Case Study

1

Developing and Implementing the Wellcome Trust's Data Management and Sharing Policy

Author: David Carr (Programme Manager – Open Research, Wellcome Trust)
Email: d.carr@wellcome.ac.uk

1.1. MAXIMISING THE VALUE OF RESEARCH DATA: A KEY PRIORITY FOR RESEARCH FUNDERS

There is a strong and growing consensus among research funders over the need to ensure that data outputs resulting from the research we support are managed and shared in ways that will deliver the greatest benefit to society. Over recent years, funders around the world have introduced policies requiring that their funded researchers make data available to others in a timely and responsible manner, and plan their approach for managing data as an integral part of planning their research.

The Wellcome Trust[1] is a global research foundation dedicated to improving health for everyone through enabling great ideas to thrive. This case study summarises our experience in implementing our data management and sharing policy over the last decade, drawing out some of the key lessons and remaining challenges.

1.2. WHY DOES DATA SHARING MATTER TO RESEARCH FUNDERS?

In common with other research funders, Wellcome's work to encourage data sharing is driven in large part by a recognition that much of the data currently generated by research represents a vast untapped resource. Enabling researchers and other users to access, combine and use data could open up new avenues for discovery and innovation that might never have been anticipated by the original data generators. In addition, access to the data underlying research findings is critical to ensure that these claims can be scrutinised and reproduced. Data sharing also holds the potential to help reduce avoidable duplication and waste – helping to enable research funds to be allocated effectively and enhancing the efficiency of the research enterprise.

1.3. DEVELOPMENT OF WELLCOME'S DATA MANAGEMENT AND SHARING POLICY

Wellcome's policy on data management and sharing[2] was published in January 2007, and was updated following a review in 2010. It followed two years after the introduction of our policy on open access to research publications[3], and built on Wellcome's work over many years to develop key data resources for the benefit of the research community – notably in the genomics field where we took a lead role in ensuring the data generated in the Human Genome Project was made immediately available, with no restrictions on its use.

Unlike our open access policy, where we were able to set out a clear mandate that all original research papers we fund be made open access within six months of publication, our data management and sharing policy allows for a case-by-case approach. Whilst a research article is a single type of output for which

a consistent rule could be applied, the optimal approach for sharing data outputs will vary dramatically depending on the nature of the data and the research context. Furthermore, there are some types of data which cannot be shared openly and where limits on sharing are required – particularly data relating to human research participants.

1.4. KEY POLICY PROVISIONS

Our data management and sharing policy is very similar to those of other major funders – including the UK Research Councils and US National Institutes of Health. The key elements are as follows:

- We expect all of our funded researchers to maximise the availability of research data with as few restrictions as possible;
- We require applicants for funding to submit a data management and sharing plan in cases where their proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community;
- We commit to review data management and sharing plans, and any associated costs to deliver them, as an integral part of the funding decision;
- We expect all users of research data to acknowledge the sources of their data and to abide by the terms and conditions under which they accessed the original data.

1.5. OUR EXPERIENCE IN IMPLEMENTING THE POLICY

Over the ten years the policy has been in place, we have undertaken periodic reviews to take stock of the data management and sharing plans being submitted and to gauge the perspectives of researchers and reviewers. Following our review of the policy in 2010, we introduced more detailed guidance for applicants[4] on developing data management and sharing plans – structured around seven key questions that plans should address (Figure 1.1). The overall quality of plans has improved over time, but plans still vary significantly in terms of their levels of detail and specificity. Particularly in areas of research where data sharing resources and practices are less well-established, many researchers and reviewers still do not feel there is sufficient clarity on our expectations of researchers or, in many cases, how best to put data sharing requirements into practice.

1. What data outputs will your research generate?
2. When do you intend to share your data?
3. Where will the data be made available?
4. How will your data be accessible to others?
5. Are limits on data sharing required (including any intended to protect research participants or safeguard intellectual property)?
6. How will key datasets be preserved?
7. What resources are required to deliver the plan?

Figure 1.1 Wellcome guidance on developing a data management and sharing plan: key questions

[1]   http://wellcome.ac.uk; last accessed 5 February 2017.
[2]   http://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing; last accessed 5 February 2017.
[3]   http://wellcome.ac.uk/funding/managing-grant/open-access-policy; last accessed 5 February 2017.
[4]   http://wellcome.ac.uk/funding/managing-grant/developing-data-management-and-sharing-plan; last accessed 5 February 2017.

At present, it is also not clear that the costs of implementing data plans are always being fully factored into funded applications. In 2016, we updated our guidance to give greater specificity on the costs that could be requested in terms of people and skills, data storage and computation, data access, data preservation and deposition.

Finally, our ability to monitor the extent to which researchers put their plans into practice is currently limited. While data sharing does form a key criterion in decisions over renewals for major resources and databases we support, we do not have a process, nor the in-house resources, systematically to track the delivery of plans across the bulk of research we support.

1.6. WIDER CHALLENGES IN DATA SHARING

Different research disciplines are at very different stages in developing the resources and practices required to support data sharing. Several major barriers persist which must be overcome if funder policies to maximise the value of research data are to be successful – key amongst these are:

- **Infrastructure and tools** – building and sustaining the technical resources and tools needed to store, access and analyse vast and complex research datasets;
- **Culture and incentives** – fostering a cultural shift to ensure data sharing is valued and rewarded appropriately;
- **Capacity and skills** – developing the skills necessary in the research community to manage and analyse data effectively;
- **Ethics and governance** – establishing the policy frameworks to ensure data sharing is ethical and fair, and maintains public trust.

1.7. WORKING IN PARTNERSHIP

To take on these challenges, Wellcome's approach over recent years has been to focus on strategically important research areas where we believe there is the potential to work with the community to advance data sharing, and to forge partnerships with other funders to drive change. For example, we have:

- established the Expert Advisory Group on Data Access[5] in partnership with MRC, ESRC and Cancer Research UK to provide strategic advice to funders on data access for cohort and longitudinal studies across genetics, epidemiology and social sciences;
- joined with a consortium of pharmaceutical companies to support the ClinicalStudyDataRequest.com[6] platform to enable research access to clinical trials data;
- taken a lead role in working with funders and journals to drive the rapid sharing of research data related to public health emergencies[7].

1.8. EMERGING PRIORITIES

Wellcome is actively exploring how we can build on the work we have done to champion data sharing. In terms of our policy, we are likely to move towards a more holistic approach for the management of outputs. Rather than request a data management and sharing plan in isolation, we would ask researchers to outline a plan for managing and sharing any outputs of value (including software and research materials, as well as data) and to describe their approach where relevant to managing any associated intellectual property. In parallel, we are also actively exploring how to strengthen implementation of our data management and sharing policy – including defining more clearly the roles of reviewers and staff in assessing plans and developing a clearer template for plans. We are particularly keen to explore the scope to work with other funders to implement machine-actionable data management plans that dynamically update as data outputs are generated.

Wellcome is also actively developing new opportunities to advance the broader goals of open research and pilot innovative models to push the boundaries of openness – building on our work to establish the Wellcome Open Research[8] publishing platform and Open Science Prize[9].

1.9. CONCLUSIONS

Over the last five to ten years, there has been a growing international recognition of the crucial importance of maximising access to research data. There is a strong policy alignment between major research funders, but significant challenges remain in implementing these policies in practice. Based on Wellcome's experience, key issues for funders to consider in developing and implementing data sharing policies include the need to:

- clarify expectations for researchers as far as possible; and to develop guidance tailored to specific research fields and data types in terms of current best practice for data management and sharing, and the resources available;
- establish a clear process for reviewing and assessing data management plans and the associated costs, and a proportionate mechanism to track plans post-award;
- develop new mechanisms for funders to recognise and reward the contributions of researchers who generate and share high quality datasets and initiate a broader cultural shift;
- consider how best to work in partnership at national and international level to:
    - build and sustain repositories, standards and tools to support data sharing;
    - develop the skills and capacity needed to manage, share and analyse data;
    - harmonise policies and practices wherever possible;
    - advocate and champion the ongoing transition to open science approaches.

[5] http://wellcome.ac.uk/what-we-do/our-work/expert-advisory-group-data-access; last accessed 5 February 2017.
[6] http://www.clinicalstudydatarequest.com; last accessed 5 February 2017.
[7] http://wellcome.ac.uk/what-we-do/our-work/data-sharing-public-health-emergencies; last accessed 5 February 2017.
[8] http://wellcomeopenresearch.org; last accessed 4 March 2017.
[9] http://www.openscienceprize.org; last accessed 4 March 2017.