

Sacrifice for the greater good?

Genome sequencers intend that their community and others should deposit data in community databases immediately, even if it risks the loss of publishing priority. But enforcement of this ideal could be a step too far.

Some US and UK genomics researchers are seeking to extend their principles of open access throughout the world of biology in unprecedented fashion. They claim that enforcing these principles would be in the best interests of science, and they may be right. But anybody believing that researchers in other disciplines and countries are ready to sign up would be wrong.

The principles concern databases — not just any databases, but those that are so widely used as repositories and sources that they are seen as ‘community resources’. A closed meeting in Fort Lauderdale last month explored the conflicts for the originators of genomics data between, on the one hand, the principles of immediate deposition and unconditional access, and on the other, the need to protect the originators’ rights to publish, and gain appropriate priority for, the outcomes of their labours.

This dilemma has been hanging over the community for some time, and has led to failed attempts to establish a set of ground rules to which the community could sign up. The problem has occasionally hit the headlines — for example, in a debate between the sequencers of the genomes of the malaria parasite *Plasmodium falciparum* and *Trypanosoma brucei*, which causes sleeping sickness (see *Nature* **405**, 601; 2000), and in conflicts over early use of the sequence data of the protozoan *Giardia lamblia* (see *Science* **295**, 1206; 2002). And the genome sequencers can cause problems themselves: sometimes they sit on data for an unreasonable period. The rat genome highlights another problem with immediate openness: the first pre-publication assembly has numerous errors, but is already deposited in public databases. Researchers who rush in and use the data may come to regret doing so.

A new attempt to establish ground rules, backed by the Wellcome Trust and the National Human Genome Research Institute (NHGRI, part of the US National Institutes of Health), is more radical than previous efforts, not only in the rules themselves but also in their enforcement and scope. So much is clear from proposals that form part of a report of the Fort Lauderdale meeting, due to be published this week on the Wellcome Trust’s website. They would remove all restrictions on the use of genome data.

Cause for concern

Deposition of data on an open-access website is, rightly, not considered to be equivalent to peer-reviewed publication. Under the new proposals, the originators of such data, if publicly funded, are in effect giving up any protection of their rights to claim priority in a publication. Anybody can download the data and publish whole-genome analyses, even if they scoop the originators. Attempts to impose licensing agreements to prevent such use (previously described by some as data misappropriation or, more bluntly, piracy) would henceforth be forbidden, the penalty being that centres attempting to impose such licences would be ineligible for funding.

The proposals are already causing concern in other countries (see page 877), where the principles of immediate and unrestricted access so successfully adopted by the international Human Genome Project are still not readily accepted. Even in Britain and the United States, not everyone is on board. After all, these principles are all fine and

dandy for the grandees of the field, who have little to lose in funding and reputation, but are tough on others who are less favoured.

The proposals are long on supporting data release, backed by funding-agency pressure, but short on any means of enforcing the other central requirement: that everyone should behave honourably and ensure that the originators of data get due credit. *Nature* is all too aware that under the pressure to publish, collegiality and seamliness frequently go out of the window.

To be fair, the proposals seek to compensate originators; they propose that genome sequencers publish statements of intent and a project description containing the scope of data and analysis that the originator expects to undertake. This would be a citable means of giving them credit in case they are scooped at the other end of the process. This is a good idea as far as it goes: one potential example has already appeared (M. V. Olson & A. Varki, *Nature Rev. Genet.* **4**, 20–28 (2003); doi:10.1038/nrg981). But will it compensate for the loss of incentive for researchers to stay in the business of data generation if they lose priority protection? Researchers now have an opportunity to feed in their views before the NHGRI’s advisory council considers the proposals for its endorsement in May.

Broader community

Most mammalian sequencing projects have adopted principles of openness and unconditional access, whereas microbial and plant sequencers have been less forthcoming. The new proposals potentially project themselves beyond genomics: they suggest that any project where a ‘community resource’ is the objective should subject itself to the same principles of immediate release without conditions. The Fort Lauderdale meeting focused on genomics, so it remains to be seen whether other branches of biology data generation will accept the fundamental principle espoused: that science will benefit from immediate openness in advance of publication even if data originators lose out. But whatever the discipline, the idea that funding centres should insist on the adoption of such principles smacks of coercion, and may drive sequencers away from public funding.

What of the role of journals? Some referees have refused to review papers from groups that have not deposited data in databases at the time of submission. That is their right, but it is the editors’ job to steer around such indirect imposition of principles. *Nature* has received papers that might have scooped genome sequencers, but judged them to be inappropriate for scientific reasons; they were published elsewhere. We expect that to change — such situations will arise more and more often as bioinformatics improves.

Editors and peer reviewers must do what they can to ensure that sufficient credit is given to originators of data, and we will generally check papers with originators anyway to ensure that the data are reliably deployed. We will try to ensure that licensing agreements (for as long as they continue) are not contravened. And we will support originators’ rights to publish, and be generous in interpretation as to whether a previous publication has scooped them or not. But if a good piece of whole-genome analysis arrives on our desks, we’ll publish it whoever it comes from. That, as the Fort Lauderdale proposals also say, is in the best interests of science. ■