



## Funders must encourage scientists to share

*To realize the full potential of large data sets, researchers must agree on better ways to pass data around, says Martin Bobrow.*

How can we make best use of the vast amounts of data on genomics, epidemiology and population-level health being collected by researchers? Maximizing the benefits depends on how well we as a scientific community share information.

The Human Genome Project set strong precedents for rapid pre-publication data sharing, and all biological research has benefited enormously from this approach. Most research-funding agencies, and most scientists, now agree that research data should be shared — provided that those who donate their data and samples are protected. This approach is strongly advocated by organizations such as the Global Alliance for Genomics and Health. But data sharing will work well only when it is streamlined, efficient and fair. How can more scientists be encouraged and helped to make their data available, without adding an undue administrative burden?

I chair an expert advisory group on data access that has examined this question. As part of our work, we surveyed current practices and questioned *Nature* readers. We saw plenty of good practice — in the UK social-sciences community, for example — but also significant inefficiencies. Both those who generate data and those who want to use them expressed frustration at the way that data-access processes are frequently opaque.

At present, mechanisms for data sharing are too often an afterthought. Access protocols are set up and managed differently from study to study, and this adds to the administrative burden for both producers and users. No one wins in this scenario, least of all those who donate their personal data.

Today, we publish our recommendations (see [www.wellcome.ac.uk/EAGDA](http://www.wellcome.ac.uk/EAGDA)). They are aimed at research funders, who are best placed to implement them. But we hope that all researchers will find them useful. A key recommendation is that data-access plans should be integral to the grant-application process. Researchers should set out what they regard as a reasonable process for governing and managing access, including estimates of the costs of making the data visible and available to other researchers. The review process should advise on this and the data-access plan should be an integral, auditable part of the funded grant.

Generally speaking, bigger studies will need more-substantial processes. Small experimental studies may reasonably do no more than make their data available on request, after the time to prepare for publication. Very large studies require a more formal data-access plan from their inception.

Many epidemiological or genomic studies establish data-access committees (DACs) to manage data release. It makes little sense for each to do this in isolation, with individual processes and policies. The information required by DACs,

and the undertakings they ask of potential data users, are usually similar across studies. Where possible, funders should encourage the streamlining and standardization of this process, while allowing for the fact that studies have their own characteristics. It would be helpful, where possible, to introduce common application forms and adjudication processes, and to allow new studies to make use of or consolidate with existing DACs. Access procedures should be made more transparent and straightforward by including an independent appeals process for settling disputes over access requests.

Protecting research participants is sometimes cited as a reason for withholding data. The risk that research participants could be re-identified from shared data must be carefully assessed, particularly when data sets are linked in novel ways. But safeguarding participants' identity should not require a complex or opaque system of data access, as often seems to be the case.

It is easier to protect subjects if researchers build data access into their studies from the beginning. Participant consent forms, for example, should be designed with data sharing in mind — granting permission for de-identified personal information to be shared safely with researchers outside the study group.

It is reasonable for scientists to impose certain conditions or restrictions on the use of their hard-earned data sets, but these should be proportionate and kept to a minimum. Justifiable conditions can range from requiring secondary users to acknowledge the source of the data in publications, to stipulating a fair embargo time on the use of new data releases. Whatever the conditions

imposed, they need to be presented clearly to data users.

Criteria used to judge academic careers still focus heavily on individual publication records and provide little incentive for wider data sharing. Scientists who let others use their data deserve reward too.

To build trust, any significant breaches of data- and material-transfer agreements should be treated seriously, with appropriate sanctions being imposed, such as prevention of future access to data sets, or forcing the withdrawal of a published paper.

Funders should expect that each data set they support will be made accessible unless there are particular, agreed reasons for it not to be. Science is increasingly a joint, international and collaborative enterprise. The emphasis now must be on encouraging scientists, with support and resources from funders, to voluntarily make their data more readily available to others. ■

**Martin Bobrow** is emeritus professor of medical genetics at the University of Cambridge, UK, and chair of the Expert Advisory Group on Data Access.  
e-mail: [mb238@cam.ac.uk](mailto:mb238@cam.ac.uk)

**SAFEGUARDING  
PARTICIPANTS'  
IDENTITY  
SHOULD NOT REQUIRE  
A COMPLEX OR OPAQUE  
SYSTEM OF  
DATA ACCESS.**

► **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/wdw9cq](http://go.nature.com/wdw9cq)