

# Leading by example

The primary safeguard against scientific misconduct is the example set every day by thousands of senior researchers in the laboratory.

Everybody likes a good scandal, and there is nothing like a fresh allegation of research misconduct to set tongues wagging in the scientific community and outside it.

There is broad agreement within the community on two main points regarding outright scientific fraud: it is rare, and it is serious. This unanimity frays at the marginal territory lying just beyond falsification and fabrication of data, however, where 'misconduct' may involve hogging the limelight for one paper too many, or tardy record-keeping at the laboratory bench. But even here, there is a reasonable consensus that scientists in every nation and every discipline will do the best for both themselves and their work by keeping all forms of substandard conduct to a minimum.

It has sometimes been argued that this is so because the credibility of the scientific enterprise is on the line. But if the truth be told, the smattering of egregious misconduct cases that have attracted widespread attention — the most recent being that of South Korea's Woo Suk Hwang — have made little ill impression on the generally favourable public attitude to science as a whole.

It is in fact the reliability of science — or to be more specific, that of the scientific record in the literature — that is the principal victim of scientific misconduct. No aircraft are going to fall out of the sky because somebody fudged a graph in a materials-science journal. Rather, the problem is the time wasted by other scientists chasing up minor and not-so-minor infractions in pursuit of their own research.

A range of tools are available to help minimize research misconduct (see page 240). The most obvious are the systems of investigation and sanction that have been established in many nations. Rather less obvious, but perhaps just as useful, are the methods being developed to secure a scientific environment that is conducive to the highest standards of research conduct.

The thorough and fair investigation of misconduct allegations is a necessary activity that is fraught with difficulty, and likely to remain so. In the United States, the Office of Research Integrity at the health department and the Inspector General's office at the National

Science Foundation have between them overseen such investigations at universities for more than a decade now. But every investigation is a new and testing experience for the institution involved.

Parallel developments have occurred in Europe, especially in Germany, where the creation of an ombudsman system at the DFG, Germany's main research agency, in 1999 has provided useful impetus for fair and thorough university-based investigations. Other nations, including Britain, still lack a national mechanism responsible for this — leaving institutions to fend for themselves, with uncertain results.

A more aggressive approach is being taken in east Asia, where major universities have responded sharply to recent misconduct allegations, constituting a major improvement on their past reluctance to act. However, it is not at all clear that this approach will extend beyond high-profile cases that have already found their way into the public domain — or that adequate checks are in place to assure the rights of the accused.

Attention should certainly be paid to nurturing the right framework for investigation and to setting up appropriate sanctions in an effective, cross-institutional manner. But hell will freeze over before an investigatory system can be developed that is universally recognized as comprehensive and fair.

In the interim, the scientific community should continue to concentrate on developing an environment that is inherently consistent with minimizing scientific misconduct.

A respectable level of ethics training for all postgraduate students is an important element of this. It needs to be introduced at all research universities — alongside stricter rules on record-keeping, and arrangements for protecting whistleblowers, where this is missing at the national level.

But most important of all, as the first scientific studies of the factors behind good conduct confirm, is the example set by senior researchers themselves. It is here in the laboratory — not in the law courts or the offices of a university administrator — that the trajectory of research conduct for the twenty-first century is being set. ■

## The database revolution

Funding agencies face conflicting challenges in supporting the databases essential to modern biology.

In molecular biology and genetics, high-throughput data gathering has progressed in a short time from a novel strategy of the few to a standard approach of the many.

Much of these data are not collected in a way that guarantees their accessibility to other researchers for future analysis. The best databases, however, are online agglomerations of raw and processed data,

analytic tools and insights that have been collectively dubbed 'knowledge environments'.

Funding agencies are slowly adjusting their priorities to reflect the importance of consistent, assured support for some of these databases. The US National Institutes of Health (NIH), for example, has recognized that it must provide guidance for its research sections and grantees so as to maximize the return, in both knowledge and public health, on its investments in biomedical research.

Which strategies best support the collection, analysis and dissemination of large databases of related information? Do knowledge environments promote the aims of the funding agencies? At a meeting in Bethesda, Maryland, last month, it was clear that the NIH is

struggling to find a middle road between two diametrically opposed approaches to the development of such databases. Top-down pressure by the agency on researchers to use certain software or formats would probably impede their development. But a bottom-up strategy that merely encourages cross-project cooperation, while allowing researchers total freedom to devise their own databases, is bound to be chaotic, does not guarantee cross-compatibility of data, and ultimately reduces the likelihood that their contents will be used to maximally benefit research.

What is clear is that individual labs can no longer make much progress alone. Currently, many researchers feel they are drowning in data. For all they know, a database might contain answers to patient safety issues or glimmers of new therapeutics — but this is being lost through an inability to effectively harvest the data already available. Other opportunities are missed because both experts and data are 'silo'-ed in isolated and often inaccessible systems. On top of these issues is the fact that neither databases nor the experts that create them are permanent or inseparable.

The NIH and its equivalent agencies elsewhere in the world are now turning their attention to working out how best to assist the growth of validated and accessible databases. This should involve, at the least, development of policies for evaluating proposals on databases and associated analytic tools, for their sustained funding, and for ensuring that the data deposited remain accessible long after the project originators have moved on.

The NIH itself, if it chose, could aim for something grander. It could take it upon itself to define a broad reference model and the basic architecture for knowledge environments. It could even build a centralized warehouse with Google-like storage, a veritable National Biomedical Resource of raw data and the tools to access and analyse them. The challenge then would be how to store, retrieve and read data created on multiple systems by diverse research groups.

But perhaps it will prove more realistic for the US agency to concentrate on improving the inter-operability of databases, rather than

pushing for their merger, and to provide incentives for building in 'joins' from the start. The NIH should work on this with industrial companies and other government agencies.

All funders should also be aware of the need to support viable career paths for the software engineers and bioinformaticians who create the knowledge environments and curate the data in them. And in order to obtain value for money, it will be vital for funding agencies to carefully select the databases they choose to support and then to support them for the long term.

They must encourage the sustained availability of these data and build incentives for the development of cross-querying capability.

As things stands, even popular databases such as PhysioNet ([www.physionet.org](http://www.physionet.org)), which is facing loss of NIH funding after seven years in development, often lack secure funding. In order to get value, the NIH and other funding agencies need to develop policies for such knowledge environments that will allow long-term support of successful ones.

Researchers also need stronger incentives to sustain their own participation in building knowledge environments. At a minimum, contributors should receive a citable acknowledgement of depositions. Leadership and trust are required to ensure that primary researchers personally benefit from storing their data in open databases.

Michael Huerta, associate director of the National Institute of Mental Health, told those at the Bethesda meeting that the NIH is keen to get more mileage out of high-throughput data acquired with public funds. But that will mean that the agency should require the databases it supports to be equipped with tools and protocols that researchers find convenient to use. Ultimately, these databases must show that their data are being used by other groups, and generating work of publishable quality. ■

**"To obtain value for money, it will be vital for funding agencies to carefully select the databases they choose to support and then to support them for the long term."**

## Capitol calling

A fresh perspective on science and politics.

Scientists typically master the requirements of their vocation early on — delivering a lecture, writing a grant proposal, setting up a lab. And yet even some of the most accomplished scientists remain largely in the dark about the most basic information underpinning their work: the circumstances surrounding the decision to pay them to do the science in the first place.

The world leader in the public funding of science is, of course, the United States. And America's purse-strings are held not by the president, but by the 435 representatives and 100 senators who make up the US Congress. Many scientists have only the vaguest idea about how that much-maligned institution does its business.

On page 248 of this issue, we introduce a new column that is intended to rectify that and to give readers of *Nature* a fuller insight into how and why different topics take their place in America's research and development agenda. Until the end of last year,

the column's author, David Goldston, served as chief of staff for the House Committee on Science, helping to set the legislative agenda for science and technology before Congress.

For the past 20 years, Goldston has been at the heart of forging US science and environmental policy. He is now leaving Capitol Hill on the retirement of his long-time boss, Representative Sherwood Boehlert (Republican, New York). In addition to writing exclusively for *Nature*, Goldston will serve as a visiting lecturer in science and environmental policy at the Woodrow Wilson School of Public and International Affairs at Princeton University, New Jersey.

Goldston's column, Party of One, will provide a seasoned insider's take on the interaction between science and politics. From stem cells and energy policy to climate change and space exploration, it will provide a unique perspective on why things work, or don't work, in Washington DC. And stripped of the congressional staffer's customary anonymity, Goldston will be able to offer his own candid opinions on what's really up, on Capitol Hill. As a result, we are confident that our readers in the United States and around the world will gain a valuable insight into some of the forces that help shape science today. ■