

Summary of Funders' data sharing policies

Public Health Research Data Forum Review of funders' data sharing policies

Introduction

This work was originally undertaken by the Wellcome Trust in 2012 as an in-kind contribution to the Public Health Research Data Forum, in order to develop a shared resource to inform best practice around the management and sharing of research data. We reviewed and updated the information in December 2013, and would like to discuss with partners how we can use and build on this work moving forward.

This report provides a short summary of the key findings from the project and highlights areas of best practice. It should be read in companion with the summary of individual funders' policies (attached as Annex A) and the summary table highlighting the main principles of the policies (Annex B).

Methodology

In compiling this report, the data sharing policies of individual funders were reviewed alongside the findings from the survey on data sharing undertaken by the Wellcome Trust in summer 2011, and follow-up communications with individual funders where needed. A number of key principles common to many of the policies were identified, and are summarised below as examples of good practice, while more detailed summaries of the policies themselves are included in the attached annexes.

Principles of good practice in data management and sharing

Of the 20 organisations that make up the forum, 11 have dedicated policies on data sharing and management. Of the remaining nine funders, some have guidance in place of an explicit policy, or have endorsed an external agreement; some of these funders also have plans to develop a dedicated policy (these are highlighted where relevant). From this broad base of information, it was possible to formulate the broad principles outlined below; it should be noted that these do not apply uniformly to every funder, but provide a basis on which to compare funders' policies and guidance, and identify commonalities and differences to inform best practice.

1. Data management plans

The majority of the policies considered require or recommend a data management plan to be submitted, as either an integral or recommended part of a research proposal. The circumstances in which a plan is required and requirements as to its content vary between funders, but most at a minimum require researchers to describe the expected data outputs of their research, how these will be collected and stored, and how they will be made available to other researchers. These requirements tend to be more descriptive guidance rather than fixed requirements; guidance from the Wellcome Trust on the development of a data management and sharing plan notes that there is no set format for a data sharing plan, and that researchers should structure their plan "in a manner most appropriate to the proposed research." In most cases, funders cite the formulation of a data management plan as a principle of good practice rather than a set of fixed requirements; the exception to this is the

UK Medical Research Council, which for MRC-funded cohort studies sets out 21 mandatory requirements that a data sharing plan must address.

2. Time frames

The question of whether specific time limitations are placed on researchers with regard to the sharing of data varies between funders. Some funders, such as the Bill and Melinda Gates Foundation and the National Institutes of Health, state that expected timescales for data sharing should be built into data sharing plans. However, others make explicit expectations for timescales, such as the Centers for Disease Control and Prevention (CDC) which states that the release of data should occur “no more than one year” after plans for data sharing are laid out in a plan.

3. Use of public databases and central repositories

Several of the policies reviewed make reference to the use of public databases and/or repositories for the depositing or storage of data, in order for them to be freely accessible to other researchers. For example, the Wellcome Trust, the CDC and the Canadian Institutes of Health Research (CIHR) all include provisions for the use of central repositories for data sharing, where these are available. The nature of these repositories or databases vary; the World Bank and the Hewlett Foundation both make reference to specific data repositories set up by their own organisations, either directly or through partnerships, while the UK Economic and Social Research Council funds the UK Data Archive, which curates the largest collection of digital data in the social sciences and humanities in the UK. Other funders do not specify the nature or location of such repositories, although most require that they be online and/or open access, to facilitate data sharing.

4. Ethics and confidentiality requirements

A major component of the policies reviewed was the inclusion of requirements around ethics and confidentiality, and the protection of research data containing identifiable patient information. A majority of funders who have data sharing policies cite the importance of ethical conduct with regard to data sharing, and of protecting confidentiality. Provisions set out for achieving this include anonymising or redacting data where research participants could potentially be identified, controlling access to sensitive data, and building provision for data sharing into consent procedures prior to the commencement of research studies.

Most policies expect researchers, at the very least, to have considered measures designed to protect the confidentiality of research participants and to ensure that their research upholds high ethical standards – examples of specific measures included in funders’ individual policies are outlined in Annex A.

5. Compliance

The question of whether and how funders ensure compliance with their data sharing policies was more varied, and thus less easy to draw broad conclusions from. Most funders who have explicit policies have at least touched on the issue of how they will track and ensure compliance with their policies, to the extent that this is required. In most cases the provision of data sharing measures as part of a data management plan is considered as part of the grant application process, and so any applicants failing to comply with data management requirements (when necessary) may face the possibility of their application not being taken forward on these grounds. Where compliance is mentioned as a specific measure, it is in

the context of past compliance as a measure of success in future funding decisions; for example, the policy of the Bill and Melinda Gates Foundation states that “a grantee’s adherence to previous [data access] plans could be a factor in ongoing funding decisions.”

In some cases, funders will make reference to specific funding which is available for data sharing and management. For example, the National Institutes of Health states in its policy that applicants may request funds to cover the costs of data sharing and management, while the Wellcome Trust will allow applicants to include such costs as part of their research proposal.

Summary

Clearly organisations will vary widely in their policies and practices, but the findings of this project show that a number of broad common principles of data sharing and management are common to many of the member organisations in the Forum. It is hoped that this report, by highlighting these principles, may help those organisations to identify areas of best practice within their own policies and guidance, and to develop these further in the future.

Annex A – Summary of funders’ data sharing policies

This annex provides a summary of the funders’ policies considered as part of this project, where available, with summaries of their key points and any associated guidance.

Wellcome Trust

Policy on data management and sharing

<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

- The Wellcome Trust expects all of its funded researchers to maximise the availability of research data with as few restrictions as possible.
- In cases where the proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community, applicants will be required to submit a data management and sharing plan to the Wellcome Trust prior to an award being made.
- The Trust will review data management and sharing plans, and any costs involved in producing them, as an integral part of the funding decision, and work with grant holders on an ongoing basis to support them in maximising the long-term value of key datasets resulting from their research.

Associated guidance

<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm>

Guidance sets out the circumstances under which data management and sharing plans are required, and expectations for the issues that these plans should address. A data management plan is expected where a proposal involves the generation of datasets that have clear scope for wider research use and hold significant long-term value.

UK Medical Research Council

Policy on data management and sharing

<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/index.htm>

- MRC’s overarching aim for data-sharing is to maximise the life-time value of research data assets for human health and to do so timely, responsibly, with as few restrictions as possible, in a way consistent with the law, regulation and recognised good practice.
- The MRC’s policy on data sharing does not, except for MRC-funded cohort studies, prescribe when or how researchers should preserve and share data but requires them to make clear provision for doing so when planning and executing research.
- Potential research benefits to patients and the public should outweigh identified risks. Risks such as inappropriate disclosure of personal information must be managed in a proportionate yet robust manner

Associated guidance

<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharing/DMPs/index.htm>

- Sets out expectations of issues to be addressed by a data management plan.
- For MRC-funded cohort studies these plans must address 21 mandatory requirements:
<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharing/policy/PHSpolicy/requirements/index.htm>

UK Economics and Social Science Research Council

Policy on data management and sharing

General information: <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>

Full policy at: http://www.esrc.ac.uk/images/Research_Data_Policy_2010_tcm8-4595.pdf

- Research data created as a result of ESRC-funded research should be openly available to the scientific community to the maximum extent possible, through long-term preservation and high quality data management.
- While ESRC strongly advocate further re-use of data it is also recognised that some research data are more sensitive than others and argue that it is a responsibility of the award holders to consider all issues related to confidentiality, ethics, security and copyright before initiating the research.
- Requires that the data must be made available for preparation for re-use and/or archiving with the ESRC data service providers within three months of the end of the award otherwise ESRC will withhold the final payment.
- It is then a responsibility of the award holder to incorporate data management as an indivisible part of the research project to increase the potential for data to be shared.

Associated guidance

http://www.esrc.ac.uk/images/Data-Management-Plan-Guidance-for-peer-reviewers_tcm8-15569.pdf

- A data management plan should include: assessment of existing data, information on new data, quality assurance of data, back-up and security of data, expected difficulties in data sharing, copyright/Intellectual Property Right, responsibilities, preparation of data for sharing and archiving
- Guidance sets out requirements with regard to formats in which data are presented, and measures to protect confidentiality.

Bill and Melinda Gates Foundation

Policy on data management and sharing

<http://www.gatesfoundation.org/global-health/Documents/data-access-principles.pdf>

- As a condition of a grant award from the Bill & Melinda Gates Foundation and in support of the foundation's Global Access policy, all prospective grantees will be required to facilitate the prompt and broad dissemination of data. For grants over \$500,000, this will begin with development of a Data Access Plan to specify how data will be managed and accessed
- Prospective grantees are encouraged to consult with their Program Officer in developing a suitable Data Access Plan.
- This plan will be guided by the following principles: Promotion of the Common Good, Respect, Accountability, Stewardship, Proportionality, Reciprocity

Associated guidance

<http://www.gatesfoundation.org/global-health/Documents/faq.pdf>

- Sets out requirements of a Data Access Plan, including timing, manner of storage and who will have access.
- Data should be made widely available as soon as possible while safeguarding the rights and privacy of individual subjects. It is the responsibility of the investigators, their IRB, and their institution to protect the rights of participants and the confidentiality of their data; guidance sets out measures that could be used to protect privacy.

US National Institutes of Health

Policy on data management and sharing

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>

- Data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health.
- Investigators submitting an NIH application seeking \$500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.
- Reviewers will not factor the proposed data-sharing plan into the determination of scientific merit or priority score.

Associated guidance

http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

- The precise content of the data-sharing plan will vary, depending on the data being collected and how the investigator is planning to share the data – applicants planning to share data may wish to describe the timescales and measures by which data may be shared, and whether a data-sharing agreement will be required.
- As NIH stated in the March 1, 2002 draft data sharing statement the rights and privacy of people who participate in NIH-sponsored research must be protected at all times.

- The “timely release and sharing” of data is defined as being no later than the acceptance for publication of the main findings from the final data set.

CDC

Policy on data management and sharing

<http://www.cdc.gov/maso/policy/releasingdata.pdf>

- CDC believes that public health and scientific advancement are best served when data are released to, or shared with, other public health agencies, academic researchers, and appropriate private researchers in an open, timely, and appropriate way.
- The interests of the public—which include timely releases of data for further analysis—transcends whatever claim scientists may believe they have to ownership of data acquired or generated using federal funds. Such data are, in fact, owned by the federal government and thus belong to the citizens of the United States.
- The goal is to have a policy on data release and sharing that balances the desire to disseminate data as broadly as possible with the need to maintain high standards and protect sensitive information. To ensure that issues of confidentiality, proprietary use, and informed consent are addressed correctly, CIOs may choose to develop specific data release plans for each data set.

Associated guidance

- Sets out guidance on what plans should cover, including: procedures to protect confidential data; instructions for non-CDC users on the appropriate use of the data; the formats in which the data will be released; and the date by which the data will be released, which should be as soon as possible after they are collected.
- This release should occur no more than one year after these activities.

<http://www.cdc.gov/od/science/integrity/docs/cdc-policy-human-research-protections.pdf>

- All of CDC’s human research activities will be guided by the ethical principles of respect for persons, beneficence, and justice as set out in the “Human Research Protections” Policy.

DFG

As part of the Alliance of German Science Organizations the DFG has published a general policy statement under http://www.allianzinitiative.de/en/core_activities/research_data/

- Data sharing policy is an integral part of the DFG's funding policy and is echoed in the DFG's guidelines. If research data will be systematically produced using DFG project funds, applicants are requested to describe what measures will be implemented to ensure their management, curation and long-term preservation for future reuse.
- Within each proposal, data management measures are requested for the individual project and those are subject to peer review. This will also be the subject of subsequent project reporting. Funding will be supplied for the measures that are suggested in the individual proposals and approved by peer review.

DFG's guidelines for project proposals also contain measures for data management and sharing:

2.4 Data handling

- Improving the handling of research data is a priority both for national and international research organisations and for science in general. In order to enhance the long-term preservation of research data, the DFG funds projects that seek to achieve an efficient and sustained use of research data.
- If research data will be systematically produced using DFG project funds, describe what measures will be implemented to ensure their management, curation and long-term preservation for future reuse. Please regard existing standards and data repositories in your discipline where appropriate.”

Doris Duke Charitable Foundation

Policy on data management and sharing

The Foundation has drafted a series of policies and guidelines for the Population Health Implementation and Training (PHIT) Partnership Data Collaborative, a key part of its African Health Initiative. These include a policy on data management and sharing, attached as Annex A at:

<http://www.ddcf.org/Global/11%200104%20PHIT%20Data%20Collaborative%20Principles,%20Policies%20and%20Procedures.pdf>

- All of the information which is regularly collected area-wide by all of the PHIT Partnerships (referred to as core data) will be submitted in a timely manner to a data repository and made available to all other PHIT Partnership members.
- An adequate data sharing plan and the timely submission of data will be an important aspect of both the grantee selection process and continued funding.
- Collaborative databases will be developed and maintained under the supervision of the Data Coordinator (the central coordinating body of the PHIT Data Collaborative) in consultation with the Collaborative Management Committee. Each Partnership will be responsible for submitting their data sets along with sufficient explanatory information to allow reanalysis.
- The Data Coordinator will perform a series of quality control checks, and consult with the Partnerships as needed.

Hewlett Foundation

The Hewlett Foundation does not currently have a specific policy on data management and sharing. However, the Foundation supports data sharing by funding the International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries (INDEPTH), a global network of members who conduct longitudinal health and demographic evaluation of populations in low- and middle-income countries (LMICs). Efforts are underway to produce an INDEPTH data sharing policy that will be subscribed to by all its member centres.

World Bank

Open Access policy

<http://documents.worldbank.org/curated/en/2012/04/16200740/world-bank-open-access-policy-formal-publications>

- The World Bank supports the free online communication and exchange of knowledge as the most effective way of ensuring that the fruits of research, economic and sector work, and development practice are made widely available, read, and built upon. It is therefore committed to open access, which, for authors, enables the widest possible dissemination of their findings and, for readers, increases their ability to discover pertinent information.
- For work carried out by Bank staff, the policy applies to manuscripts and all accompanying data sets (a) that result from research, analysis, economic and sector work, or development practice; (b) that have undergone peer review or have been otherwise vetted and approved for release to the public; and (c) for which internal approval for release is given on or after July 1, 2012. For external research funded by the Bank, for which funding was approved on or after July 1, 2012, the policy applies to the final report provided by the researchers to the funding unit within the Bank.

Canadian Institutes of Health Research

CIHR Policy on Access to Research Outputs - <http://www.cihr-irsc.gc.ca/e/32005.html>

CIHR believes that greater access to research publications and data will promote the ability of researchers in Canada and abroad to use and build on the knowledge needed to address significant health challenges. Only when research findings are widely available, enabling open scrutiny, will this evidence be translated into policies, technologies, health-related standards and practices, and new avenues of research that will benefit the health of Canadians and others.

Researchers awarded new or renewed funding from CIHR are reminded to adhere with the following responsibilities:

- to ensure that all research papers generated from CIHR funded projects are freely accessible through the Publisher's website or an online repository within 12 months of publication;
- to deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database (e.g. gene sequences deposited in GenBank) immediately upon publication of research results;
- to retain original data sets for a minimum of five years (or longer if other policies apply);
- to acknowledge CIHR support by quoting the funding reference number in journal publications.

Associated guidance (FAQ) - <http://www.cihr-irsc.gc.ca/e/35664.html>

National Health and Medical Research Council (Australia)

Revised Policy on Dissemination of Research Findings -

<http://www.nhmrc.gov.au/media/notices/2012/revised-policy-dissemination-research-findings>

- NHMRC acknowledges that researchers take into account a wide range of factors in deciding on the best outlets for publications arising from their research.
- Such considerations include the status and reputation of a journal, book, publisher or conference, the peer review process of evaluating their research outputs, access by other stakeholders to their work, the likely impact of their work on users of research and the further dissemination and production of knowledge.
- Taking heed of these considerations, NHMRC wants to ensure the widest possible dissemination of the research supported by NHMRC funding, in the most effective manner and at the earliest opportunity.
- NHMRC therefore requires that any publications arising from an NHMRC supported research project must be deposited into an open access institutional repository within a twelve month period from the date of publication.
- The revised policy is due to come into force on 1 July 2012.

See also: Australian Code for the Responsible Conduct of Research -

<http://www.nhmrc.gov.au/guidelines/publications/r39>

Section 4 covers publication and dissemination of research findings:

- Researchers have a responsibility to their colleagues and the wider community to disseminate a full account of their research as broadly as possible.
- The account should be complete, and, where applicable, include negative findings and results contrary to their hypotheses.

INSERM (France)

A general charter (available in French only) on open cohort data has been established by the French Public Health Research Institute (IReSP) as part of the national coordination of public health cohorts – this is a generic charter from which individual cohorts may establish their own charter and outline general principles that cohort teams must respect, and proposes principles on how to make cohort data available to the scientific community.

<http://iresp.net/index.php?goto=actions&rub=ecbd&prjld=4>

MRC South Africa

No specific policy, but book 1 of the Guidelines on Ethics for Medical Research (General Principles) contains guidance on research results and publishing:

<http://www.mrc.ac.za/ethics/ethicsbook1.pdf>

10.5 Research results

10.5.1 Responsibilities of investigators

Investigators have responsibilities to share possible benefits of research results with participants.

10.5.2 Rights of participants to results of research

The benefits of research are to be made available to the research population and the local communities from which they were drawn, and adequate reports of the research must be made publicly accessible within a reasonable period of time.

10.7 Publication and authorship

10.7.1 Introduction

Those engaged in research have a moral obligation to share their findings with other investigators, clinicians and society, for the mutual benefit of all. There are personal pressures on investigators to publish, and institutions in South Africa also benefit from the volume and quality of their research output.

Scottish Chief Scientist's Office

Standard conditions of grants contain guidance on data sharing:

<http://www.cso.scot.nhs.uk/wp-content/uploads/2013/02/Standard-Conditions-of-a-Research-Grant.doc>

Section 16 on Preserving and Sharing Research Data states that:

16.1 CSO, in common with other public research funders, strongly encourages the sharing of data from research it supports. Where the data may be of interest to researchers other than the original investigators, consent from research participants should be worded in terms that enable the data to be used for secondary analysis, and datasets should be preserved in a way that encourages other analysts to use them. The best method for ensuring this is to deposit the data with full supporting documentation in a public archive, such as the UK Data Archive.

16.2 CSO recognises that the original investigator has a right to a limited period of exclusive use of the data that secondary analyses may be most fruitfully conducted in collaboration with the original investigator, and that publications making secondary use of the data should acknowledge the intellectual property of the original investigator.

16.3 Whether or not the data is likely to be used for secondary analysis, the Chief Investigator must ensure that the raw data or results are stored for a minimum period of 5 years after completion of the project. At any time during this period the data or results may be requested by CSO. If a longer period of storage is required this will be indicated in the notice of funding.

Annex B – Summary table of data sharing policies

Organisation name	Does the policy describe requirements or guidance for a data management plan?	Does the policy specify time frames for data sharing?	Does the policy have requirements for the use of public databases or repositories?	Does the policy have requirements regarding ethics or the protection of confidentiality?	Does the policy have specific measures to track or ensure compliance?
Bill & Melinda Gates Foundation	Yes (for grants over \$500k)	No	Yes	Yes	Yes
Canadian Institutes of Health Research	No	Yes	Yes	No	No
Centers for Disease Control and Prevention	Yes	Yes	Yes	Yes	Yes
DFG, German Research Foundation	Yes	No	No	No	Yes
Doris Duke Charitable Foundation	Yes	Yes	Yes	Yes	Yes
Economic and Social Research Council, UK	Yes	Yes	No	Yes	Yes
Medical Research Council, UK	Yes	No	No	Yes	No
NHMRC, Australia	No	Yes	Yes	No	No
National Institutes of Health	Yes	Yes	No	Yes	Yes
Wellcome Trust	Yes	Yes	Yes	Yes	Yes
World Bank	No	Yes	Yes	No	No

The following organisations do not currently have specific policies on data management and sharing, but in many cases are working to develop policies and guidance in this area:

- Chief Scientist's Office, Scotland
- FAPESP, Brazil
- Health Research Council of New Zealand
- Hewlett Foundation
- INSERM, France

- MRC South Africa
- UNICEF
- USAID
- WHO