

# Information overload

A report released last week by the US National Academies makes recommendations for tackling the issues surrounding the era of petabyte science.

Geneticists spent more than a decade getting their first complete reading of the 3 billion base pairs of the human genome, which they finally published in 2003. But today's rapid sequencing machines can run through that much DNA in a week, and are busily churning out multiple sequences from an ever-expanding list of species. Meanwhile, astronomers working with the Sloan Digital Sky Survey telescope in New Mexico have mapped some 25% of the sky since 2000, obtaining data on more than 200 million objects. The Large Synoptic Survey Telescope, scheduled for completion atop Chile's Cerro Pachón in 2015, will gather that much data in one night.

Statistics tell a similar story in many scientific fields. This is great news for research: data glut is always better than data famine. But it is also cause for concern, because investigators' ability to amass huge quantities of data has accelerated much faster than have policies and practices for handling those data. Journal editors, in particular, have found themselves grappling with issues such as image manipulation, the preservation of original data, assuring continued access to large data sets, and standards for algorithm and code sharing.

In 2006, these concerns led a number of scientific societies and research journals, including *Nature*, to ask the US National Academy of Sciences to look at the problem. This resulted in the formation of a National Academies study committee, sponsors of which included Nature Publishing Group. The committee was headed by cancer researcher Phillip Sharp and physicist Daniel Kleppner, both of the Massachusetts Institute of Technology in Cambridge, and its report was published on 22 July (see <http://tinyurl.com/datasteward>).

The report makes 11 recommendations, organized around three major principles: integrity, access and stewardship. The integrity principle affirms that each researcher is ultimately responsible for ensuring the truth and accuracy of the data he or she produces. Individual investigators should adhere to the professional standards in

their fields, and institutions should ensure that training is in place to make this possible.

The access principle asserts the value of openness: only if results are shared can other researchers check the data's accuracy, verify analyses and build on previous work. So unless there are very good reasons for researchers to withhold data — reasons that should be publicly posted and available for comment by other researchers — they should make provisions to supply public access in a timely manner, possibly as early as their grant proposals.

Finally, the stewardship principle addresses the need for long-term preservation. Scientific societies and communities need to provide guidelines on which data are worth retaining for future analysis; institutions and funding agencies need to address and support these needs. Journals can play a part in the preservation of the published record, and in the dissemination and enforcement of guidelines. And data professionals should be recognized for their crucial role in stewardship: certainly they deserve more respect and support than researchers sometimes give them.

The authors of the report readily admit that they have provided an overview, rather than a resolution, of the complexities that surround digital data. What is needed now is for institutions, consortia and scientific societies to find individual solutions that will work in their fields and physical settings. Funders must take up their responsibilities and increase investment in the upkeep of data, from the individual grant onwards. The scientific enterprise requires that the integrity of its data forms a bond of trust with the public. It is time to strengthen that bond with action. ■

**"Each researcher is ultimately responsible for ensuring the truth and accuracy of the data he or she produces."**

## The shale revolution

The vast reserves of US natural gas must be used judiciously to ease the transition to clean energy.

Several years ago, it looked as though the United States was running short of natural gas. Prices spiked as declining production in old fields collided with increasing industrial demand. Electric utilities shifted from 'clean' gas back to cheap coal, and suppliers began building terminals to import liquefied natural gas from abroad. Yet today, coal-fired power is again on the wane, ports for liquefied natural gas are idling below capacity, and the nation is awash with gas.

So what happened? Clearly, the threat of carbon regulation has curbed industry's appetite for coal, and the sagging economy has

depressed energy demand across the board. But just as importantly, natural-gas production is again on the rise. Thanks to advances in drilling technology, including horizontal drilling and more effective rock fracturing, producers have at last unlocked the vast quantities of gas trapped underground in impermeable strata of shale.

The Potential Gas Committee, a volunteer group of industry, government and academic experts headquartered in Golden, Colorado, increased its estimate of recoverable gas reserves by 39% in its biennial report released last month, mostly because of shale gas. The new total, almost 60 trillion cubic metres, is equivalent to about a century's worth of gas at current usage rates.

Policy-makers everywhere should take note. Shale formations similar to those that have upended the US natural-gas market exist all over the world. Early explorations are already under way in Canada and several European countries, many of which are overly reliant on