

Let data speak to data

Web tools now allow data sharing and informal debate to take place alongside published papers. But to take full advantage, scientists must embrace a culture of sharing and rethink their vision of databases.

Upload and share your raw data, and have a high impact factor for your blog — or perish? That day has not yet come, but web technologies, from personal publishing tools such as blogs to electronic laboratory notebooks, are pushing the character of the web from that of a large library towards providing a user-driven collaborative workspace (see page 547).

This will in turn expose many fields of research to changes that are already sweeping disciplines such as bioinformatics and high-energy physics. A decade ago, for example, astronomy was still largely about groups keeping observational data proprietary and publishing individual results. Now it is organized around large data sets, with data being shared, coded and made accessible to the whole community. Organized sharing of data within and among smaller and more diverse research communities is more challenging, owing to the plethora of data types and formats.

A key technological shift that could change this is a move away from centralized databases to what are known as 'web services'. These are published interfaces that serve to simplify access to data and software (for an example of such services in action, see www.ebi.ac.uk/xembl/index.html). Until recently the preserve of expert programmers, such interfaces now mean that anyone with even a basic knowledge of programming can automate data processing and analysis.

Various sorts of data are increasingly being stored in formats that computers can understand and manipulate, allowing databases to talk to one another. This enables their users quickly to adapt technologies to extract and interpret data from different sources, and to create entirely new data products and services.

In biodiversity research, for example, rather than creating centralized monolithic databases, scientists could tap into existing databases wherever the data are held, weaving together all the relevant data on a species, from its taxonomy and genetic sequence to its geographical distribution. Such decentralization also helps to solve the problem that databases are often the fruits of individual or

lab research projects that are vulnerable to the vagaries of funding, and to people and labs moving on to pastures new.

Although discipline-specific databases have an indisputable role, science also needs to capitalize on large common repositories for data, whose preservation is guaranteed, and where the data can easily be used by anyone. If that sounds utopian, consider OurMedia, a service created by the Internet Archive and the Creative Commons, which allows anyone to store and share permanently and free of charge any digital work — even their videos and holiday photos. And last month Google launched Google Base, which also allows anyone to upload anything to its massive platform.

Such services will also require new thinking on open data. Web services are dependent on computers being able to freely access data in real time. Although GenBank and many large databases allow unhindered access to their data, many research organizations still cling to obsolete manual data permission policies, which prevent their data being used by web services.

Scientists may be justified in retaining privileged access to data that they have invested heavily in collecting, pending publication — but there are also huge amounts of data that do not need to be kept behind walls. And few organizations seem to be aware that by making their data available under a Creative Commons licence (see <http://creativecommons.org/license>), they can stipulate both rights and credits for the reuse of data, while allowing its uninterrupted access by machines.

As web services empower researchers, the biggest obstacle to fulfilling such visions will be cultural. Scientific competitiveness will always be with us. But developing meaningful credit for those who share their data is essential, to encourage the diversity of means by which researchers can now contribute to the global academy. ■

"By making data available under a Creative Commons licence, scientists can stipulate rights and credits for the reuse of data."

Life at the edge

Successes in structural studies of membrane proteins deserve to be celebrated.

Sealed membrane systems are a defining feature of cellular life. Membranes provide a barrier between the cell and its external environment and, in many organisms, divide the interior of the cell into functionally distinct compartments. The barrier, comprising lipids that are impenetrable to electrically polarized molecules, has proteins inserted within it that allow the selective transport of

ions and molecules. These proteins enable cells to ingest nutrients, excrete metabolic waste, sample the environment for the sake of the immune system, and store energy by means of ion electrochemical gradients. They mediate molecular signalling across the barrier. And they are the very devil to study.

Genome sequencing projects have highlighted the central role of membrane-linked processes in cells. They have revealed that membrane proteins represent about a third of the gene products in most organisms. Unfortunately, our molecular knowledge of these membrane proteins lags far behind that of proteins found in the cell cytoplasm and in external environments. This is primarily due to the difficulty in obtaining high-resolution structural information on

which to build a mechanistic understanding. For example, the purification of membrane proteins for structure determination requires them to be removed from their native membrane environment using detergents. This renders the proteins less stable.

Notwithstanding this and other technical obstacles, isolated successes in the determination of membrane protein structures were reported as early as 1977 for bacteriorhodopsin, the light-powered ion pump in the membranes of archaebacteria, and 1983 for the photosynthetic 'bacterial reaction centre'. However, it is only in the past five years that significant numbers of membrane protein structures have been determined, including structures of ion channels, most components of the mitochondrial and photosynthetic electron transfer chains, and proteins that mediate the transport of small molecules across membranes.

Given this progress, we can now be said to be entering the golden age of membrane protein structure. A flavour of the excitement of membrane proteins can be obtained in many of the articles in the Insight on membrane biology in this issue of *Nature* (see page 577) and by reading the landmark paper on the aquaporin structure embedded in a lipid bilayer (see pages 633 and 569).

But what are the reasons for this recent explosion in membrane protein structures? It is partly driven by technological innovation — the availability of microfocussing synchrotron beamlines suitable for data collection from small crystals, advances in the ability to express membrane proteins to high levels and, increasingly, the use of high-throughput screening methods.

In order to obtain crystals one needs to stimulate cells, often of a different organism, to generate, or 'express', large amounts of the required protein. Difficulties in expressing eukaryotic membrane proteins remain the most significant bottleneck in the field. Expressing such proteins in bacterial cells has not been achieved, for reasons that are unclear, and researchers are turning to alternative systems such as insect and yeast cells to obtain their proteins of interest.

But this remains a challenge: milligram quantities are typically required. These challenges in expression, coupled with the difficulty in obtaining diffraction-quality crystals, mean that there is no guarantee of success.

It can take several years, in many cases longer than the standard three-year postdoctoral contract, so embarking on such a quest is a risky business.

Despite the difficulties, there is one area of structural biology where membrane proteins are at the cutting edge: protein structure prediction. Membrane protein structures should be easier to predict than those of water-soluble proteins because the structural possibilities are constrained by the membrane environment.

The prospects for membrane biology are bright, not only thanks to technical breakthroughs but also because of a sense of adventure. The field's success is due in part to the willingness of scientists to dedicate their careers to this challenging endeavour. ■

"The prospects for membrane biology are bright, not only thanks to technical breakthroughs but also because of a sense of adventure."

Stem-cell probe needed

South Korea would benefit from investigating what went wrong in its leading stem-cell lab.

Last week, Woo Suk Hwang of Seoul National University finally admitted using eggs donated by graduate students and paid donors in his embryonic stem-cell research (see page 536). The admission raises pointed questions of the stem-cell research community worldwide and of the South Korean government and media. Each of these groups should be asking themselves why it has taken them so long to take this matter seriously.

In the stem-cell research world beyond South Korea, the adverse publicity generated by Hwang's decision to resign as head of the World Stem Cell Hub should serve as a reminder — as if one was really needed — of the importance of transparent and stringent ethical behaviour by practitioners in this field.

Most of Hwang's international colleagues were slow to accept that anything was amiss in his laboratory. Even as the Korean authorities failed to properly investigate the allegations first made in this journal 18 months ago (*Nature* 429, 3; 2004), researchers seemed almost universally eager to establish fresh collaborations with the laboratory and to laud the quality and integrity of its work.

The withdrawal of collaboration by Gerald Schatten of the University of Pittsburgh on 12 November and Hwang's subsequent confession on 24 November are likely to quench some of that enthusiasm. These events have also opened the doors for South Korea's

media to start looking more closely at the situation, something they were unwilling to do last year.

There are already signs, however, that some people in Korea are drawing the wrong conclusions from the episode. The government has promised to maintain financial support for Hwang, and the organizers of the stem-cell hub say they will refuse his resignation. The Munhwa Broadcasting Corporation, a Korean television network that broadcast a documentary critical of Hwang last week, has been hit by accusations that it is being unpatriotic, and several of the corporation's advertisers have dropped their accounts. On Internet chatboards, producers of the programme have been threatened with violence, and demonstrators have gathered outside the company's Seoul headquarters.

Roh Moo-hyun, the country's president, added his own thoughts to his presidential website on 27 November, calling for calm and branding the actions against the broadcasting corporation "absurd".

Many in Korea still argue that Hwang's ethical lapse was not a serious violation, but merely reflects a difference in culture between Korea and the West. But how can we know what ethical violations — however they are defined — took place, given that there have been so many obstacles to finding the truth? The Korean national interest would best be served not by more flag-waving, but by the rigorous, official inquiry that has yet to be instigated into exactly what went on at Hwang's lab. ■

"How can we know what ethical violations took place, given that there have been so many obstacles to finding the truth?"