

## Sharing research data to improve public health



The purpose of medical research is to analyse and understand health and disease. A key and expensive element is the study of populations to explore how interactions between behaviour and environment, in the context of genetic diversity, determine causation and variation in health and disease. As funders of public health research, we need to ensure that research outputs are used to maximise knowledge and potential health benefits. In turn, the populations who participate in research, and the taxpayers who foot the bill, have the right to expect that every last ounce of knowledge will be wrung from the research.

Ensuring data are made widely available to the research community accelerates the pace of discovery and enhances the efficiency of the research enterprise. In many research fields—from genetics and molecular biology to the social sciences—data sharing is already

ingrained in how researchers work. In genetics and genomics, the pooling of studies of different populations has led to an explosion of knowledge on the genetic determinants of human variation in health and disease.<sup>1</sup> Well-established repositories and tools enable researchers to access and interrogate shared data resources, and build on one another's work.<sup>2</sup>

By contrast, this culture has yet to be widely embraced by the public health research community. Much of the infrastructures, technical standards, and incentives that are needed to support data sharing are lacking, and these data can hold particular sensitivities. And some researchers are reluctant to share data. Too often, data are treated as the private property of investigators who aim to maximise their publication record at the expense of the widest possible use of the data. This situation threatens to limit both the

Published Online  
January 10, 2011  
DOI:10.1016/S0140-  
6736(10)62234-9

### Panel: Joint statement of purpose—vision, principles, and goals

#### Vision

We intend to work together to increase the availability to the scientific community of the research data we fund that is collected from populations for the purpose of health research, and to promote the efficient use of those data to accelerate improvements in public health

#### Principles

Funders agree to promote greater access to and use of data in ways that are:

- *Equitable*: it should recognise and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data, and communities and funders who expect health benefits to arise from research
- *Ethical*: it should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data
- *Efficient*: it should improve the quality and value of research, and increase its contribution to improving public health; approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition

#### Immediate goals

- Standards of data management are developed, promoted, and entrenched so that research data can be shared routinely and reused effectively

(Continues in next column)

(continued from previous column)

- Funders and employers of researchers recognise data management and sharing of well-managed datasets as an important professional indicator of success in research
- Researchers creating datasets for secondary analysis from shared primary data are expected to share those datasets and act with integrity and in line with good practice, giving due acknowledgment to the generators of the original data

#### Longer-term aspirations

- Data collected for health research are made available to the scientific community for analysis which adds value to existing knowledge and which leads to improvements in health
- The research community, particularly those collecting data in developing countries, develop the capacity to manage and analyse those data locally, as well as contributing to international analysis efforts
- To the extent possible, datasets underpinning research papers in peer-reviewed journals are archived and made available to other researchers in a clear and transparent manner
- The human and technical resources and infrastructures needed to support data management, archiving, and access are developed and supported for long-term sustainability

The full statement is online: <http://www.wellcome.ac.uk/publichealthdata>. Other funding organisations are invited to join as signatories and partners in this work.

progress of this research and its application for public health benefit.

In May, 2010, the Wellcome Trust and the Hewlett Foundation convened a workshop in Washington, DC, to explore how funders could address these issues. The meeting brought together representatives of global health funding agencies and other communities, including academic researchers, international organisations, and journals. It built on earlier discussions at the Global Ministerial Forum in Bamako in 2008 and broad consultation with stakeholders in developing and developed countries over the past 2 years.<sup>3</sup>

The funders represented in Washington committed to work together to increase the availability of data generated by their funded research, subject to appropriate safeguards. Consensus crystallised around a series of high-level principles and goals for advancing this vision (panel). So far, 17 funding organisations have signed up to the statement, committing to further these goals within the context of their legal and operating frameworks.<sup>4</sup>

Any discussion on increasing access to research data on public health typically raises three key concerns. The first is that researchers in resource-poor settings doing much of the crucial work to generate public health research datasets will lose out to better-resourced researchers overseas, who have the skills and tools to rapidly analyse data. We are committed to advancing data sharing in a way that balances the rights and responsibilities of those who generate and those who use data, and which recognises the contributions and expectations of the individuals and communities who have participated in the research—fair trade, not free trade.

Second, there are fears that increased data sharing will create unacceptable risks for research participants. But data should only be shared if adequate safeguards are in place, and in a manner fully consistent with the terms of the consent under which the data were provided. Although safeguarding privacy is paramount, confidentiality can be maintained by meticulous handling of research records and anonymisation or pseudonymisation. Meanwhile, consent can and must be obtained in a fashion that enables participants to understand that the value of their participation will be maximised. Indeed, it is unethical for an ethics committee to allow a study to go ahead that does not maximise this potential value, while also protecting confidentiality.

Third, data sharing carries a substantial cost in terms of money and time. No research funder wishes to support data sharing for its own sake: any shared data must have a value to other users that will justify the resources to make them usefully available. Many funders require researchers to set out their approach for sharing data and the resources they will need. These plans are considered as an integral part of the funding decision.

Although identifying high-level principles is an important first step, the challenges in building the culture and resources needed to support data sharing are considerable. We must build the capacity and skills in the research community to manage and analyse data, particularly in low-income and middle-income countries. We need to create incentives for researchers to share data and shift a culture in which rewards are almost exclusively based on publications to one in which those who collect and curate the data are valued equally. And we need to develop the data infrastructure and technical tools needed to store, preserve, and analyse research datasets safely and securely.

The partners in this initiative have established working groups that will develop joint activities to address these challenges, working with others as required. We want to involve as broad a base of funders in these discussions as possible, and welcome participation from other organisations who are committed to maximising the full potential of public health research data to generate better health. This is an urgent problem and we call on researchers and funders to mend their ways.

*\*Mark Walport, Paul Brest*

Wellcome Trust, London NW1 2BE, UK (MW); and Hewlett Foundation, Menlo Park, CA, USA (PB)  
d.carr@wellcome.ac.uk

Other signatories to the joint statement of purpose include: Warwick Anderson, National Health and Medical Research Council, Canberra, Australia; Alain Beaudet, Canadian Institutes of Health Research, Ottawa, Canada; Paul Boyle, Economic and Social Research Council, Swindon, UK; Carolyn M Clancy, Agency for Healthcare Research and Quality, Rockville, USA; Francis S Collins, National Institutes of Health, Bethesda, USA; Thomas R Frieden, Centers for Disease Control and Prevention, Atlanta, USA; Ed Henry, Doris Duke Charitable Foundation, New York, USA; Pamela S Hyde, Substance Abuse and Mental Health Services Administration, Rockville, USA; Matthias Kleiner, Deutsche Forschungsgemeinschaft, Bonn, Germany; Tamar Manuelyan Atinc, World Bank, Washington DC, USA; Robin Olds, Health Research Council of New Zealand, Auckland, New Zealand; John Savill, Medical Research Council, London, UK; André Syrota, INSERM, Paris, France; Mary K Wakefield, Health Resources and Services Administration, Rockville, USA; Tachi Yamada, Bill & Melinda Gates Foundation, Seattle, USA. We thank Elizabeth Pisani, Katherine Little, Sara Seims, Helena Choi, David Carr, Nicola Perrin, Robert Terry, and Jimmy Whitworth for their contributions to this work. MW is the Director of the Wellcome Trust; PB is the President of the Hewlett Foundation.

- 
- 1 McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–69.
  - 2 Cochrane GR, Galperin MY. The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Res* 2010; **38** (suppl 1): D1–4.
  - 3 Pisani E, Whitworth J, Zaba B, AbouZahr C. Time for fair trade in research data. *Lancet* 2010; **375**: 703–05
  - 4 Wellcome Trust. Sharing research data to improve public health: joint statement of purpose. Jan 10, 2011. <http://www.wellcome.ac.uk/publichealthdata>.