

SAVING AND SHARING RESEARCH DATA: ISSUES OF POLICY AND PRACTICE

Peter Davis
Professor of Public Health
University of Otago
Christchurch

Late in November 2003 Wellington hosted a one-day conference on the policy and practice of saving and sharing research data. Sponsored by the Health Research Council,¹ the conference brought together speakers – both local and international – drawn from research funding agencies, government, and the scientific community. Over 70 attended, mainly from government agencies and policy ministries, but also with a smattering of researchers.

The meeting was prompted by two trends. Firstly, there is growing public investment in scientific research. Secondly, information and communication technologies are transforming the scientific enterprise and the management of research data, including its storage, access, analysis and distribution. Accordingly, the purpose of the meeting was threefold:

- to hear from key actors about the potential for saving and sharing research data
- to consider issues that follow from the OECD-backed principle that “publicly funded research data should be openly available to the maximum extent possible”
- to advance the issue in the science and science policy community.

The conference was opened by the Honourable John Tamihere, Minister of Statistics. The morning session (chaired by Steve Thompson, Chief Executive of the Royal Society of New Zealand) provided an international perspective from three keynote speakers, followed by reactions from two New Zealand scientists, with stakeholders and panellists in the afternoon session (chaired by Ray Delany, Chief Executive of the New Zealand Health Information Service).²

¹ The Health Research Council provided funding, administrative and organisational support (in the person of Fiona Gordon), and top-level assistance and advice (from Chief Executive, Bruce Scoggins). SPEaR, New Zealand Health Information Service, Ministry of Research, Science and Technology, and the Foundation of Research, Science and Technology provided funding for visiting speakers.

² PowerPoint versions of all these presentations are available on the SPEaR website.

INTERNATIONAL SPEAKERS

The international dimension was provided by Dr Peter Dukes of the United Kingdom's Medical Research Council, and Dr Deborah Mitchell and Sophie Holloway from the Australian National University, Canberra.

Peter Dukes discussed why the Medical Research Council has moved to develop an explicit policy of active data preservation. In part, this was a matter of prudent asset management. After decades of research funding and building up an impressive inventory of data sets, the Medical Research Council began to appreciate that it had a significant stewardship role for this substantial investment in the British scientific infrastructure. With the retirement of a generation of scientists came the realisation that resources of incalculable intellectual value were in danger of being lost for want of an adequate asset management policy (and reliance on the unguided, enlightened self-interest of scientists).

The speakers from the Australian National University arrived at a similar conclusion from a different starting point. The Australian National University already has a data archive for the social sciences, and Sophie Holloway is the university's first digital data archivist/professional librarian. Technological developments have made a vast range of data amenable to digital preservation and rendered the traditional central data repository somewhat anomalous. It is now possible to preserve almost any cultural artefact in digital form, including qualitative and quantitative research data. At the same time, connectivity and data grids of modern information technology mean that linkages can be made across the internet between data sets, investigators and computer analytical power – in real time. All that is required are protocols for preservation and access, together with the necessary expertise and software. This opens up the issues of preservation and access of research data in a hitherto unimagined way. Deborah Mitchell, who is Director of the ACSPRI Centre for Social Research, welcomed New Zealand participation in addressing these issues.

NEW ZEALAND RESPONSES

The contributions from the United Kingdom and Australia represented an international science momentum in which, until now, New Zealand has taken little part. The Economic and Social Research Council in the United Kingdom, the National Science Foundation in the United States, and the Australian Research Council all have policies encouraging data preservation and access among their grantees. No such development has occurred in New Zealand. How do New Zealand researchers and stakeholders respond to this trend?

Associate Professor Ritchie Poulton, Director of the Dunedin Multidisciplinary Health and Development Study, presented the perspective of an investigator who is the custodian for one of the most valuable research data sets in New Zealand. Dr Poulton maintained that his group adhered in large measure to the principles established by the OECD report on the saving and sharing of research data, asserting that the group has collaborated with a number of investigators outside the immediate team. However, this data access has been carefully contained and vetted. A policy of this kind, argued Dr Poulton in his concluding remarks, was necessary in order to maintain respondent and stakeholder confidence, to secure long-term investigator commitment, and to enhance the quality of data and research output.

Associate Professor Chris Cunningham presented a Māori viewpoint. Māori have a particularly protective attitude to data access, particularly where data tap Māori cultural and linguistic sources. Any New Zealand policy of enhanced data preservation and access would have to take these views into account. Despite this, Dr Cunningham remained optimistic that such issues could be resolved to the satisfaction of Māori.

Following these cautionary contributions, speakers from SPEaR, the Foundation of Research, Science and Technology, the Ministry of Research, Science and Technology, Statistics New Zealand and the Ministry of Social Development provided brief insights into the perspectives of a range of government stakeholders. While there was evidence of hopeful initiatives – for example, Statistics New Zealand’s proposed Official Statistics Research and Data Archive and SPEaR’s advocacy of an on-line data-sharing portal – the abiding impression was more of a potential unfulfilled. This sentiment was captured by Henry Barnard of Massey University and pioneer there of the now-defunct Social Science Data Archive. He displayed evidence since the 1970s of reports and bureaucratic interest in data preservation and sharing, repeatedly advocated but never acted upon, over that 30-year period.

THEMES AND ISSUES

The motivation for the conference was the need to bring the New Zealand science and science-policy community up-to-date with a major international development. A number of factors lie behind this momentum. One of them is a new professionalism about science management and policy, together with the high stakes that governments now have in the sector for economic, policy and even security reasons. However, it is also true that new technologies emerging in the natural sciences have revived interests in old collections from which new value can be extracted (in genomics, for example). Add to this trends in “e-science” (science harnessing the potential of the internet and information technology), digitisation, and grid technology, and we in the New Zealand science and policy community suddenly realise that there is a major international movement in full swing about which we need to be better informed. Thus the

conference raised the issue of how New Zealand can retain its international standing when important developments in science, sometimes right out of our league, can gain momentum so fast that they can nearly pass us by altogether.

For all the positives about data saving and sharing – cost effectiveness, transparency, scientific asset management, maximising the enormous potential of new technology – there were also some snags and snares, most of which came in the implementation of these policies. It was evident that it was essential for funding agencies to take their scientific communities with them, as reflected in the misgivings aired by Ritchie Poulton in his presentation. Commitment to data preservation and sharing has hardly been instantaneous, entirely willing, or universally complied with. Nevertheless, the centre of gravity has certainly shifted towards more active data management, and the move is being reinforced by linking it to continued funding.

At the same time that both technological possibility and explicit policy are moving in the direction of saving and sharing research data, there is an opposite trend of almost equal force in public attitudes to the strengthening of data privacy, a trend that is being reinforced by active, professional advocates in law and ethics. Therefore, just as it is important to take the scientific community along with policy development in this area, it is perhaps even more vital to encourage informed public debate on the topic. The misgivings articulated by Chris Cunningham from a Māori perspective do not look very different from those presented on behalf of the wider community by advocates of data privacy. We should be able to learn from experience elsewhere and avoid some of the more obvious semantic and ethical pitfalls on this journey.

Finally, what of the mechanics of implementation – how expensive is it, which data sets to start with, and which to target for treatment? Our best estimate to date is that the cost of saving data adequately amounts to about 1% to 2% of overall research expenditure (in the aggregate), although there are undoubtedly up-front costs that need to be met to set up systems and deal with backlogs. To decide which data sets to start with requires a judgement on scientific value. There may even be circumstances where data sets have to be rescued, and perhaps these should be among the first to be targeted, along with collections of historic significance. It is also evident that some data sets in archives are barely touched, while others are heavily accessed. Again, scientific judgement needs to be applied. Nevertheless, it would appear that good scientific practice in any investigation would require standards of protection, documentation and ease of access that should not be too far below those required of specially archived data sets. Grid technology and e-science developments may well produce for us distributed networks of access to scientific data without the need for major investment, but rather the establishment of standardised protocols.

SOME CONCLUDING PERSONAL OBSERVATIONS

One intriguing insight from the excursion into the world of digital data and data preservation, sharing and analysis was the appreciation that “archiving” applies as much now to qualitative as to quantitative data. Indeed, there is a UNESCO requirement on us to protect our cultural heritage, and digital technologies are likely to be a key mechanism. Protecting data applies not just to matters of research interest, but right across the cultural spectrum, and qualitative data are likely to outweigh quantitative and more traditional scientific items in this respect.

A second insight – at least to this observer – was the fragmented nature of the New Zealand science estate. For a small country we have a remarkable proliferation of single-purpose agencies in the research field. Each must have its CEO, each its own funding and assessment policies, each its own circle of lobby groups. If New Zealand is to assess this international science trend and respond to it, a much more united scientific response is required, and also a “whole of government” approach. Not only the traditional science agencies, but also the National Library, Statistics New Zealand, even local government, all may need to contribute in some way and assist with a consensual and comprehensive solution.

Finally, a striking aspect of this whole episode is the abiding weakness of the social sciences alongside the heavy hitters in the natural sciences: biomedicine, the Crown Research Institutes, the environmental sciences and industrial research. Public funding of social science research – at least at any organised and visible level outside the Marsden fund – amounts to little more than 1% of the funding of the Foundation of Research, Science and Technology, to be compared with nearly two-thirds going to business and technical research and development and one-fifth to the environment. How can a tiny portfolio, bolted on to the business and environment portfolio, gain the leadership and nourishment it needs? As the presentation by Henry Barnard demonstrated, this has been an issue of long standing. Its legacy is weak infrastructure and an inadequate contribution to informing policy and public debate.