

## Time for fair trade in research data



Geneticists, astrophysicists, and molecular biologists routinely share research data with colleagues and rivals alike. The reason is that scientists and their funders know we will understand complex issues sooner if people build on one another's work.<sup>1,2</sup> Yet scientists in the complex area of public health have been left behind in the data-sharing revolution.

If health researchers made their data available to colleagues, there would be less duplication of research and fuller use of study results. Data could be combined across time and countries to answer new questions, improving health policy. Data sharing would save time, effort, and money—it would probably also save lives.

Why do researchers not share data more? The obstacles are ethical, technical, and professional. Science funders, wanting more public health bang for each research buck, believe the obstacles can be overcome with the right investments and incentives. Researchers and journal publishers will play crucial roles. Many funding bodies are now reviewing their data-sharing policies. Researchers should engage to ensure that emerging policies meet their needs. A draft code on sharing public health data, the result of consultations between funders and international researchers, was discussed at the International Ministerial Meeting on Health Research in Bamako, Mali, in November, 2008.<sup>3</sup>

The ethical hurdles to sharing data are thrown up by concerns that secondary users might not respect the promises of confidentiality made to participants. But anonymisation and encryption technologies have come a long way: with sensible data access policies, data can be shared with minimum risk to individuals. Broad consent policies are already becoming common, while failure to maximise the use of data to improve people's health is under increased ethical scrutiny.<sup>4</sup>

The social sciences have shown that data sets containing personal information can be shared with minimum fuss. Biomedical data might require extended metadata standards and additional anonymisation to safeguard sensitive health information, but most of the hard work has been done by pioneers in other fields. The major technical hurdle for epidemiologists is to raise standards in the woefully neglected area of data management, which is no small task. In public health

research, data-management capacity is limited; in developing countries, it is virtually non-existent.<sup>5</sup>

Data management is rarely treated as a discipline in its own right, so such management remains undervalued and underfunded, shoring up the professional hurdles to data sharing. Epidemiologists gain no credit for publishing datasets and data managers are rarely authors on publications. As long as funding and promotion depend on publishing papers in peer-reviewed journals, giving away data equates to giving away job prospects. The emphasis on publication discourages researchers from allowing others to analyse data they have collected, and stacks incentives against wringing the greatest knowledge from data in the shortest time.

Explicit policies from funders, journals, and universities laying out requirements and rewards for data sharing might coax more epidemiologists into the data-sharing age. Funders of public health research are increasingly requiring grantees to say how they expect to share research data. The US National Institutes of Health now require that data from their larger grants be made available to other researchers.<sup>6</sup> Some biomedical journals require a statement about data availability; in other fields, journals encourage researchers to submit a replication dataset with articles.<sup>7,8</sup>

No one is talking of the instantaneous release of machine-readable data. Protected fair-use periods for primary investigators and bona-fide access restrictions will probably become norms. Still, epidemiologists remain concerned about "giving away" data. Most worried of all are field researchers in developing

Published Online  
November 13, 2009  
DOI:10.1016/S0140-  
6736(09)61486-0

The printed journal  
includes an image merely  
for illustration

countries, who do much of the hard graft in collecting data of interest to the global public health community. Senior scientists guiding small overworked teams in places with erratic electricity supplies and limited computing power do not have the time or the pool of skills available to do all the analysis they would like to. It will not help if they have to use their limited resources to manage data for analysis by academics from well-funded institutions in the developed world.

Sharing data can lead to new collaborations and increased funding, but examples are few and researchers remain wary.<sup>9</sup> With public health data, we need fair trade, not free trade. If funders wish developing world scientists to make their data available to others for secondary analysis, they must invest to give those scientists the skills to do primary analysis more rapidly. Secondary users and their funders will have to contribute, collaborating with primary researchers, learning about the dataset and passing on analysis skills. A history of publishing data must be recognised when reviewing grant applications. Metadata and archiving standards must be developed, data managers trained and supported, and data storage infrastructure expanded.

These developments will cost money, but many funders of research are prepared to make the investment. Genomics has taught us that investing in data sharing cuts duplication, speeds progress, and increases career opportunities for researchers. In public

health, the dividend will also be better policies and healthier people.

\*Elizabeth Pisani, James Whitworth, Basia Zaba, Carla Abou-Zahr

London N16 7LE, UK (EP); Wellcome Trust, London, UK (JW); Centre for Population Studies, London School of Hygiene and Tropical Medicine, London, UK (BZ); and Department of Health Statistics and Informatics, WHO, Geneva, Switzerland (CA-Z) [pisani@ternyata.org](mailto:pisani@ternyata.org)

We thank Ties Boerma, Tamara Fox, Katherine Littler, Nicola Perrin, and Sara Seims for comments. EP has worked as a consultant for the Wellcome Trust on data-sharing projects. JW is employed by the Wellcome Trust, a foundation that funds health research, including in developing countries. BZ holds a grant funded by the Wellcome Trust, under which data collected in field sites in developing countries will be shared. CA-Z declares that she has no conflicts of interest. The views in this Comment do not necessarily reflect those of WHO.

- 1 Chokshi D, Parker M, Kwiatkowski D. Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. *Bull World Health Organ* 2006; **84**: 382–87.
- 2 Arzberger P, Schroeder P, Beaulieu A, et al. Promoting access to public research data for scientific, economic, and social development. *Data Science Journal* 2004; **3**: 135–52.
- 3 Wellcome Trust, WHO. Sharing public health data: a code of conduct. 2008. <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/index.htm> (accessed Aug 3, 2009).
- 4 Mascalzoni D, Hicks A, Pramstaller P, Wjst M. Informed consent in the genomics era. *PLoS Med* 2008; **5**: e192.
- 5 Chandramohan D, Shibuya K, Setel P, et al. Should data from demographic surveillance systems be made more widely available to researchers? *PLoS Med* 2008 **5**: e57.
- 6 SHERPA Project. Research funders' open access policies. <http://www.sherpa.ac.uk/juliet/index.php> (accessed Aug 3, 2009).
- 7 Groves T. Managing UK research data for future use. *BMJ* 2009; **338**: b1252.
- 8 Annals of Applied Statistics. Manuscript submission. <http://www.imstat.org/aoas/mansub.html> (accessed Aug 3, 2009).
- 9 Piwowar H, Day R, Fridsma D. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2007; **2**: e308.