A new kind of science: research in the age of big data

By Timo Hannay

Published on **EuroScientist**: www.euroscientist.com

Open science requires greater sharing and crediting for all research contributions

Call it what you like: Science 2.0, eScience or even the Fourth Paradigm. One thing is certain: research is changing profoundly. Driven by the same revolution in information technology that is transforming almost every aspect of human life, tomorrow's science will be different from yesterdays. And the shift is arguably an epochal one. Though change in academia comes slowly, we are now firmly on the road to a new digital and much more open future for research.

Fourth paradigm

To some observers of science this change ushers in a fourth great age of research fourth great age of research. To review, the first, which started over two millennia ago in Greece, was based on observation of the natural world and the formulation of qualitative theories about how it works. The second, in the 17th century, arose from the quantitative, mathematical worldview of the Enlightenment. Then, during the second half of the 20th century, the third age was associated with electronic computers, which enabled new methods of investigation, especially sophisticated numerical simulations.

The scientific age we are now entering will be every bit as disruptive and empowering as the previous three. What is sometimes called the Fourth Paradigm is characterised by the use of ubiquitous networked computers, vast quantities of data and industrial-scale research programmes involving hundreds, or even thousands of

scientists. There are particularly prominent examples in astronomy (the Sloan Digital Sky Survey), particle physics (the Large Hadron Collider), molecular biology (the Human Proteome Project HapMap) and neuroscience (the Human Brain Project). Yet, every discipline is affected in one way or another.

Like previous paradigm shifts, these advances afford new possibilities, thanks in this case to our newfound ability to accumulate and analyse vast amounts of data. Many new discoveries being made today in areas such as genomics and subatomic physics were not merely more difficult in the past, they were literally impossible. These advances will also change the way we do science and what it means to be a scientist.

Global collaboration

One such development is the increase in global collaboration. This is manifest not only in the kinds of high-profile initiatives mentioned above, but at almost every level of research. Co-authorship statistics show that in Western European countries, international collaborations rose from under 20% of research papers in 1981 to around 50% by 2011.

This is a welcome change – a reflection of the fact that researchers are increasingly able to tap ideas and expertise from around the world almost as easily as they can collaborate with someone down the corridor. It is also entirely in keeping with the global, collaborative ethos of science itself.
But this shift also brings its own challenges. One issue arising is linked to research groups becoming ever larger and the roles within them being ever more specialised. How then can we effectively assign credit to every member of the team according to his or her contribution? Witness, for example, the recently published — and much discussed — 1,000-author genomics paper.

Credit attribution

Initiatives by publishers and others to clarify and codify each researcher's contribution are welcome, but insufficient on their own. In particular, the scientific establishment, from funders and universities to societies and publishers, needs to acknowledge the essential contributions of those who gather and share data or computer code, not just those who draw out the insights and write up the papers.

Adequate level of credit is not only important in the interests of fairness. It is also a prerequisite for an even more significant change: the move to a greater culture of openness. The power of information increases when it is shared. Yet, too often researchers hold their data close to their chests, unwilling to put in the extra effort to make it useful to others. This is because they are afraid of being scooped to the only thing that currently seems to matter: a high-profile peer-reviewed paper.

This, in turn, has contributed to a corrosive combination of insufficient transparency, publication bias and, in a few high-profile cases, outright fraud. It is no exaggeration to say science itself now faces a credibility crisis. Fortunately the new digital, networked kind of science enables us to share information more freely. It also makes it possible to open up research to greater scrutiny than ever before. We can and must grasp that opportunity, and funders in particular are waking up to this fact. In time, I hope, the term ‘open science’ will come to be seen as a redundant expression.

The founding essences of science are ones of open and honest enquiry, shared observations and insights, and collective progress; stretching back through the Enlightenment to Ancient Greece. It is perhaps ironic—yet a welcome one all the same—that 21st-century technology undreamt of by our scientific forebears may bring us closer than ever to the intellectual ideals they bequeathed us.

Timo Hannay

Timo is Managing Director of Digital Science.

Photo credit: Rawpixel via Shutterstock