

# Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.

**M**ore and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community. Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries — especially when data are combined from many laboratories and analysed in ways that the original researchers could not have anticipated.

All but a handful of disciplines still lack the technical, institutional and cultural frameworks required to support such open data access (see pages 168 and 171) — leading to a scandalous shortfall in the sharing of data by researchers (see page 160). This deficiency urgently needs to be addressed by funders, universities and the researchers themselves.

Research funding agencies need to recognize that preservation of and access to digital data are central to their mission, and need to be supported accordingly. Organizations in the United Kingdom, for instance, have made a good start. The Joint Information Systems Committee, established by the seven UK research councils in 1993, has made data-sharing a priority, and has helped to establish a Digital Curation Centre, headquartered at the University of Edinburgh, to be a national focus for research and development into data issues. Other European agencies have also pursued initiatives.

The United States, by contrast, is playing catch-up. Since 2005, a 29-member Interagency Working Group on Digital Data has been trying to get US funding agencies to develop plans for how they will support data archiving — and just as importantly, to develop policies on what data should and should not be preserved, and what exceptions should be made for reasons such as patient privacy. Some agencies have taken the lead in doing so; many more are hanging back. They should all be moving forwards vigorously.

What is more, funding agencies and researchers alike must ensure that they support not only the hardware needed to store the data, but

also the software that will help investigators to do this. One important facet is metadata management software: tools that streamline the tedious process of annotating data with a description of what the bits mean, which instrument collected them, which algorithms have been used to process them and so on — information that is essential if other scientists are to reuse the data effectively.

Also necessary, especially in an era when data can be mixed and combined in unanticipated ways, is software that can keep track of which pieces of data came from whom. Such systems are essential if tenure and promotion committees are ever to give credit — as they should — to candidates' track-record of data contribution.

Who should host these data? Agencies and the research community together need to create the digital equivalent of libraries: institutions that can take responsibility for preserving digital data and making them accessible over the long term. The university research libraries themselves are obvious candidates to assume this role. But whoever takes it on, data preservation will require robust, long-term funding. One potentially helpful initiative is the US National Science Foundation's DataNet programme, in which researchers are exploring financial mechanisms such as subscription services and membership fees.

Finally, universities and individual disciplines need to undertake a vigorous programme of education and outreach about data. Consider, for example, that most university science students get a reasonably good grounding in statistics. But their studies rarely include anything about information management — a discipline that encompasses the entire life cycle of data, from how they are acquired and stored to how they are organized, retrieved and maintained over time. That needs to change: data management should be woven into every course in science, as one of the foundations of knowledge. ■

**"Data management should be woven into every course in science."**

## A step too far?

The Obama administration must fund human space flight adequately, or stop speaking of 'exploration'.

**A**fter the space shuttle *Columbia* burned up during re-entry into Earth's atmosphere in 2003, the board that was convened to investigate the disaster looked beyond its technical causes to NASA's organizational malaise. For decades, the board pointed out, the shuttle programme had been trying to do too much with too little money. NASA desperately needed a clearer vision and a better-defined mission for human space flight.

The next year, then-President George W. Bush attempted to supply that vision with a new long-term goal: first send astronauts to build

a base on the Moon, then send them to Mars. This idea immediately set off a debate that is still continuing, in which sceptics ask whether there is any point in returning to the Moon nearly half a century after the first landings. Why not go to Mars directly, or visit near-Earth asteroids, or send people to service telescopes in the deep space beyond Earth?

Yet that debate is both counter-productive — a new set of rockets could go to all of these places — and moot, because Bush's vision never attracted the hoped-for budget increases. Indeed, a blue-ribbon commission reporting to US President Barack Obama this week (see page 153) finds the organizational malaise unchanged: NASA is still doing too much with too little. Without more money, the agency won't be sending people anywhere beyond the International Space Station, which resides in low Earth orbit only 350 kilometres up. And even the ability to do that is in question: Ares I, the US rocket that would return