

The potential of internet-based technologies for sharing data of public health importance

Greg Fegan,^a Michael Mouldsdales^a & Jim Todd^b

As professionals in data processing, analysis and information technology, we read with interest the *Bulletin's* coverage of the barriers to data sharing in public health.¹ We contend that there are many existing solutions for low-cost, high-quality data collection, management and analysis. Many of these systems are built on open-source technologies and thus are more amenable to receiving input for their design and operation from information technologists and researchers in low-income countries. The information technology gap between rich and poor countries may not be as large as some may think. For example, a recent map of Facebook "friend" linkages shows areas of high internet connectivity in low-income countries, with specific interest to us being the connection into Rwanda from Mombasa, Kenya.²

In our institutions, after some consideration,³ we adopted a web-based, open source clinical trials package called OpenClinica (Akaza Research, Waltham, MA, United States of America) for a large ($n > 3000$) multi-site trial (more information available at: <http://www.feast-trial.org>). For observational epidemiological and clinical studies that rely on, or are derived from, surveillance systems there are several free, easy-to-install systems such as RedCap,⁴ OpenMRS⁵ and OpenXdata. Indeed, there are novel technologies that have come from low-income countries that are now in use in high-income countries, e.g. software developer Ushahidi.⁶ As Tom Smith of the Swiss Tropical Institute said at the Pan-African Malaria Conference in 2009 when introducing a presentation on a mobile phone-based system for malaria surveillance in Zanzibar, United Republic of Tanzania: "Africa is in the vanguard in the use of such technology." Indeed the Kenyan mobile phone-based money transfer system, M-Pesa, is highly regarded and has spread very quickly.⁷ Given the ubiquity of mobile phones, the use of developing technologies, such as lens-free microscopy,⁸ are likely to have major impacts soon on disease surveillance in low-income countries.

With such technologies comes the need for a tool to effectively analyse and

disseminate the data generated. One such tool is the open-source software package, simply named R, arguably⁹ the most rapidly evolving and powerful data analytical engine with which the major statistics software packages can integrate. A most useful input from the World Health Organization and the Special Programme for Research and Training in Tropical Diseases has been sponsorship of the online R course at Thailand's Prince of Songkla University and publication of a free book. R's potential for data storage, use and analysis is well illustrated by Zack Almquist¹⁰ who has created a freely available set of tools that interrogates data from the year 2000 census in the United States of America.

Pisani & AbouZahr discuss how collaboration allows researchers to stand on the shoulders of giants.¹ We think this R package does just that. The field of genetics, which they cite as one that public health and epidemiological researchers should emulate, has been greatly assisted by R through the BioConductor project. Robert Gentleman, who developed R with Ross Ihaka,¹¹ is a notable contributor to this project.¹²

With powerful data programmes freely available, data can be shared more widely, but this also depends on skilled personnel. In east Africa alone, we need to train 100 new data managers and statisticians every year so that researchers, data managers and analysts can benefit from the wider availability of data and to ensure its quality. Public health researchers need to be trained in good data collection methods, mentoring with researchers and partnering with data analysis experts from developed countries. Systems are required that extend beyond research projects into the collection of routine data for public health systems that can be used to monitor the health of the whole population.

We think technology and shared data have the potential to radically transform the health systems of low-income countries within our lifetime. We believe that technology and data should be shared equitably

across all countries and that everyone should be enabled to use the results from the acquired knowledge. ■

References

1. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;88:462–6. doi:10.2471/BLT.09.074393 PMID:20539861
2. Smith D. Facebook's social network graph. *R-Bloggers*, 14 December 2010. Available from: <http://www.r-bloggers.com/facebook-social-network-graph/> [accessed 10 January 2011].
3. Fegan GW, Lang TA. Could an open-source clinical trial data-management system be what we have all been looking for. *PLoS Med* 2008;5:e6. doi:10.1371/journal.pmed.0050006 PMID:18318594
4. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81. doi:10.1016/j.jbi.2008.08.010 PMID:18929686
5. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D et al. The OpenMRS Implementers Network. *Int J Med Inform* 2009;78:711–20. doi:10.1016/j.ijmedinf.2008.09.005 PMID:19157968
6. Ushahidi [corporate website]. Available from: <http://ushahidi.com/> [accessed 10 January 2011].
7. Mas I, Radcliffe D. *Mobile payments go viral: M-PESA in Kenya*. Washington DC: The World Bank; 2010. Available from: http://www.microfinancegateway.org/gm/document-1.9.43376/Mobile%20Payments%20Go%20Viral_M-PESA%20in%20Kenya.pdf [accessed 10 January 2011].
8. Tseng D, Mudanyali O, Oztoprak C, Isikman SO, Sencan I, Yaglidere O et al. Lensfree microscopy on a cellphone. *Lab Chip* 2010;10:1787–92. doi:10.1039/c003477k PMID:20445943
9. Wilkinson L. The future of statistical computing. *Technometrics* 2008;50:418–35. doi:10.1198/00401700800000460
10. Almquist ZW. US Census spatial and demographic data in R: the UScensus2000 suite of packages. *J Stat Softw* 2010;37:1–31.
11. Vance A. Data analysts captivated by R's power. *New York Times*, 6 January 2009.
12. Gentleman R, Temple Lang D. Statistical analyses and reproducible research. *J Comput Graph Statist* 2007;16:1–23. doi:10.1198/106186007X178663

^a Kenya Medical Research Institute, Centre for Geographic Medicine Research (Coast), PO Box 230, Kilifi, 80108, Kenya.

^b National Institute of Medical Research, Tazama Project, Mwanza, United Republic of Tanzania.

Correspondence to Greg Fegan (e-mail: gfeagan@kilifi.kemri-wellcome.org).