

insights from people with diverse expertise.

Similar open-source techniques could be applied in fields such as theoretical physics and computer science, where the raw materials are informational and can be freely shared online. The application of open-source techniques to experimental work is more constrained, because control of experimental equipment is often difficult to share. But open sharing

of experimental data does at least allow open data analysis. The widespread adoption of such open-source techniques will require significant cultural changes in science, as well as the development of new online tools. We believe that this will lead to the widespread use of mass collaboration in many fields of science, and that mass collaboration will extend the limits of human problem-solving ability. ■

**Timothy Gowers** is in the Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK, and a Royal Society 2010 Anniversary Research Professor. **Michael Nielsen** is a Toronto-based writer and physicist working on a book about the future of science. e-mails: W.T.Gowers@dpmms.cam.ac.uk; mn@michaelnielsen.org

## Stitching science together

Google Wave is the kind of open-source online collaboration tool that should drive scientists to wire their research and publications into an interactive data web, says **Cameron Neylon**.

Science communication today remains firmly wedded to its print origins. We cling to the notion that ‘the real version’ exists on the page. Beyond ease of delivery, we take very little advantage of the potential of the World Wide Web to transform the way we store and transfer knowledge. We rarely take the opportunity to update material with new data, or to provide a record of how a document or data set has changed. Gene names and protein structures should be routinely linked to database entries through hyperlinks. The outputs of computational processes should be connected to their inputs, so analyses can be redone. If we can make these records accessible to humans and readable by machines, then whole new types of analysis will become possible, indeed standard.

Many of these things are possible today. But they are hard to achieve. Much effort has gone into solving parts of the problem, by big players such as Microsoft and Amazon as well as by smaller organizations. Electronic lab notebooks can help to capture the details of science, and databases can make it available to the user. Reference-management tools such as Delicious, semantic data stores and Wikipedia can help to wire up and monitor knowledge. But the tools are often difficult to use and don’t ‘talk’ to each other. There is no single framework that makes it easy to link all the steps of science. Scientists do their analysis and writing using different software, and prepare graphs and record data using different tools.

Very few companies worldwide have both the expertise and resources to take on the task of stitching this together. So it is with great interest that I have watched Google develop its product, Google Wave. The company describes Google Wave as “what e-mail would look like if it were invented today”. It blends elements of e-mail with instant messaging and online collaborative authoring. The big change is that the

‘document’ or ‘wave’ is shared between all the participants and updates flow in real time. You no longer need to worry about which version of a document you have e-mailed around. This is helpful for scientists, but not revolutionary. Where Wave offers a big step for science is in two other functionalities.

### Two steps forward

First, Wave introduces the idea of robots: automated agents that can be invited into a document. Robots could look through your paper checking for Protein Data Bank codes or gene names, for example, and putting in links to the databases. A robot might represent a lab instrument, adding data automatically to your laboratory record when they become available. You can easily add maps, video or three-dimensional graphics to your work using ‘gadgets’ or ‘applications’, familiar from services such as iGoogle and Facebook. Robots can interact with this information, making it

possible to have a dashboard in your inbox to monitor and control instruments in the lab.

The second step forward is using versions. Each wave maintains a record of every change. It could be possible to check each step from data collection to drawing a graph and its publication. This would allow a reader to step through an analysis to see where conclusions have come from, and would make detecting fraud — or honest mistakes — much easier.

Google has done a good thing in making the protocol and programming tools open source, enabling people to test and build. Perhaps 50 people, myself included, from experimental scientists to journal publishers, have been testing the prototype system for science applications since June, building robots that link chemical information, visualize data and format references. Since 30 September, a much bigger group has been testing. But real benefits will come only if the system is widely adopted. Perhaps a new generation of scientists will be required to exploit the power that working with these dynamic documents and tools offers.

Solving the current problems in science communication requires the intervention of strong companies such as Google. But it will take more than technical advances to provoke scientists into taking full advantage of the web. We need pressure, and perhaps compulsion, from journals and funders to raise publishing standards to the new level made possible by such tools. Google Wave may not be, indeed is probably not, the whole answer. But it points the way to tools that build records and reproducibility into every step. And that has to be good for science. ■

**Cameron Neylon** is senior scientist in biomolecular sciences at the Science and Technology Facilities Council Rutherford Appleton Laboratory, Didcot OX11 0QX, UK. e-mail: cameron.neylon@stfc.ac.uk

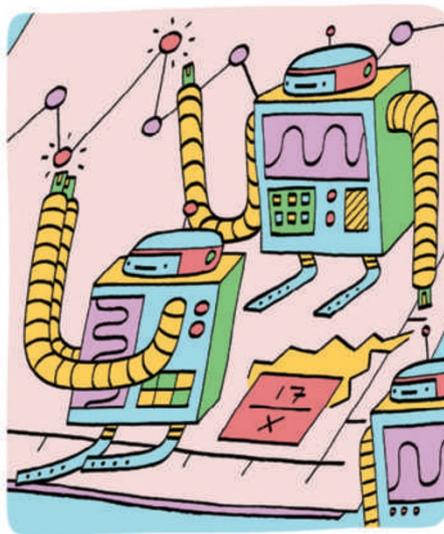


ILLUSTRATION BY M. HODSON