

Supplementary data need to be kept in public repositories

SIR — The reality of the genomics age is that there are many very large data sets that are most usefully saved and manipulated in electronic form. Many journals add online ‘supplementary material’ to articles as a service to authors wishing to publish volumes of such data that cannot be accommodated within the body of an article.

Supplementary-material collections maintained by publishers serve as archival repositories directly connected with the peer-reviewed scientific literature, often competing with or substituting for the deposition of data in public repositories.

To assess the use of these, we investigated supplementary-data archives for gene-expression profiling data, a widely used experimental protocol for which international standards for data representation have been developed.

We anticipated that such archives might be a useful source of data. But to our dismay, it was impossible to systematically analyse our sample, taken from 10,128 papers in 139 journals. No standards for organizing supplementary-data collections have been adopted either across journals or even for supplementary-data collections associated with articles in the same journal.

Data are represented in an enormous range of different file formats, from raw data files (such as Affymetrix.cel files) to spreadsheets (xls file extensions), documents (doc and pdf) and text files (txt and cvs). Within documents there are no standards for data organization: different documents provide different numbers of columns, contain both differential and absolute expression values, and often have few details about the signal processing applied to obtain data. We also encountered a significant number of typographic errors in gene names, database accession numbers and data-set identifiers.

There are public repositories for gene-expression profile data (Stanford MicroArray Database, the US National Center for Biotechnology Information Gene Expression Omnibus and the ArrayExpress repository at the European Bioinformatics Institute). We compared the accessibility of gene-expression profile data in public repositories with accessibility of data in supplementary-data archives. The public repositories provide numerous search and retrieval tools, including unique accession numbers and the ability to search by specimen, platform and profile data. Publishers’ supplementary-materials archives provided none of these features. As a result, relevant data are far harder to locate than in public repositories.

These findings are not limited to

gene-expression data. Even within the same journal, there is no consistency in reporting or format among bioinformatics resources. File extensions for documents, figures and movies include xls, doc, eps, jpg, tif, gif, pdf, ppt, qt, asf, wma and wmv. They may or may not include long lists of links, be compressed into zip files or offer the option of including the supplementary material as part of the downloadable document containing the printed version of the article.

Supplementary data often represent the raw experimental values and are especially important for researchers in the same field. Among the advantages of storing these data in public repositories are the integration of information with the community knowledge resources and the ability to track and maintain computer-readable associations between data sets.

On the basis of our analysis, we recommend that scientific journals adopt a policy, similar to *Nature*’s (see www.nature.com/nature/authors/policy/index.html#a7.2), of requiring that authors submit data to public repositories, if relevant repositories exist, and that the journal version should contain accession numbers, URLs and other appropriate specific indicators to the data source in the repositories.

Journals’ supplementary-data archives should be restricted to idiosyncratic and nonstandard data types for which no public repository exists. Only then can community standards emerge.

Carlos Santos*, **Judith Blake†**, **David J. States*‡**
*Bioinformatics Program, University of Michigan, Ann Arbor, Michigan, USA

†The Jackson Laboratory, Bar Harbor, Maine, USA

‡Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

Turkish science needs more than membership of the EU

SIR — Your Editorial “Turkey’s evolution” (*Nature* **438**, 1–2; 2005), about the country’s efforts to join the European Union (EU), states that “the opening of negotiations for EU membership offers the best hope for the continuing development of science in Turkey”. This view is common in Europe, but I believe the assumptions behind it lack solid support.

First, you assume that EU policies adopted by Turkey during membership negotiations will lead to more economic investment in Turkish science. Such investment is needed if Turkey is to close the gap with more developed countries. But the increase in the science budget, to US\$300 million in a country of 70 million, is inadequate. The €250 million (US\$292 million) that Turkey contributed towards the EU’s Sixth Framework programme is not expected to be recouped. And even though policies prescribed by

the International Monetary Fund (IMF) have reduced investment in the country’s educational infrastructure (E. Voydova and E. Yeldan *Comp. Econ. Stud.* **47**, 41–79; 2005), keeping to an IMF programme is a condition for Turkey’s acceptance into the EU.

Second, although international scientific collaboration is crucial for scientific development in any country, the extent to which knowledge sharing and cooperation depends upon international economic and political relations is less clear. Some countries, such as Cuba, India and China, have achieved scientific progress in relatively independent economic or political circumstances. Political and cultural relations among countries at dissimilar levels of development might even impede progress on the weaker side — for example, through a ‘brain drain’ effect.

Last, I fear that entrusting all hope of development to the ambiguous political process of EU membership may undermine Turkey’s existing — albeit weak — resolution to advance science.

The country needs a firm political resolution to implement long-term public investments in education and science, regardless of EU membership negotiations.

Mehmet Somel

Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany

Flu virus will not be sent in the regular US mail

SIR — The headline and photographs of your News story “Deadly flu virus can be sent through the mail” (*Nature* **438**, 134–135; 2005) are misleading with respect to the policy of the Centers for Disease Control and Prevention (CDC) regarding the transfer and use of the 1918 pandemic influenza virus. They could give the erroneous impression that the virus will be made widely available and sent through the regular US mail.

The CDC has not yet received any requests to work with the 1918 virus at a non-CDC facility and I have made it clear we currently have no plans to send the virus anywhere. Any requests we do receive will be considered on a case-by-case basis, taking into account scientific merit, biosafety and biosecurity concerns, as well as any additional standards deemed appropriate for this particular virus.

The CDC is the only agency that currently possesses this virus, and we have a special responsibility to balance the importance of scientific progress and collaboration with the moral and scientific imperatives of biosafety and biosecurity.

Julie Louise Gerberding

Centers for Disease Control and Prevention, 1600 Clifton Road Northeast, Atlanta, Georgia 30333, USA