

nature medicine

Supporting data

Everybody agrees that ensuring the integrity and accessibility of research data is crucial for scientific progress. Agreeing on the best way to do so is the hard part.

Technological advances have enabled researchers to tackle questions that involve generating vast amounts of data. Data generation at this scale poses a series of challenges concerning data analysis, manipulation, annotation, sharing and storage that researchers, institutions, funders and journals have not yet fully grasped. How should data be annotated before being stored in a database so that it can be as useful as possible to other researchers? Should data-sharing requirements be extended to the computer codes that were used to analyze the data? Who should have access to the data, and who pays for data storage and management?

These questions will become more pressing as further technological advances make it even easier to produce ever larger data sets, and it won't be simple come up with the answers. Last year, the US National Academy of Sciences, the National Academy of Engineering and the Institute of Medicine published a report called *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*, which provides a useful framework around which to organize what has become an urgent dialogue.

The report focuses on three aspects of the problem: data integrity, access and long-term preservation. The authors organize the discussion around three broad principles (the Data Integrity Principle, the Data Access and Sharing Principle and the Data Stewardship Principle) to then make a set of specific recommendations to assist in the development of policies, standards and infrastructure necessary to make the most of scientific data.

It is to the authors' credit that they took on such an ambitious task and succeeded in identifying the key issues at stake, drawing on examples from disciplines ranging from the social sciences to astronomy. But, perhaps precisely because they were charged with such a broad assignment, the principles and recommendations of the report end up being somewhat anticlimactic.

For example, the Data Access and Sharing Principle states that "Research data, methods and other information integral to publicly reported results should be publicly accessible." To that end, the report's four recommendations are that researchers make data accessible (or explain why they cannot do so, if there are "compelling reasons" to withhold them), that research communities develop discipline-specific sharing standards, that funders and journals promote data sharing and that research institutes establish clear sharing policies.

All of this is well and good, but it will hardly be news to those who have pondered these issues. At the *Nature* journals, for example, data sharing has long been a requirement for publication, and we have gone as far as directly urging authors to fulfill their commitment to sharing when other researchers have requested our involvement. So the merit of the report does not lie in its recommendations but in its disciplined

analysis of the current state of play, its multidisciplinary perspective on the problems and its identification of the tough questions that scientists, institutions, funders and journals need to answer to move forward, even though it provides little in terms of answers.

For example, a key aspect to ensure access to research data is ownership. The report duly provides a thorough discussion of copyright and patent issues, the legal aspects of data sharing and the responsibilities of journals. However, the authors of the report do not make specific suggestions about the type of actions that institutions or funders might take to encourage scientists to share their data. Thus, although the report makes clear that, if you receive federal funds for your research, you don't really own your data and must share them, the authors provides little guidance for cases in which your funds also come from an academic-industry partnership (which have an increasingly important role in science funding) or in which you collaborate with an institute with a different sharing policy from that of your own.

Also, the report quite understandably focuses largely on the US, shedding little light on integrity and accessibility issues that surround international collaborations. In Europe, for example, where international projects are the norm, the responsibility for the integrity and accessibility of the data rests upon the labs that produced them. However, papers often have a single corresponding author who assumes the responsibility for the whole manuscript. Should the corresponding author then be held accountable if a collaborator in a country with different rules does not want to share data?

In sum, the report is useful for bringing together many of the factors that need to be taken into account as the community finds the best way to ensure the integrity, accessibility and preservation of scientific data, but it does not provide an authoritative view on the way forward.

What, then, should the next steps be? Should an Asilomar-style conference take place to come up with formal recommendations that would then be globally adopted? Is that even feasible if we consider that, as technology advances, a set of rules set in 2010 may prove obsolete by 2012? Should we lean, as is often the case, on the federal government to take the lead and provide the financial and logistical resources to encourage and ensure data accessibility and preservation?

Perhaps we simply need a cultural shift to take up where recommendations leave off. In other words, scientists are the ones who have the most at stake in terms of ensuring that data are reliable and useful for future use. Scientists should therefore be the ones to take on the job and develop the right standards, lobby for the resources to set up the appropriate infrastructure and decide on the right measures to deter other scientists from data mismanagement. Data may not be the legal property of scientists, but looking after the data is certainly their responsibility.