nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2009

Data's shameful neglect

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

More 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 being 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.

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

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. ■

A step too far?

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

After 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-riband 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

If Algenol gets the grant, it will construct a pilot plant with Dow at Dow's manufacturing site in Freeport, Texas, with the goal of capturing industrial carbon dioxide and producing alga-derived ethanol to generate ethylene, a building block for plastics.

Meanwhile, Sapphire Energy has garnered more than $100 million from bigwig investors, including Gates's Cascade Investments and the Rockefeller family's venture-capital firm Venrock. Sapphire is using genetic engineering to boost several algal traits, including improved protection from predators and low-cost harvestability. It is also working to genetically manipulate the algae to produce oils that are nearly identical to crude oil as extracted from the ground.

And Solazyme's contract with the Navy is the first contract anywhere to manufacture commercial-scale quantities of next-generation biofuels. The contract requires that Solazyme deliver some 75,000 litres of F-76 renewable fuel, which is similar in composition to diesel fuel, over the next year. "This really raises the bar in what constitutes a true production capability versus an interesting research direction," says Dillon.

Still, many challenges remain. In May GreenFuel Technologies, a front-runner on the algal scene that had amassed some $70 million in investments since 2001, announced that it was closing down. Sam Jaffe, an energy analyst with IDC Energy Insights, a research and analysis firm based in Framingham, Massachusetts, says that GreenFuel pursued too many different technologies, including expensive greenhouses to control algal growth conditions. "Growing algae is easy," says Jaffe. "Growing it as a business and making money off of it is about getting the costs down."

One of the biggest challenges is to reproduce laboratory conditions on a large scale. In the lab, it can be easier to control algal growth and to find strains that produce copious amounts of oil. "But it's a totally different story when you take this organism that behaves well in the laboratory and you put it in acres' worth of outdoor ponds," says Darzins. For this reason, some companies have opted to grow their algae in enclosed 'bioreactors'. But the costs of building bioreactors can be prohibitively expensive. The algae community is "still torn" between open ponds and closed bioreactors, Darzins says.

With so much enthusiasm and investor interest in algal technology, new companies have sprung up almost overnight. Some experts say that because much of the science behind these technologies is not peer reviewed and is done through privately held companies, it can be difficult to gauge their progress. "On the one hand you get their hype, and on the other hand they're guarding everything so closely that you can't evaluate it," says Martha Groom, a conservation biologist at the University of Washington in Bothell. "I find that fairly frustrating."

Experts say that a few companies have made questionable assertions about how much fuel they can reap from their algae. "Unfortunately, a lot of people tout these technologies and yet don't have the production data to back it up," says Doug Henston, chief executive of Solix Biofuels, a renewable-energy company based in Fort Collins, Colorado, that opened an algal oil-production demonstration facility in July at a coal-bed methane plant in southwestern Colorado. "That's the unfortunate case because it clouds the picture and builds unrealistic expectations," he says. Solix hopes to push its production capacity from its current rate of about 14,000 litres per hectare per year to between 37,000 and 47,000 litres per hectare per year. However, some start-ups have claimed that they can reach oil-production capacities as high as 900,000 litres per hectare per year, which, says Henston, is "thermodynamically impossible".

Fuel source of the future?

E. DRAPER/*THE NEW YORK TIMES*/REDUX/EYEVINE

Dillon, of Solazyme, says that the recent involvement by big-league investors, oil giants and the US military will help sort out approaches that are leading somewhere from those that aren't. "I think it's a good thing that we've got some real expectations coming on," he says. "There's been a lot of hype. That has a time window on it, and that type of time window tends to close when major players with real expectations start getting involved." ■

Amanda Leigh Mascarelli

NEXT WEEK: CELLULOSIC ETHANOL

Correction
The Editorial 'Data's shameful neglect' (*Nature* 461, 145; 2009) stated that the Joint Information Systems Committee was established by the seven UK research councils. It was, in fact, established by the three Higher Education Funding Councils.