AN ACADEMIC SPRING

by

Dymphna C. van den Boom

Speech on the occasion of the 384th *dies natalis*
of the University of Amsterdam
on Friday 8 January 2016

An academic spring

Ladies and gentlemen,

Tim Gowers is not your typical blogger. Gowers, a mathematician at Cambridge University, is a recipient of the Fields Medal. His blog radiates mathematical ideas and insight. In January 2009, Gowers decided to use his blog to run a very unusual experiment. He picked out an important and difficult unsolved mathematical problem, a problem he said he'd 'love to solve.' But instead of attacking the problem on his own, or with a few close colleagues, he decided to attack the problem completely in the open, using his blog to post ideas and partial progress. What's more, he issued an open invitation asking other people to help out. Anyone could follow along and, if they had an idea, explain it in the comments section of the blog. Gowers hoped that many minds would be more powerful than one, that they would stimulate each other with different expertise and perspectives, and collectively make easy work of his hard mathematical problem. He dubbed the experiment the Polymath Project.[1]

The Polymath Project got off to a slow start. Seven hours after Gowers opened up his blog for mathematical discussion, not a single person had commented. Then a mathematician named Jozsef Solymosi from the University of British Columbia posted a comment. Fifteen minutes later, an Arizona high-school teacher named Jason Dyer chimed in with a thought of his own. And just three minutes after that, UCLA mathematician Terence Tao – like Gowers, a Fields medalist – added a comment. The comments erupted: over the next 37 days, 27 people wrote 800 mathematical comments. Reading through the comments you see ideas proposed, refined and discarded, all with incredible speed. You see top mathematicians making mistakes, going down wrong paths, getting their hands dirty following up the mundanest of details. And through all the false starts and wrong turns, you see a gradual dawning of insight.[2]

Just 37 days after the project began Gowers announced that he was confident the polymaths had solved not just his original problem, but a harder problem that included the original as a special case. He described it as 'one of the most exciting six weeks of my mathematical life.' Since Gowers's original project, nearly a dozen Polymath-like projects have been launched.[3]

What Gowers did was to efficiently bring together the brainpower of a large group of experts. It's a great example of the potential of open science. Open science is a broad concept: it encompasses open educational resources, open access, open peer review, open methodology, open source and open data. The term was coined in 2003 by the economist Paul David[4] to

[1] M. Nielsen. *Reinventing Discovery: The New Era of Networked Science.* Princeton NJ: Princeton University Press, 2011 (p.1).
[2] Idem (p.1)
[3] Idem (p.1-2)
[4] Paul A. David. 'Understanding the Emergence of "Open Science" Institutions: Functionalist Economics in Historical Context.' *Industrial and Corporate Change*, 13, 4, pp. 571–589.

describe a movement which makes scientific research, data and discussion accessible to an interested community ranging from amateurs to professionals.

A second revolution

The advent of academic journals is also referred to as the first open science revolution. Prior to the 17th century, society forced academics to maintain a certain level of secrecy. They had little to gain, but a lot to lose from sharing their scientific breakthroughs. That all changed with the rise of the academies and scientific publications. In 1660, England established the Royal Society, and six years later the Academy of Sciences was founded in France. After this, around 70 additional scientific organisations received official recognition. In 1665, Henry Oldenburg became the editor of *Philosophical Transactions*, a publication by the Royal Society, the first journal to be exclusively devoted to science. It was the starting point for an explosive growth in academic publishing. By the end of the 17th century, there were already 30 academic journals. By around 1790 this number had increased to as many as 1052, and scientific publishing has been on the rise ever since.

We now find ourselves on the brink of a second open science revolution,[5] which will have at least as big an impact as the first. The changes it will bring go to the heart of the scientific endeavour and extend far beyond the requirement to publish research or make data publicly available.

Firstly, open science changes how people work together. Second, it fundamentally changes the traditional way of doing science. And third, it impacts on the relationship between science and society. I will discuss these three aspects in succession.

The Polymath Projects illustrate the first point, namely the ways in which scientists can go about working together differently. They demonstrate that new communication technologies have the potential to create a new social dynamic in science, as long as scientists are willing to share every step in the process and let go of the traditional way of doing things.

The second change relates to the spectacular growth in the amount of available data. Computers make it possible to create huge datasets and explore them in ways that reveal inherent, but unsuspected patterns and correlations. The computer scientist Don Swanson, who is retired but still active in the field, provides an example of this. He used Medline, a database containing millions of medical articles, and made several discoveries without having ever studied medicine in a formal setting. Among other things, he discovered a connection between migraines and magnesium by looking for correlations between migraines and other conditions. No single individual could ever grasp the millions of studies contained in the Medline database, but one individual can discern patterns, the existence of which could not have been previously suspected.

---

[3] *Science as an open enterprise*. The Royal Society Science Policy Centre Report 02/12.

Another example is the joint collection and mapping of data, such as that undertaken in the GenBank[6] by biologists, or in HapMap[7], a catalogue of genetic variations among people. Scientists are working together online in a wide range of other fields to create these types of databases – databases which chart the structure of the cosmos, the worldwide climate, languages and global flora and fauna. Using these integrated databases, everyone can use algorithms to look for connections that no one could otherwise have anticipated.

Some people refer to this as the fourth paradigm in science.[8,9] The classic duo of experiment and theory were joined by a third paradigm of science, that of simulation, after the advent of the modern computer. The fourth paradigm turned everything on its head. Rather than hypotheses being tested and developed from data collected for that purpose, hypotheses are constructed *after* identifying relationships in the dataset, on a scale far greater than hitherto possible. This not only turns the method of doing research on its head, but also brings about a sea-change in the sort of findings that are possible. That's why it represents a fourth paradigm of science.

The third big shift resulting from open science is a change in the relationship between science and society. An example is the website Galaxy Zoo, which has recruited more than 200,000 online volunteers to help astronomers classify galaxies. You can think of Galaxy Zoo as a cosmological census, the largest ever undertaken, a census that has so far produced more than 120 million galaxy classifications and 25 scientific publications based on Galaxy Zoo data. Other examples are Foldit, a computer game which challenges volunteers online to figure out how amino acids form proteins, and eBirds, which asks amateur ornithologists to record information about birds they spotted on a website. Tens of thousands of bird enthusiasts regularly use the site. Of course, the potential of this so-called citizen science is still unclear. But it's certain that the growth of the movement can bring about a wholly different dynamic in the scientific community. You could say that the boundary between professional and amateur is becoming blurred; at any rate, the nature of the public's involvement in science is changing.

I will finish with an example in which all of these strands come together: open science in real time during an outbreak of a severe gastro-intestinal infection in Hamburg in 2011.[10] This infection spread through several European countries and the US, affecting about 4,000 people and resulting in 50 deaths. All patients tested positive for an unknown variant of the E.coli

---

[6] D.A. Benson, M. Cavanaugh, K. Clark, I Karsch-Mizrachi, D. J. Lipman, J. Ostell & E. W. Sayers (2013). GenBank. *Nucleic Acid Research*, 41, Database issue.

[7] The International HapMap Consortium (2003). 'The International HapMap Project.' *Nature*, 426, 18–25 December.

[8] J. Gray (2009). 'E-Science: a transformed scientific method.' In *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Hey, T., Tansley, S. & Tolle, K. (Eds.). Microsoft Research: Washington.

[9] Shadbolt, N., Berners-Lee, T. & Hall, W. (2006). 'The Semantic Web Revisited.' *IEEE Intelligent Systems*, 21, 3, 96–101. Available at: http://eprints.soton.ac.uk/262614/.

[10] World Health Organization (2011). *Outbreaks of E. coli 0104:H4 infection*. Available at: http://www.euro.who.int/en/health-topics/disease-prevention/food-safety/outbreaks-of-e.-coli-o104h4-infection.

bacterium. The strain was initially analysed by scientists in Shenzhen, China, working together with those in Hamburg, and three days later a draft genome was released under an open data licence. This generated interest from bioinformaticians on four continents. 24 hours after the release of the genome it had been assembled. Within a week two dozen reports had been filed on an open-source site dedicated to the analysis of the strain. They produced results in time to help contain the outbreak. It's an impressive example of a new way of working together, in combination with open data and the sharing of databases.

Progressive goals, conservative means

So far I have presented you with an optimistic account of open science, but there are also obstacles that I haven't mentioned yet. Sharing data, at least at this point in time, is a strange choice, because as a scientist it's to your advantage to keep your data secret as long as possible. Why would you share your data online before you've had the chance to publish it yourself? Why would you share your hypothesis so that someone else can do the analysis? Why would you put a possible approach to a problem out there and risk helping someone else on their way?

In the scientific world it is publications that lead to jobs and promotions. That's how academia works. Making data available to the public doesn't exactly contribute to your academic career and may even impede it by helping the competitor. The system of incentives, recognition and promotion within universities and research institutes therefore presents a significant obstacle to the principles of open science achieving general acceptance.

It is therefore crucial that the production and management of important datasets and the open approach to scientific activity are recognised and rewarded. The skill and creativity that are required to successfully create large datasets also need to be acknowledged as such. Citations from open data must be placed on a par with those from conventional academic publications. New approaches such as science wikis, science blogging and open peer review also require this kind of shift in mindset. Altmetrics[11] could help with that. This term refers to a range of different ways of measuring the impact of research, approaches that don't just look at citations in journals but also take into consideration factors such as the open use and sharing of data. Although science has progressive goals, it often still uses conservative means, despite the fact that improvements to the scientific process itself should be considered to be of greater importance than individual discoveries.

A transformation in science

Ladies and gentlemen,

[11] R. Kwok (2013). 'Research impact: Altmetrics make their mark.' *Nature*, 500, 22 August.

The examples I mentioned before only form a small part of a much bigger story, a story about how open science is transforming the scientific process. Just as the first open science revolution gradually transformed the scientific process, the second revolution will radically change scientific practice over the next twenty years.

A scientific culture will have to be established in which as much information as possible is openly available. And by that I mean *all* information that is of scientific value: from raw data to computer codes and all the questions, ideas and speculations that currently aren't accessible.

That means a huge cultural shift. The importance of the first open science revolution was that it brought together the individual and the collective interest by rewarding scientists for sharing their results in academic journals. Although it is now in the collective interest of scientists to embrace new technologies, it is still in the individual interest to focus on publications. This individual interest should once again be aligned with the collective interest. Scientists cannot do this by themselves. Different institutions will have to play a role, and everyone has to adjust.

The academies need to create worldwide benchmarks for open data and metadata. Universities will have to adjust their HR policy so that the innovations generated by open science are recognised. These types of adjustments are also required where research assessment procedures are concerned. Scientists will have to be willing to make their data openly available – of course, after a period of exclusive access to their own data, and with due regard for privacy and security.[12] Ministries will have to create policy for the sharing of research data. Funding bodies will have to do the same. They can also join forces in supporting the development of the tools, software and staff that are vital for the success of this endeavour. I feel this is a topic should be treated as a key priority on the National Research Agenda, both on a content level and financially. This is in the interest of all universities in the Netherlands. With the right policy and the right resources, the Netherlands can play a leading role in this area.

We now have the opportunity to change the way knowledge is being created. Investing in this approach, and leading by example in making it part of the Dutch National Research Agenda, could pave the way for a second open science revolution in the Netherlands.

Thank you for your attention.

[12] OECD (2015). 'Making Open Science a Reality.' *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en.