values that have existed since the first scientific revolution in the 17th century. Instead of giving an exhaustive list of sound open science projects, some have tried to embrace the diversity and blurry definition. Benedikt Fecher and Sasha Friesike, at the Alexander von Humboldt Institute for Internet and Society in Berlin, have identified five schools of thought (see Table 1), each concerned with improving a different aspect of scholarship (Fecher & Friesike, 2013). Their categories help understand the value of the various approaches. After a critical assessment — beware of "openwashing", a term derived from "greenwashing", describing the act of portraying a product or company as open, although it is not — there is no doubt that research institutions will benefit from adopting the initiatives that have the clear objective of improving the quality and transparency of research practices and outputs.

Table 1: Open Science: Five Schools of Thought

Pragmatic school	Better, more efficient and collaborative research
Infrastructure school	Technological architecture supporting open science
Measurement school	Alternative impact evaluation methods
Democratic school	Unrestricted access to knowledge
Public school	Public participation to knowledge production

IF OPEN SCIENCE IS THE SOLUTION, WHAT IS THE PROBLEM?

A majority of open science initiatives have emerged from academia itself, usually under the impetus of researchers frustrated with a particular aspect of the scientific enterprise. Understanding the somehow unrelated — yet intertwined — concurrent situations that have led to the issues described below is important to evaluate the potential transformation that open science represents. Research evaluation, career promotion, access to the literature, funding allocation, are all being criticized, and whether open science initiatives represent answers to these critiques remains to be seen. This section summarizes the roots of the growing frustration in the scientific community that led to the emergence of the open science movement.

A Growing Commodification of Knowledge

The first — and most important — source of frustration is one of increasingly hindered access to information, and the consequence of private, for-profit companies taking an overly large responsibility for organizing the quality control and the distribution of scientific literature over the course of the 20th century. In a capitalist tradition, the near-monopolistic position of publishers allowed them to exploit a system in which researchers give away their intellectual property for free, while the dynamic molecules of the research process are fragmented into static and pay-walled atoms of knowledge, mostly documents in the PDF format.

Abusing their dominant position, several publishers have charged libraries ever-increasing fees to access new research, ultimately leading to the exclusion of institutions with limited resources. In the early 2000s, open access emerged as a promising solution to this problem. The movement proposed to reform the publishing industry and challenged funding agencies and research institutions to make all their outputs available online, free from all restrictions on access (e.g. access tolls) and free of many restrictions on use (e.g. licence restrictions). The three influential events that led to the establishment of the open access movement were the Budapest Open Access Initiative (Budapest Open Access Initiative, 2002) (see Box 1), the Bethesda Statement on Open Access Publishing (Suber *et al.*, 2003) and the Berlin Declaration (Berlin Declaration, 2003).

Although many expected it would kill two birds with one stone, open access is a disappointment to some. It is encouraging that, thanks to one open science project entitled Unpaywall, the share of legal open access (in opposition with illegal sharing platforms such as SciHub) is now believed to reach nearly half of the total volume of existing literature (Piwowar *et al.*, 2017). But for those who criticized the for-profit objectives of private commercial publishers, open access failed to break their monopolistic position. Indeed, the prerequisites for open access (accessible, reusable) are perfectly compatible with a for-profit approach. In a transition from one revenue model to the next — trading subscription fees, site licences or pay-per-view charges against article processing fees — the costs of publishing incurred on research institutions may even have increased (Cambridge Economic Policy Associates Ltd., 2017).

And because of the prestige one obtains when publishing in top-ranked closed journals, the moral imperative to make all research freely accessible was never completely met. An evaluation system based entirely on the reputation of publication venues and in which quantity prevails over content quality has another dramatic consequence: universities are in effect outsourcing their talent management to journals that use marketing strategies to compete for visibility. Because it failed to drift away from the notion of impact factors, the current implementation of open access does not completely solve the issues related to the loss of control over the allocation of scientific merit.

Box 1: *Excerpts from the Budapest Open Access Declaration*

> *An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.*
>
> *By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.*

The Reproducibility Crisis

The second crisis that academia is facing is a consequence of the decline in the reliability of the knowledge it produces. The reputation economy that drives scientific careers in academia has been using and abusing metrics that often reflect quantity more than quality. There is increasing evidence that using such proxy for productivity contributed in part to the lack of reproducibility for published scientific results (Ioannidis, 2014). Some disciplines are more affected than others, but, in general, both academic and corporate researchers now question the validity of what they can read in the literature.

Whether it is a consequence of fraud, honest mistakes or underpowered studies, the lack of reproducibility is likely rooted in the pressure to produce and publish (positive) results. With journals emphasizing the need for scientific originality in submissions, and a majority of science career moves requiring a long publication list in the most prestigious journals, making one's research stand out can come at the cost of cutting corners, or worse, fabricating results. This could explain why journals with higher impact factors have higher retraction rates. It could of course also be the consequence of an extensive scrutiny by more, and more careful, readers, who tend to notice mistakes more often (Fang, Casadevall & Morisson, 2011). In any case, the explosion of the retraction rates across all disciplines (Marcus & Oransky, 2015; Nature News, 2014) calls for a re-evaluation of peer-review process.

Altogether, this means that current research is less efficient than it could be. And the lower-than-expected quality of scientific facts per part of budget invested has led to frustrations. For example, a provocative estimate recently suggested that $28 billion a year is spent on irreproducible biomedical research in the US alone (Freedman, Cockburn & Simcoe, 2015). The reasons for the irreproducibility of scientific studies are diverse, but they are all rooted in an insufficient quality control and an incentive system that increasingly appears to be flawed. Fixing incentives, adopting standards, carefully documenting and sharing all methods and results; a list of solutions to the reproducibility problem is relatively easy to draft, but much more difficult to implement.

Rampant Digital Frustrations

Paradoxically, the advent of digital technologies has not always translated into an improvement of the scientific method. The very same academic community that invented email and the World Wide Web primarily to share scientific discoveries has been reluctant to explore the full potential of these technologies. In contrast with the fast digitalization of virtually every corner of society, the slow adoption by research communities has created tensions. Three aspects are particularly problematic.

First, while the technologies behind instrumentation have seen tremendous improvements, the format of scientific documentation — laboratory notebooks and scientific reports — is mostly unchanged since the 17th century. Pen and paper are still the norm in a majority of disciplines — although most observations are made with instruments that produce some form of digital medium — and the PDF has been a very poor and static digital substitute for printed documents. There is therefore an urgent need to fill the gap between the expectations and the reality of the current knowledge dissemination model. A generation of computer-literate researchers used to the web 2.0 is asking for a change in the way research results are communicated.

Second, with a peer-review system flooded with manuscripts, there are long delays from discovery to dissemination that are difficult to justify. A recent analysis of thousands of journals revealed the time between submission and acceptance and that between acceptance and publication (Woolston, 2015). For popular open access journals, the former was 75-175 days, and the latter 5-55 days. Although they have been used in some areas of physics and mathematics since the dawn of the Internet, pre-prints are now becoming increasingly popular in other disciplines, including the social sciences. One needs to welcome this with caution: pre-prints are not peer-reviewed articles and should not be considered as such. But, in light of the time it takes to formally publish a scientific discovery, pre-prints may be an opportunity to get results disseminated faster, prior to formal validation.

Last, with the emergence of information technologies, the scientific method is expending, however we fail to share the research output in formats beyond traditional publications. The two recent additions — computational and data-driven research (see Box 2) — have triggered an explosion in the number of computational methods and digital artefacts scientists use in their research projects. They can be new software, custom code, large data sets, photographs, sound and video recordings, etc. New platforms need to be developed in order to share them with the rest of the community and get credit for it. Although it could be at the expense of creativity, there is a need for some standardization in the way scientists deal with digital data in order to guarantee reusability and interoperability. Programming and statistics, life cycle management and database maintenance, all have become a crucial part of good scientific practice, yet very few scientists get trained accordingly.

Box 2: *The Four Branches of the Scientific Method*

Branch 1: Deductive (mathematics, formal logic)
Branch 2: Empirical (controlled experiments, statistical analysis)
Branch 3: Computational (simulations)
Branch 4: Data driven (aka "Big Data")

OPEN SCIENCE IN PRACTICE

Despite all the promises for a better, efficient, more inclusive scholarship, the adoption of open science principles at research institutions is still marginal. And EPFL is no exception. Apart from situations that call for legitimate exceptions — i.e. intellectual property, privacy and security — research should be an inherently open and global endeavour. However, we are in the presence of a series of paradoxes that introduce major obstacles to the widespread implementation of open science initiatives. Various factors play important roles in enabling or inhibiting their adoption.

To achieve a cultural change, EPFL is investigating actions that could lead to an increased awareness among researchers. We also need to guarantee the availability of infrastructure, training and career incentives. The following sections suggest how this can be done.

Building the right incentive frameworks

Several reports and studies that investigated the current state of data sharing have pointed out to the same issue: within the current incentive framework, what is in the best interest of the scientific community — not to mention that of the whole of society — is not necessarily in the best interest of individual scientists trying to build a career. Known as "the prisoner's dilemma", this paradox emerges in the reputation economy of science. Even if we disregard the fear of being scooped by other researchers, the practice of open science often represents a significant opportunity cost. The curation of increasingly complex and voluminous research data requires learning new skills and spending time not devoted to producing new data. It will not be encouraged unless it is recognized as a significant contribution to research. New forms of incentives will be necessary to promote this cultural change, while new infrastructures, tools and methods will contribute to an effortless transition (see below).

The San Francisco Declaration on Research Assessment (DORA, 2012) that was initiated by the American Society for Cell Biology represents one first step towards a change in evaluation methods. There are many further steps to take: promoting and rewarding reproducibility studies will improve the number of trusted results; enabling data citation and taking data sharing into consideration during evaluation will encourage the reuse and pooling of these valuable resources, with the potential of significantly improving the efficacy of science budgets.

Supporting bottom-up initiatives

Innovation often arises from frustrations and users are usually the best source of clever solutions to systemic problems. Open science initiatives are typically community-driven solutions but the majority of its most active supporters are not being recognized for their contributions. Institutions need to find ways to distribute resources to support the initiatives that are aligned with their values. This means that the research community has to investigate new revenue models for publishing services, new criteria for funding allocation, and new career paths for individuals making significant contributions to scientific best practice.

For example, one solution to the commodification of knowledge is the re-appropriation of the means of production. Until the 19th century, scientists controlled their journals entirely through learned societies. Many of these societies have now sold these publishing activities, outsourced them. When they are still independently publishing original research, they often fear open access publishing because it requires a new revenue model. Some have envisaged flipping journals, not just to meet the new open access paradigm, but also to put new governance models in place. However, the long hours put into editing, reviewing and formatting research articles written by other scientists is rarely viewed as a criterion for promotion. These tasks are crucial to the quality of academic research and should be rewarded as such.

Providing training and support

Putting open science into practice will require a continuous investment in training and support for our research communities. While librarians have been part of the research environment for a long time, there still is no equivalent for the management of digital scholarship. Training researchers how to properly generate, analyse and share their data is one crucial step towards reproducibility, but a career path for data scientists and statisticians, similar to that of librarians, has become crucial. The para-academic communities — data engineers who can write and maintain the code used to organize data, data analysts who can build models and visualizations, data stewards or information specialists — have become increasingly important to the research process and their work must be recognized.

EPFL and ETH Zürich have recently and jointly created the Swiss Data Science Center (SDSC) to make the barriers to best practice in data management and sharing as low as possible. With offices in both Lausanne and Zurich, its role will be to foster innovation in data science, catalyse multidisciplinary research and promote open science by providing tools to its users. The SDSC team will not only produce support to researchers, but also provide education at both institutions in the form of courses in data science at Master level.

Another initiative launched three years ago at EPFL is the deployment of electronic laboratory notebooks (ELN). The goal was to obtain a robust traceability of experiments and samples and to facilitate data management and further publication. Because each discipline has different requirements, there is no "one-size-fits-all" solution and researchers must have the freedom to choose the most appropriate tool. However, providing human resources to help them learn best practices has proven extremely efficient in facilitating the adoption of ELN across campus.

OPEN SCIENCE AS A GLOBAL ENTERPRISE

As a combination of both top-down and bottom-up initiatives, Open Science is a change in the way scholarship is produced, disseminated and evaluated. It represents a chance for the scientific community to increase the transparency and impact of research, as well as claim back ownership over quality control and talent management. A few years back, a report (The Royal Society, 2012) insisted on the fact that open enquiry is at the heart of the scientific enterprise. In 2017, it is time to reaffirm the global dimension of the scientific enterprise. We have to acknowledge that each of the challenges described earlier has a better chance to be tackled if institutions from around the world work together rather than in isolation.

Although one can only speculate about the reasons behind the absence of willingness to reward researchers who adopt open practices, it is likely to be due to the perceived high risk for a country or an institution to take this step and see its researchers being excluded from a competitive arena. Global and national institutions, whether they are research universities or funding agencies, need to take the following points seriously if they want to improve the quality of the research output and support general openness in science. In the coming years, EPFL will put in place initiatives that support the following actions:

- Value quality over quantity. In the spirit of the San Francisco Declaration on Research Assessment, we propose to use alternative assessment methods — i.e. other than impact factor and publication list — for evaluation, in order to promote the rights incentives and avoid outsourcing talent management to for-profit publishers.
- Increase, accessibility and visibility of all research outputs, beyond scientific articles. It has become crucial to explore novel knowledge dissemination routes and to enact sharing policies and standards that correspond to the requirements of different disciplines.
- Promote reproducibility and reuse of digital materials. Being open, in machine-readable formats and under appropriate licences is not sufficient. There must be incentives and rewards for those who create value and impact with scholarship provided openly by others.
- Support bottom-up initiatives — such as databases, journals, tools, etc. — that empower researchers by providing them with resources, training and infrastructures that enable them to share their research results.

The future of the scientific endeavour depends on it.

REFERENCES

Berlin Declaration (2003). [Online] Available from: https://openaccess.mpg.de/Berlin-Declaration [Accessed 9 August 2017].

Budapest Open Access Initiative (2002). [Online] Available from: http://www.budapestopenaccessinitiative.org [Accessed 9 August 2017].

Cambridge Economic Policy Associates Ltd. (2017). Financial Flows in Swiss Publishing. [Online] Available from: https://zenodo.org/record/240896 [Accessed 9 August 2017].

DORA (2012). The San Francisco Declaration on Research Assessment [Online] Available from: http://www.ascb.org/dora/ [Accessed 9 August 2017].

European Space Agency. (2008). ESA: Missions, Earth Observation: ENVISAT. [Online] Available from: http://envisat.esa.int/ [Accessed 3 July 2008].

Fang, F. C., Casadevall, A. & Morisson, R. P. (2011). "Retracted science and the retraction index". *Infect Immun.* 79, 3855-9.

Fecher, B. & Friesike, S. (2013). "Open Science: One Term, Five Schools of Thought." In Bartling, S. & Friesike, S. (Eds.), *Opening Science* (pp. 17-47). New York, NY: Springer.

Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. (2015). "The Economics of Reproducibility in Preclinical Research." *PLOS Biology*, 13(6): e1002165.

Ioannidis J. P. A. (2014). "How to Make More Published Research True." *PLOS Medicine*, 11 (10): e1001747.

Marcus, A. & Oransky, I. (2015). "What's Behind Big Science Frauds?" *New York Times*, 22 May.

Nature News (2014). "Why high-profile journals have more retractions," *Nature Publishing Group*, 17 September [Online] Available from: http://www.nature.com/news/why-high-profile-journals-have-more-retractions-1.15951 [Accessed 9 August 2017].

Piwowar H. *et al.* (2017). "The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles." *PeerJ Preprints* 5:e3119v1 (Online) Available from: https://doi.org/10.7287/peerj.preprints.3119v1 [Accessed 9 August 2017].

Suber, P. *et al.* (2003). *Bethesda Statement on Open Access Publishing.* (Online) Available from: http://www.earlham.edu/~peters/fos/bethesda.htm [Accessed 9 August 2017].

The Royal Society (2012). "Science as an Open Enterprise", *The Royal Society Policy Centre Report* [Online] Available from: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/ [Accessed 9 August 2017].

Woolston, C. (2015). "Long wait for publication plagues many journals", *Nature*, 523, 131.