

9. Mascalzoni D, Hicks A, Pramstaller P, Wjst M. Informed consent in the genomics era. *PLoS Med* 2008;5:e192. doi:10.1371/journal.pmed.0050192 PMID:18798689
10. Homer N, Szelling S, Redman M, Duggan D, Tembe W, Muehling J et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008;4:e1000167. doi: 10.1371/journal.pgen.1000167 PMID:18769715
11. *Application to use restricted microdata*. Minneapolis: IPUMS International. Available from: <https://international.ipums.org/international/> [accessed 26 February 2010].
12. UK Data Archive. End user licence. Colchester: University of Essex; 2008. Available from: <http://www.data-archive.ac.uk/aandp/access/licence.asp> [accessed 26 February 2010].
13. Pisani E, Whitworth J, Zaba B, AbouZahr C. Time for fair trade in research data. *Lancet* 2010;375:703–5. doi:10.1016/S0140-6736(09)61486-0 PMID:19913902
14. *OECD Principles and guidelines for access to research data from public funding*. Paris: Organisation for Economic Co-operation and Development; 2007.
15. *Global strategy and plan of action on public health, innovation and intellectual property*. Geneva: World Health Organization; 2008.
16. How to encourage the right behaviour. *Nature* 2002;416:1. doi:10.1038/416001b
17. Data's shameful neglect. *Nature* 2009;461:145. doi:10.1038/461145a
18. PLoS Medicine Editors. Next stop, don't block the doors: opening up access to clinical trials results. *PLoS Med* 2008;5:e160. doi:10.1371/journal.pmed.0050160 PMID:18630986
19. Laine C, Berkwits M, Mulrow C, Schaeffer MB, Griswold M, Goodman S. Reproducible research: biomedical researchers' willingness to share information to enable others to reproduce their results. In: *Sixth International Congress on Peer Review and Biomedical Publication, Vancouver, Canada, 10–12 September 2009*. Available from: <http://www.ama-assn.org/public/peer/abstracts-0910.pdf> [accessed 26 February 2010].
20. *NIH guide: final NIH statement on sharing research data*. Bethesda: National Institutes of Health; 2003. Available from: <http://grants.nih.gov/grants/oe.htm> [accessed 26 February 2010].
21. *MRC Policy on data sharing and preservation*. London: Medical Research Council; 2008. Available from: <http://www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing/PolicyonDataSharingandPreservation/index.htm> [accessed 26 February 2010].
22. *Policy on data management and sharing*. Wellcome Trust; 2007. Available from: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm> [accessed 26 February 2010].
23. *Sharing public health data: a code of conduct*. London: Wellcome Trust; 2008. Available from: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/index.htm> [accessed 26 February 2010].
24. Supari SF. *Saathya dunia berubah: tangan Tuhan di balik virus flu burung / Siti Fadilah Supari* [in Indonesian]. Jakarta: Sulaksana Watinsa Indonesia; 2008.
25. *International Comprehensive Ocean-Atmosphere Data Set*. Washington, DC: National Oceanic and Atmospheric Administration; 2009. Available from: <http://icoads.noaa.gov/> [accessed 26 February 2010].
26. *International Clinical Trials Registry Platform, WHO Registry Criteria, version 2.1*. Geneva: World Health Organization; 2009. Available from: http://www.who.int/ictrp/network/criteria_summary/en/index.html [accessed 26 February 2010].
27. DataCite - International initiative to facilitate access to research data. Hannover: German National Library of Science and Technology; 2009. Available from: <http://www.datacite.org/> [accessed 26 February 2010].
28. Paskin N. Digital Object Identifier (DOI) System. In: *Encyclopedia of library and information sciences*. New York: Taylor & Francis; 2008.
29. Altman M, King G. A proposed standard for the scholarly citation of quantitative data. *D-Lib* 2007. 13.
30. Lowrance W. *Access to collections of data and materials for health research*. London: Medical Research Council; 2006.
31. Pisani E. *OpenEpi: a new culture for public health data?* London: Wellcome Trust; 2008.
32. National Academy of Sciences. *Ensuring the integrity, accessibility and stewardship of research data in the digital age*. Washington, DC: National Academy Press; 2009.

Round table discussion

Publishing research data on a professional basis

Toby Green^a

As Pisani & AbouZahr have identified, there are many obstacles to the publishing of data: social (incentives for researchers to make the effort to publish), financial (having adequate financing to cover short-term publishing and long-term curation costs), and technical (standards and systems).¹ This paper looks at some of the technical challenges of publishing data professionally and describes the discoverability and citability benefits that follow.

Let's take it as read that publishing research data is a "good thing," that researchers are as willing to publish data as they are research papers and funding is in place to make them available online in the long run. Job done? Well, no, not by a long chalk.

Just as loading a journal article onto a web site somewhere isn't the same as publishing it properly, so the same is true for data. To be as discoverable and as citable as research articles, data sets need to be published using an infrastructure that is compatible with research articles. It is not enough that data sets hang like dongles off a research article; they need to be discoverable and citable in their own right – just like a journal article. This means the metadata must be compatible with existing bibliographic management and citation systems like Ref Works[®] and CrossRef[®]. Users will expect search engines, abstracting and indexing services and library catalogues to reference data sets, so, for example, librarians will need MARC (MACHine-Readable Cataloging) records.

Is this overkill? Well, the Organisation for Economic Co-operation and Development (OECD) doesn't think so. OECD publishes more than 390 data sets as stand-alone objects, as well as thousands of data sets as supplemental data to its books and journal articles. Sub-sets of the data sets are also posted on the web as stand-alone objects too. So it is no surprise that, in the absence of good discovery metadata and systems, the number one complaint from users is the challenge of finding a relevant data set. They know the data is there, but they can't find it – even with Google's help.

To solve this problem, OECD's Publishing Division has spent the past three years grappling with the challenge of how to publish these many thousands of data objects so that users can not only find the data they need, but can then cite and manage the data sets using the same tools that they already use to manage journal articles or book chapters. The first result was a white paper,² first released in March 2009, which described this challenge and proposed a set of metadata schema for databases in their own right, sub-sets of databases and supplemental data.

More significantly, was the launch of OECD iLibrary, OECD's new publishing platform, in July 2009. OECD iLibrary³ hosts all OECD books, working papers, journals and data sets in a seamless manner. OECD iLibrary puts the white paper's proposed bibliographic schema for data objects into practice. Search for "health data" and the search results include data sets, book chapters – even individual tables found inside books.

^a Organisation for Economic Co-operation and Development, 2 rue André Pascal, 75775 Paris Cedex 16, France. Correspondence to Toby Green (e-mail: toby.green@oecd.org).

OECD's data sets can now be discovered more easily and they can be cited as simply and as easily as a research article using the downloadable citation provided. Later in 2010, librarians will be supplied with MARC records and the bibliographic records for OECD data sets will be shared with discovery platforms like RePEc (Research Papers in Economics)⁴ – the world's largest collection of economics grey literature – enabling visitors to find data objects alongside working papers and journal articles. Imagine being able to discover and cite data sets as easily and as simply as journal articles. Imagine no more. ■

Competing interests: None declared.

References

1. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;88:462–6.
2. Green T. *We need publishing standards for datasets and data tables*. Paris: Organisation for Economic Co-operation and Development; 2010. Available from: <http://dx.doi.org/10.1787/603233448430> [accessed 1 March 2010].
3. OECD iLibrary [Internet site]. Paris: Organisation for Economic Co-operation and Development; 2009. Available from: www.oecd-ilibrary.org [accessed 19 May 2010].
4. RePEc (Research Papers in Economics) [Internet site]. Available from: www.repec.org [accessed 1 March 2010].

Sharing data for public health: where is the vision?

Alan D Lopez^a

“By refusing to share data, researchers are slowing progress towards reducing illness and death.” Pisani & AbouZahr are making a big claim in this round table.¹ Is this claim sensationalist or does it have some basis? Can we argue that data from public health research really affect the ways prevention and control programmes are designed? Lives have become longer and healthier in the past 50 years, despite an arguably poor evidence base for health and an even poorer appreciation by policy-makers of the value of reliable health information.^{2,3} Pisani & AbouZahr are arguing that such gains would have been bigger, faster and more equitable had the world had better information about what works and does not work in public health; lost ground is partly due to widespread hoarding of research findings, particularly primary data.

They have a point. Restricting access to data to only those scientists directly engaged in a research project limits the scope of legitimate scientific enquiry and the potential for research to influence policy and practice. No individual scientist who collects or collates data has all the possible analytic methods, expertise and time to extract key public health messages from research or routine data sets.^{4–7} Lost opportunity for analysis is the main consequence of poor data sharing practices.

Yet, as Pisani & AbouZahr argue, it is unreasonable to expect data collectors to share without adequate incentives. Incentives could include professional recognition for well collected and documented data, appropriately disseminated using good data management practices. Data collectors too need assurance that their efforts will be respected and that errors in data are inevitable and rarely disastrous. Experienced researchers are aware of

these risks and can use a range of quality assessment techniques to deal with errors.

Mentoring is one incentive that is missing from the otherwise excellent set proposed by Pisani & AbouZahr. Partnerships between researchers and data collectors, including intensive methodological workshops, are feasible and can help ensure that those who collect data realize the public health potential and value of their efforts. Such an approach could rapidly increase analytical capacity and diversify the analysis of rich, but underutilized, data sets. Funding such collaborations would be an innovative and constructive use of research funds. Competent analysts should be able to resolve potential challenges in interpreting data because of specific local conditions surrounding their collection. Restricting access on this basis reflects a lack of confidence, imagination or trust by those who collect data and should be questioned when used to preclude further analysis.

The authors propose an urgent agenda for action to improve data sharing practices that will benefit all stakeholders – data collectors, analysts, the policy community and, ultimately, the public. This is admirable but, for such a plan to succeed, funders, researchers and data collectors alike need to understand its benefits. That will only happen with effective and committed leadership. What better role for the World Health Organization? ■

Competing interests: None declared.

References

1. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;88:462–6.
2. Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S et al.; Monitoring of Vital Events. A scandal of invisibility: making everyone count by counting everyone. *Lancet* 2007;370:1569–77. doi:10.1016/S0140-6736(07)61307-5 PMID:17992727
3. Horton R. Counting for health. *Lancet* 2007;370:1526. doi:10.1016/S0140-6736(07)61418-4 PMID:17992726
4. Murray CJL, Lopez AD. The utility of DALYs for public health policy and research: a reply. *Bull World Health Organ* 1997;75:377–81. PMID:9342897
5. Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring under-five mortality: validation of new low-cost methods. *PLoS Med* 2010;7:e1000253. doi:10.1371/journal.pmed.1000253 PMID:20405055
6. Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD et al. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974–2006. *PLoS Med* 2010;7:e1000260. doi:10.1371/journal.pmed.1000260 PMID:20405004
7. Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Med* 2010;7:e1000262. doi:10.1371/journal.pmed.1000262 PMID:20405002

Data sharing: reaching consensus

Jimmy Whitworth^b

Pisani & AbouZahr write passionately about the need to change the culture of data sharing in public health research.¹ They explain why this is in everybody's best interests and outline ways in which the main obstacles can be overcome. This is laudable and much appreciated; it is time for a change, the current situation is unacceptably inefficient in terms of scientific progress and value for money from research.

^a School of Population Health, University of Queensland, Herston Road, Herston, Qld., 4006, Australia (e-mail: a.lopez@sph.uq.edu.au).

^b Wellcome Trust, Gibbs Building, 215 Euston Road, London, NW1 2BE, England (e-mail: j.whitworth@wellcome.ac.uk).