



Open your minds and share your results

An open approach is the best way to maximize the benefits of research for both scientists and the public, says Geoffrey Boulton.

There is a compelling case for having open access to scientific papers, to enhance the efficacy and reach of scientific communication. But important though this is, the open-access debate has drawn attention away from a deeper issue that is at the heart of the scientific process: that of 'open data'. In an attempt to focus much-needed attention on this subject, I chaired a group that produced *Science as an Open Enterprise*, a policy report from the Royal Society in London, published last week.

Open enquiry has been at the heart of science since the first scientific journals were printed in the seventeenth century. Publication of scientific theories — and the supporting experimental and observational data — permits others to identify errors, to reject or refine theories and to reuse data. Science's capacity for self-correction comes from this openness to scrutiny and challenge.

Modern techniques to gather, store and manipulate data make this more difficult. In the 1980s, I published a paper that presented seven hard-won data points showing the relationship between stress and velocity beneath a glacier. Two years ago, I was involved in an analogous experiment on the Antarctic ice sheet that created more than a billion times more data points. No journal could publish these data, so for them to be accessible, the only option was to deposit the information in a recognized repository, complete with metadata (data about data), and to signpost it in published papers, preferably through live links in the papers' electronic versions.

In the Royal Society report, we argue that this procedure must become the norm, required by journals and accepted by the scientific community as mandatory. As scientists, we have some way to go to achieve this. A recent study of the 50 highest-impact journals in biomedicine showed that only 22 required specific raw data to be made available as a condition of publication. Only 40% of papers fully adhered to the policy and only 9% had deposited the full raw data online (A. A. Alsheikh-Ali *et al.* *PLoS ONE* 6, e24357; 2011).

We also need to be open towards fellow citizens. The massive impact of science on our collective and individual lives has decreased the willingness of many to accept the pronouncements of scientists unless they can verify the strength of the underlying evidence for themselves. The furore surrounding 'Climategate' — rooted in the resistance of climate scientists to accede to requests from members of the public for data underlying some of the claims of climate science — was in part a motivation for the Royal Society's current report. It is vital that science is not seen to hide behind closed laboratory doors, but engages seriously with the public.

There is, of course, a problem in making data sets open to non-specialists. They are rarely in the form of an Excel spreadsheet, an illusion

under which many politicians labour in their laudable but problematic calls for open data. True openness requires data to be not only accessible, but also intelligible, assessable (who produced the data, what are their qualifications, do they have conflicts of interest?) and reusable.

Everyone will benefit from a more open approach. The digital and communications revolutions bring opportunities for research that demand openness and a willingness to share data. These include the assembly of massive data sets from diverse sources, and linking them to allow data integration, dynamic updating and the manipulation of data within electronic publications. Such data-led science offers ways to explore massive data sets for patterns and relationships.

Yet this, too, presents a problem. Too often, we scientists seek patterns in data that reflect our preconceived ideas. And when we do publish the data, we too frequently publish only those that support these ideas. This cherry-picking is bad practice and should stop.

For example, there is strong evidence that the partial reporting of the results of clinical trials, skewed towards those with positive outcomes, obscures relationships between cause and effect. We should publish all the data, and we should explore them not just for preconceived relationships, but also for unexpected ones. Without rigorous use and manipulation of data, science merely creates myths. At the same time, communications technologies are displacing the printed page from its dominant role as the medium of scientific communication. This is already exploiting the collective intelligence of the scientific community and shifting the social dynamic of research towards collaboration.

This shift has not been mandated by research councils, governments or national academies, but is the consequence of scientists finding more productive and creative ways to do science. Pathfinder disciplines include bioinformatics, astronomy, mathematics, nanotechnology and social and health statistics. Likewise, to extend the reach and depth of these approaches does not need top-down orchestration. It merely requires some constraints to be removed and some enabling changes to be made.

What about costs? Data curation should be viewed as a necessary cost of research. Creative data generation should be a source of scholarly esteem and a criterion for promotion. We need a revolution in the role of the science library, with data scientists supporting the management of data strategies for both institutions and researchers. We need strategic funding to develop software tools to automate and simplify the creation and exploitation of data sets. And above all, we need scientists to accept that publicly funded research is a public resource. ■

Geoffrey Boulton is regius professor of geology emeritus at the University of Edinburgh, UK.
e-mail: g.boulton@ed.ac.uk

SCIENCE'S
CAPACITY FOR
SELF-CORRECTION
COMES FROM
ITS OPENNESS TO
SCRUTINY AND
CHALLENGE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/u5lazz