

# Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers

Cameron Neylon ‡

‡ Centre for Culture and Technology, Curtin University, Perth, Australia

Corresponding author: Cameron Neylon ([cn@cameronneylon.net](mailto:cn@cameronneylon.net))

Reviewable v1

Received: 23 Jun 2017 | Published: 23 Jun 2017

Citation: Neylon C (2017) Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. Research Ideas and Outcomes 3: e14673.

<https://doi.org/10.3897/rio.3.e14673>

## Abstract

There is a wide and growing interest in promoting Research Data Management (RDM) and Research Data Sharing (RDS) from many stakeholders in the research enterprise. Funders are under pressure from activists, from government, and from the wider public agenda towards greater transparency and access to encourage, require, and deliver improved data practices from the researchers they fund.

Funders are responding to this, and to their own interest in improved practice, by developing and implementing policies on RDM and RDS. In this review we examine the state of funder policies, the process of implementation and available guidance to identify the challenges and opportunities for funders in developing policy and delivering on the aspirations for improved community practice, greater transparency and engagement, and enhanced impact.

The review is divided into three parts. The first two components are based on desk research: a survey of existing policy statements drawing in part on existing surveys and a brief review of available guidance on policy development for funders. The third part addresses the experience of policy implementation through interviews with funders, policy developers, and infrastructure providers.

In our review we identify, in common with other surveys, that RDM and RDS policies are increasingly common. The most developed are found amongst funders in the United States, United Kingdom, Australia, and European Union. However many other funders and nations have aspirational statements or are developing policy. There is a broad pattern of policy development moving from aspiration, to recommendations, to requirements, and finally reporting and auditing of data management practice.

There are strong similarities across policies: a requirement for data management planning, often in grant submissions, expectations that data supporting published articles will be made available, and in many cases requirements for data archiving and availability over extended periods beyond grants. However there are also important differences in implementation.

There is essentially no information available on the uptake and success of different policies in terms of compliance rates, or degrees of data availability. Many policies require a Data Management Plan as part of grant submission. This requirement can be enforced but there is disagreement on the value of this. One view is that requirements such as DMPs are the only way to force researchers to pay attention to these issues. The other is that such requirements lead to a culture of compliance in which the minimal effort is made and planning is seen as a “tick-box” exercise that has no further value. In this view requirements such as DMPs may actually be damaging the effort to effect culture change towards improved community practice.

One way to bring these two views together is to see DMPs as living documents that form the basis of collaboration between researchers, funders, and data managers throughout the life of a research project. This approach is reflected in guidance on policy development that emphasises the importance of clarifying responsibilities of various stakeholders and ensuring that researchers are both recognised for good practice and see tangible benefits.

More broadly this points to the need for the program of improving RDM and RDS to be shared project with the incentives for funders and researchers aligned as far as is possible. In the interviews successful policy implementation was often seen to be dependent on funders providing the required support, both in the form of infrastructure and resourcing, and via the provision of internal expertise amongst program managers. Where resources are limited, leveraging other support, especially from institutional sources, was seen as important as was ensuring the scope of policy requirements were commensurate with the support available and readiness of research communities.

Throughout the desk research and the interviews a consistent theme is the desire for cultural change, where data management and sharing practices are embedded within the norms of behaviour for research communities. There is general agreement that progress from aspirational policies to achieving compliance is challenging and that broad cultural change, with the exception of specific communities, is a long way off. It is interesting to note that discussion of cultural change is largely externalised. There is little engagement with the concept of culture as an issue to consider or work with and very little engagement with models of how cultural change could be enabled. The disagreement over the value of

DMPs is one example of how a lack of active engagement with culture and how it changes is leading to problems.

## **Key Findings**

Policies on RDM and RDS are being developed by a number of agencies, primarily in the Global North. These policies are broadly consistent in aspiration and outlines but differ significantly in details of implementation.

Policies generally develop along a path starting with aspirational statements, followed by recommendations, then requirements, and finally auditing and compliance measures.

Measurement of policy adoption and compliance in terms of the over goals of increased availability and re-use of data is not tracked and is likely unmeasurable currently.

Data Management Plans are a central requirement for many policies, in part because they can be made compulsory and act as a general focus for raising awareness.

There are significant differences in the views of stakeholders on the value of Data Management Planning in its current form.

Some stakeholders regard them as successful in raising awareness albeit with some limitations.

Some regard them as actively damaging progress towards real change in practice by making RDM appear as one administrative activity among the many required for grant submission

Successful policy implementation is coupled with funder support for infrastructure and training. Seeing RDM as an area for collaboration between funders and researchers may be valuable

Internal expertise and support within a funder is often a gap which becomes a problem with monitoring and implementation

DMPs can be a helpful part of process but it will be important to make them useful documents throughout and beyond the project

If the object of RDM and RDS policy is cultural change in research communities then direct engagement with understanding the various cultures of researcher and other stakeholder communities, alongside frameworks of how they change is an important area for future focus.

## **Keywords**

review, data management planning, data sharing, policy, implementation, culture change, policy design

## Introduction: Aims and Scope

This review was written in support of the International Development Research Center (Canada) program *Exploring the opportunities and challenges of implementing open research strategies within development institutions* (Neylon and Chan 2016). The project was constructed as a pilot in which the proposed IDRC Data Sharing policy is tested in the context of eight funded research projects. The intent of the review is to support the project and its participants by examining the existing literature on funder policies on research data management and sharing, and, using interviews with relevant experts, to develop an understanding of the current state of policy implementation and its challenges.

The review is focussed on funder policy, as opposed to policies of research institutions or disciplinary communities. It aims to develop an overview of the funder policy landscape as a whole, examine existing guidance for funders on policy development, and to probe the issues that funders, and the researchers and institutions that they fund, are facing in the implementation and adoption of policies.

As the development of data sharing and data management policy and practice amongst funders of development research has been limited to date, the review does not focus specifically on data sharing in a development research context. However it does draw out issues that are likely to be relevant in this context, including the provision of infrastructure, (mis)alignment between culture of research disciplines, institutions and stakeholders, and motivations of researchers and funders. The surfacing of these issues is intended to guide the research program in its examination of data sharing in the real world of the projects that are taking part.

## Research Funder Policies on Data Sharing and Management

Research funders are increasingly developing policies on research data management (RDM) and research data sharing (RDS) or Open Data. The oldest and most developed policies are in the United States, Australia and the United Kingdom. The European Union has imposed new requirements for the Horizon 2020 program and various member states are developing policies. Canada's Tri-Councils have been developing policy and infrastructure over the past several years.

A range of development research funders have also developed RDM and RDS policies. The UK Department for International Development included requirements for data management and sharing within the DFID Research Open and Enhanced Access Policy that came into effect in November 2012. The requirements included the provision of a Data Management Plan when proposals were submitted and datasets to be made available through institutional and subject repositories within 12 months of final data collection.

The World Bank Open Access Policy also includes requirements for datasets associated with formal publications to be made available through the Bank's Open Knowledge Repository. The World Bank also has a strong record of making general economic and research data available through its data portal and has a very strong disclosure policy in general relative to other similar organisations.

## Review methodology

Information on funder policies was sourced from the *Comprehensive Brief on Research Data Management* (Shearer 2015), *Current Best Practice for Research Data Management Policies* (Hodson and Molloy 2015), as well as two online collations of funder policies with a United States federal agency (Whitmire et al. 2015) and United Kingdom focus (Digital Curation Centre 2012). Specific policy documents for the U.K. Department for International Development (DFID 2013), the World Bank (World Bank 2012), the Australian Research Council and National Health and Medical Research Council, were also examined.

A focus was retained on funder policies as opposed to those of research institutions. National policies (with the exception of the United States OSTP Memo) and statements from trans-national bodies are very high level and generally aspirational rather than specific. As the review focusses on the impact of implementation details these were not considered except where they had a direct effect on the details of research funder requirements.

## Funder Motivations

The motivations for funders to develop data management and sharing policies are generally focussed around two broad issues. The first of these is maximising the impact and reach of research, the second is concerns around citizen access and engagement with research. Fecher and Friesike (Fecher and Friesike 2013) identified five categories of framing within the broader Open Science Movement; Democratic, Pragmatic, Infrastructure, Public, and Measurement. While issues might be raised with this classification, these framings are a useful lens for examining funder motivations to support data management and sharing. Funder policies and the documents around them often refer to the motivations for developing policy and guidance, and in the area of data these public motivations fall largely into the Democratic and Pragmatic categories with gestures towards the Public category. These motivations tend to also align with public statements of governments addressing broader Open Data agendas.

Other categorisations of drivers have been described (e.g Leonelli 2013, Borgman 2012) and policy documents enumerate a range of positive outcomes for RDS (OECD 2007, Royal Society 2012, Shearer 2015). The value of the categorisation of Fecher and Friesike is its basis in the analysis of the discourse of these documents as a means of interrogating the underlying motivations.

The Democratic framing is described as being concerned that access to knowledge, and the ability to re-use it, should be equitably distributed. It is distinguished from the Public framing, which is concerned with the public, that is non-professional engagement in the consumption and production of knowledge, rather than equity per se. Reference to these framings in policy statements is generally political, aligning with government language on democratisation and engagement.

For instance the UK Department for International Development in its Research Open and Enhanced Access Policy (DFID 2013) states that “DFID is committed to greater transparency in its activities and spending, and is working to make data more accessible to the public” and that “[the U.K. g]overnment is also committed to expanding access to publicly-funded research”. Similarly the World Bank policy on access to formal publications states that the 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” and that 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”.

In most cases the focus of policy language aligns strongly with the Pragmatic framing and elements of this are also seen in the examples above. This is often also described as “the impact agenda”; the goal of maximising value creation through funder investment in research. In the *Concordat on Open Research Data* (UK Open Data Research Forum 2015), a U.K. document developed by a multi-stakeholder group and endorsed by the UK Research Councils the first motivation for data sharing is that “[t]he societal benefits from making research data open are potentially very significant; including economic growth, increased resource efficiency, securing public support for research funding and increasing public trust in research”.

The Office of Science and Technology Policy of the U.S. Executive Administration starts its memorandum on *Increasing Access to the Results of Federally Funded Scientific Research* (White House Office of Science and Technology Policy 2013) with the statement that the “Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to *and useful* for the public, industry, and the scientific community” [emphasis added]. In the European Union the focus on “Big Data” as an economic good, in particular Vice President Kroes’ description of it as “the new oil” is also an element of this Pragmatic discourse.

Funder policy statements less frequently involve language that invokes the Infrastructure or Measurement framing identified by Fecher and Friesike. This raises questions as infrastructure and technology are clearly required to deliver on the goals of the Democratic and Pragmatic agendas and measurement of policy compliance is a key tool for driving uptake, or at least monitoring adoption.

A useful analysis is therefore to consider how, for any given policy, the language reflects the discourse associated with each of the five sets of discourse, the extent to which the substance of the policy addresses each school, and whether a funder is tracking success against the goals of each framing. A well balanced policy would require actions that align with the stated goals and measure success against those same goals. In practice most policies invoke the aspirations of the Democratic and Pragmatic discourse, require actions more aligned with the Infrastructure discourse and have weak reporting requirements. Where reporting requirements are stronger they rarely provide data that would support assessment of progress towards the goals of the discourse describing the Democratic and Pragmatic framing.

### **Specific funder policies**

A series of summaries of funder policies and their implementation exist, generally with a geographical focus (Shearer 2015, Whitmire et al. 2015, Digital Curation Centre 2012). The latter two of these are currently committed to maintaining updates and therefore will be a more reliable future source of information than the current review. Appendix 1 contains a table with summary information for a number of funders. Here a small selection of specific funder policies are discussed in detail as examples focussing on requirements for Data Management Planning, the support of costs for data management, and reporting and compliance requirements.

#### Engineering and Physical Sciences Research Council (UK)

The United Kingdom's EPSRC differs from most other funders in the focus of their policy implementation. It is unique in placing obligations for policy compliance firmly on the research institution as opposed to the researcher. EPSRC expects that data management planning will take place but does not require a DMP as part of a grant proposal. The institution is responsible for ensuring that data is stored appropriately and made available for a period of ten years, or for ten years after any third party request for access is made. The costs of data management are allowable expenses on grants, however the provision is necessarily institutional.

This focus on institutional provision, and the consequent threat that funding might be withdrawn at the institutional level for non-compliance is unique. The policy has definitely driven institutions to respond, although the effectiveness of those responses has been questioned by some. The EPSRC case demonstrates how the details of implementation, and particularly the assignment of responsibility can change outcomes, even when the actual requirements of the policy in terms of practice are very similar.

It is also instructive to consider the timing of responses. The council notified institutions in 2011 of the requirement to have a roadmap in place by May 2012 to deliver on the policy by May 2015. Nonetheless researchers and institutions largely did not engage with this until very late. Researchers are only now becoming aware of the requirements. This may in part be a consequence of not using a requirement for DMPs as part of grant submission to raise awareness amongst researchers.

## European Commission Horizon 2020 Program

Horizon 2020 is a seven year research program supporting around €70B of research across the European Research Area. The legislation framing the program called for a data sharing pilot program, following the success of an Open Access pilot in the previous research program (Framework Program 7).

Some data management planning is required for all proposals submitted within the main research programs of Horizon 2020. This is described as a “short, general outline” of the plans and is included within the assessment of Impact for the proposal, i.e. not as part of the assessment of research interest or quality. For grant calls included within the pilot (roughly 20% of the overall program) a more extensive DMP is required within six months of the project commencing (European Commission 2015a). There is emphasis on the role of the DMP as a living document to support the research project throughout its life.

The requirements place on the project are then to observe the DMP (although mechanisms for monitoring compliance are not specified), to deposit data in an appropriate repository, and to “as far as possible[...] take measures to enable for third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data” (European Commission 2015b). Timing of data release is not specified. Costs of data management and sharing are allowed as part of the grant and the Commission has supported a range of relevant data infrastructures.

As the program is still relatively new there have been few reports on progress within the Horizon 2020 Open Data pilot. While a proportion of projects have opted out of the pilot a substantial, though not equal, number have also opted in (Spichtinger 2015). It is likely too early in the implementation process to expect detailed information on progress but data is currently being collected on DMPs (OpenAccess\_EC 2016).

## Department for International Development (DFID, UK)

The United Kingdom’s main development funder developed a general policy on access to research outputs, including data in 2012 (DFID 2013). The policy requires datasets to be placed in an appropriate repository within 12 months of final data collection or on the publication of outputs underpinned by that data. Researchers are also “required to retain raw dataset for a minimum of five years at the end of a project and to make them available on request for free at any time after 12 months from final data collection (unless exempted by DFID)”.

DFID provides a repository, R4D for published outputs, which can accept “simple data sets” in some cases but the responsibility for storage and management is left to researchers and their institutions. An Access and Data Management Plan is required as part of proposals, and includes sections on data management. Publications are also required to a data availability statement stating where data is available and any restrictions on access.

As with many other policies the Open Data provisions are found within a wider Access Policy that includes access to other research outputs. The guidance provided specifically for data is limited and there is limited discussion of monitoring or compliance. A brief search for articles supported by DFID published in 2015 provided no examples where links were provided to data freely available in a repository and only one case where there was a formal third party to request access from. However a large proportion of articles appeared to provide free to read access to the full text in some form thus complying, in spirit at least, to that aspect of the policy.

#### World Bank

The World Bank has a strong record on making data available through its data portal. However it's policy on access to data generated by external research that it funds is limited to the inclusion of "associated data sets" in the definition of manuscripts that are included in its Open Access policy (World Bank 2012). This means requirements on external researchers are limited. The World Bank has both an extremely strong policy on public access to its internal information and records and as noted, a strong program on making internally generated data available. A similar informal test as for DFID found no cases where links to data were made clearly available in World Bank funded research.

#### National Institutes of Health (U.S.A.) and the OSTP Memorandum

The U.S. National Institutes of Health has a long history as a provider of data infrastructures and through its intramural research the generation of important research data for the community. It has a general policy on data sharing dating from 2003 requiring the preparation of a Data Management Plan for sufficiently large grants. The NIH with its strong history of funding the production of and access to *specific* types of data – in common with many other large biomedical funders – focusses more than most other policies on the existing repositories for specific data types.

The NIH in common with other large U.S. Federal Funders is subject to the Office of Science and Technology Policy memorandum on *Increasing Access to the Results of Federally Funded Research* (White House Office of Science and Technology Policy 2013). The memorandum is best known for its requirements on Open Access to articles but also includes a series of directives on the availability and management of research data. In its response, the NIH commits to expanding the requirement for DMPs to all grants.

The NIH already has a substantial investment (several hundred million dollars) in providing technical and infrastructural support for data management and sharing. It is unusual in this respect in having the scale and resources to tackle the support requirements head on. The response to the OSTP Memorandum also notes that Data Management Plans can form part of the formal Notice of Award and therefore be the basis for compliance action, including withholding funding. However monitoring is still a challenge and the response notes the importance of enhancing discoverability, particularly through metadata that "can be used to verify that novel data sets are registered in accordance with the applicable NIH

policy” (National Institutes of Health, USA 2015). It is notable that even the NIH has challenges in tracking data outputs of its funded research.

The OSTP Memorandum also includes a directive to “develop...appropriate attribution to scientific data sets that are made available under the plan”, focussing attention on the need for credit and recognition of good practice in data management and sharing (see also Hodson and Molloy 2015, Shearer 2015 and Section below on issues of credit and attribution). This is connected to the discoverability issue and again, the challenges that even an organisation with the scale of NIH faces in doing this well are worthy of note.

Australian Research Council and National Health and Medical Research Council (Australia)

Australia’s two main government funders, the ARC and NHMRC co-developed the *Australian Code for the Responsible Conduct of Research* (National Health and Medical Research Council, Australia et al. 2007), which provides the main policy framework for data sharing in Australia. This places limited requirements on researchers or institutions. Both have obligations to ensure the management and care of data but in terms of sharing have larger focus on ethical and ownership issues than data access. The single recommendation on data availability is that “[r]esearch data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters”.

Nonetheless as Shearer (2015) notes Australia is generally regarded as a leader in the data sharing space due to provision of infrastructure through the Australian National Data Service and good connections between institutions. The *Code* is in practice not directly monitored or enforced with expectations on detailed policy implementation deferred to institutions. Australia has some of the best discovery infrastructure for data at a national level through Research Data Australia and the National Library. Nonetheless it is still challenging to determine levels of uptake or compliance in what is a complex policy landscape.

## **Patterns in the policy landscape**

Within the set of funder policies there are two main patterns. There are consistent patterns in how policies and implementation evolve over time and there are patterns that relate to disciplinary differences. These patterns are logical and expected but they also are the root of a range of implementation challenges.

### **Policies develop over time**

Funder policies follow a quite consistent development pattern. They start as an articulation of aspirations or points of principle. These policy articulations are often the ambitious in aspiration but very limited in implementation details, often limited to statements that researchers “should” share data with limited specifications if any. These policies achieve relatively little uptake and are essentially never monitored. The second stage of development is to specify requirements in more detail, often focussed on data supporting

published reports and articles. Sometimes, but not in all cases, this includes a requirement for Data Management Plans at the point of grant submission, but often with limited clarity on how such plans are judged as part of the grant review process. In a few cases DMPs are required but data sharing is not.

A problem with this pathway of development is that the early stages are performative. To the extent that new requirements are imposed they are additional to, and generally separate from, the main flow of the research process. This means that monitoring is challenging, and recording and reporting minimal. Often DMPs are required but do not actually play any significant part in the review process. This can contribute to a researcher view of requirements as not connected to their core concerns and any requirements as a tick box administrative exercise to be completed.

This can create problems as policies move to the next state where requirements are imposed and some process of monitoring introduced. Firstly because in previous stages most researchers have not engaged deeply there has been little serious preparation to support data management and sharing, and little consideration of the changes in practice required. The frantic scrabbling of UK institutions in response to the EPSRC requirements is an example of this. On the funder side there has generally been little consideration of how to monitor progress, and the systems for monitoring are generally inadequate.

The final phase is one where monitoring and strict requirements are put in place, often with the threat of sanctions. There have not been explicit reports of sanctions being carried out to date. This is the stage at which it is generally realised that compliance monitoring is challenging or impossible across the community. Generally some form of audit process is suggested with limited random checks. A number of funders also report that at this stage the capacity and expertise of staff to monitor policy uptake and application is a challenge (Fig. 1).

Category	Aspiration	Recommend	Requirement	Compliance
Language				→
researchers...	can...should...might...	should...can...	must...will...	will...report...
practices are...	supported...examples are...	recommended...	acceptable....required...	minimal...standards...
funder will...	discuss...	review...	monitor...test...	audit...sanction...
Examples		FCT (Portugal), Genome Canada	BBSRC (UK), NIH (US), SSHRC (Canada), DFID (UK)	Wellcome (UK), Gates (US), EPSRC (UK)

Figure 1. [doi](#)

Illustration of the categories through which many research data management and sharing policies develop, with examples of the language used.

## Disciplinary variation

Strong policies are most common in areas of the biomedical sciences with a history of data sharing, most notably human genomics and structural biology, and areas of social sciences. These areas of strength are associated with the existence of long term data archives within the disciplines (ENA, NCBI, PDB, UKDA, ICPSR), often dating from the 1980s and 90s. Within the biomedical and structural sciences successful sharing is focussed very specific data types and data archives often specialise on a very small number of data types.

Funders in the biomedical sciences tend to have some of the strongest data sharing requirements and some of the most specific requirements in terms of appropriate data forms and archives. However in practice effective data sharing is largely focussed around those specific data types for which repositories already exist. Other forms of data are often seen as unimportant or too difficult to manage. For example, while there is extensive data supporting the claimed structures of biomolecules the data that describes the purification of samples prior to structure determination is rarely made available.

In the social sciences, where more heterogeneous data is common, a larger proportion of the core data is often available. However here effort has often been focussed on particularly data series which often have long term funding or at least grant support. An exception is the ESRC (UK) requirements associated with the UK Data Archive. All ESRC grant recipients are required to *offer* data collected for curation and management by UKDA. However the UKDA is not obliged to take it. The Social Sciences and Humanities Research Council (Canada) also has a long standing policy on data archiving and availability.

At the other end of the scale data sharing in the humanities is extremely limited. In part this is due to humanities scholars having the sense that they do not work with data as it is understood in the social and natural sciences. However digital data are increasingly being generated in humanities scholarship and more traditional approaches also generate notes and archives that could be seen as falling into a similar category. The International Digging into Data Challenges (<http://diggingintodata.org>) and the development of Digital Humanities more generally are an important signal of developing change. However the broader engagement of the humanities disciplines with what data sharing might be relevant or appropriate for them has been limited by comparison with other disciplines to date.

Interdisciplinary researchers and funders crossing a wide range of disciplines therefore face a substantial challenge in working across communities that have different histories, expectations, and infrastructure in place. Differing cultures have entirely different perspectives on what data is important, how it might be shared, and indeed what data even is. Therefore care is needed in developing policy to support these diverse groups and communities.

## Likely future directions

In the near term future there are three broad directions of travel that can be discerned. Many funders, particularly large government funders, will continue to develop down the path laid out above from aspiration to compliance tracking. These policies currently focus on DMPs at the individual grantee level. This development pathway runs the risk of driving a compliance culture in which researchers and their institutional support mechanisms adopt a minimal compliance strategy, delivering on the precise requirements but not engaging further with the agenda. This approach is most successful where the funder is supporting the provision of infrastructure, training, and education

The second approach, exemplified by the EPSRC also follows the same path but focusses compliance requirements on the institution, rather than the researcher. DMPs are often still required but in the EPSRC case hard requirements are placed on institutions. Although at this stage there is no global reporting regime, there will be audits. Failure to pass an audit could put an entire institution's EPSRC funding at risk and this is driving rapid changes within UK institutions. Again there are risks that this compliance culture is not optimising the behaviours and culture change that are desired by the funder. Nonetheless it is extremely effective and driving institutional change and motivating institutional research support to monitor researcher practice.

The final path is not well represented amongst current funder practice but is seen in a small number of research institutions. Here there is a shift away from the compliance focus to a partnership that drives adoption. This is often characterised by an effort to put infrastructure and support in place prior to policy announcements. This is most frequently successful where there are long standing data infrastructures and a disciplinary culture of data management and sharing.

A clear challenge that emerges from existing analyses as well as the interviews conducted for this review (below) is the balance between compliance monitoring and sanctions, and supporting a slower process of cultural change. Particularly where a funder is engaged with a diverse group of research communities there will be substantial differences between readiness and knowledge across those communities. These differences are often mirrored by capacity and knowledge within the funder itself.

Because policy tends to follow a pathway starting with aspiration and only gradually moving to concrete requirements the ability to monitor progress is often limited. While requiring DMPs is often seen as a universal and relatively low burden means of engaging researchers in data sharing, follow up on performance against the DMP or support on delivering it appears a high burden for both researchers and funders. Both the financial and political costs of putting systems in place for further tracking of data management and sharing mean these systems are not in place when they are needed. This in turn makes both support from the funder and ongoing engagement with the DMP difficult, focussing compliance tracking on very specific points of the project life cycle. This tends to exacerbate the risk of developing a compliance culture.

## A brief review of current guidance for policy development

While there is a substantial literature on data policies and extensive discussion of (and often objections to) their implementation there is a limited quantity of guidance on the development of policies, particularly for funders. Amongst these, the most important is a white paper *Current Best Practice for Research Data Management Policies* (Hodson and Molloy 2015) Fig. 2.

**Issues that RDM/RDS Policies Should Address**

The CODATA Paper on *Current Best Practices for Research Data Management Policies* (Molloy and Hodson, 2015) details the core issues and principles an RDM policy should address. It divides these into general principles and key policy elements.

**General Principles to be addressed**

1. An account of the general drivers and principles
2. A discussion of the requirements for effective data sharing
3. A statement of the necessary limits of openness

**Key Policy Elements to be addressed**

1. A definition of research data
2. An overview of the data within the scope of the policy: which generally includes two definitions:
  - a) The data that directly underpin or substantiate published research findings (i.e. those that are required for validation). Such data should be made available concurrently with the research publication.
  - b) The data assets that are created by the research project, but which may not directly underpin the published research findings.
3. An indication of general criteria for the selection of research data
4. A summary of responsibilities
5. An indication of the availability of infrastructure and responsibility for costs
6. An overview of data management planning requirements
7. Recommendations on enabling discovery and reuse
8. Stipulations to encourage recognition and reward for data providers
9. A summary of reporting requirements, compliance monitoring and any possible sanctions

Figure 2. [doi](#)

The core issues and principles that a Research Data Management policy should address, adapted from Hodson and Molloy 2015

The other main pieces of guidance are a report commissioned by SPARC, focussed on policy development for Health Research Funders that has a strong U.S. focus (Tananbaum 2014), European Commission guidance to member states on implementation of Open Access policies that touch on research data availability, and the *Comprehensive Brief on Research Data Management* (Shearer 2015) prepared for the Canadian Government and therefore with an emphasis on implementation of data policies in the Canadian context, focusing on the three federal funding agencies.

There are a range of other documents that develop policy positions and statements. These include the OECD Reports on Open Science and Open Data, the statement by the International Council of Scientific Unions, the Office of Science and Technology Policy statement on Open Access to research outputs, as well as statements by specific funders. However as these statements address motivations and requirements and do not provide a critical analysis of implementation issues specifically they are not covered in any detail here.

The Canadian Brief (Shearer 2015) provides a list of key implementation policy challenges that parallel those discussed in the CODATA paper (Hodson and Molloy 2015). Shearer lists disciplinary contexts, researcher preparedness, incentives, costs and the institutional role as the key challenges (section 7). The CODATA paper while organised differently dedicates a section to disciplinary differences, including the preparedness of researchers within disciplines, and covers recognition and reward, the provision of infrastructure and its costs, and the importance of clarifying the roles and responsibilities of the various stakeholders (including institutions) amongst the key issues that a policy should address. The SPARC Primer (Tananbaum) addresses many of the same issues in a specifically U.S. context although it takes a strong stance on specific issues that the other reports raise as potential challenges.

Other relevant documents that provide guidance are largely focussed on government and organisational data rather than research data. Examples include the Sunlight Foundation *Open Data Policy Guidelines* (<http://sunlightfoundation.com/opendataguidelines/>) and the United Nations *Guidelines on Open Government Data for Citizen Engagement* (United Nations 2013). Although these focus on large scale institutions making data available many of the same issues are raised in these documents including responsibilities, the provision of appropriate infrastructure and requirements for reporting.

A range of political critiques of Open Government Data have been made which might also be usefully reflected on in the context of research data sharing. Questions are raised as to the extent to which Open Government Data is made accessible to “ordinary citizens” versus those who already have access to the technical capacity and education which is often associated with (relatively) ready access to political influence. At its best data can democratise knowledge and empower citizens but it can also provide advantages to specific subgroups.

In the context of research data the usefulness and accessibility of shared data was raised most effectively in the Royal Society (U.K.) report *Science as an Open Enterprise* (Royal Society 2012) where the concept of “Intelligent Openness” was developed. The concept of Intelligent Openness for data is described as encompassing data that is accessible, intelligible, assessable and usable. While the Royal Society report focusses more on technical aspects of these characteristics and their application most specifically by other researchers, they might also be interrogated with respect to a wider spectrum of users, particularly in development contexts.

These issues are beyond the scope of the current review but they point to a gap in many of the documents on guidance and indeed wider literature on data sharing in general. While culture (of research communities, funders, wider publics) is often gestured towards as a challenge there is little serious engagement with how data accessibility and usability is affected in differing cultural contexts and little or no development of frameworks that could model how cultures might actually change. This is beyond the scope of the current report but an important area for future critical studies.

In summary, although there is limited documentary guidance on the design of policies strong themes emerge from the guidance that is available. Key elements of good policy design include stating the framework, clearly dividing up responsibilities, clear statements of scope, incentives for adoption and descriptions of monitoring and compliance arrangements.

One aspect of guidance documents and indeed the wider discussion on data sharing is that they rarely treat issues of culture seriously. Given that many of the challenges that are raised are cultural in nature and the aspirations for culture change that are a large part of the agenda there is a need for more direct engagement with community cultures as part of the broader implementation project.

## Findings

Developing data sharing policies is a growing concern for funders globally. Large funders in the United States and United Kingdom have had policies for some time with biomedical and social sciences funders in the United Kingdom having the longest standing policies. Most policies develop along two strands, requiring some form of Data Management Plan for grant submission, and expressing a requirement, or aspiration for data directly underlying published outputs to be made available. A small number of the most developed policies also place requirements on researchers to archive and store data generated for extended periods and/or to make this available in some form.

Policies generally develop from aspirational, to imposing requirements, to requiring reporting on the implementation of those requirements. No systems currently exist that allow funders to track uptake of policies and there are challenges in defining how levels of compliance could be calculated. The scope of policies as to what data outputs are included is often vague. The most developed policies are seen within large funders. Smaller funders are starting to follow with policy statements usually focussed on Data Management Plans for funding applications or with encouragement for data sharing found within broader Open Access policies.

In general policies are moving from the aspirational to imposing requirements. Amongst the interviewees there was a strong split between those who saw requirements as a necessary part of raising issues amongst researchers and those concerned that this was leading to a compliance culture where data management was seen as a required administrative exercise rather than an integral part of research practice. Discussion of this issue generally focussed on Data Management Plans and the tools supporting their production, but the implications were wider in terms of monitoring, policy scope and support.

Concern for the provision of support in general; awareness raising, education and training, infrastructure, and expertise was a consistent theme amongst interviewees. This was aligned with a sense of a need to move from requiring action of researchers to supporting cultural change towards data sharing in general. While the details varied, a concern with the provision of the right form of support at the right point in the research process was a

consistent theme and provides a means of integrating otherwise disparate views. The key common concern was the balance between imposing requirements and supporting organic change within communities.

### **Implications: The risk of compliance culture slowing cultural change**

The most striking finding in this study is the tension between the momentum in policy development towards requiring, and in some cases auditing, data management planning and data sharing, and a concern that these requirements are actually damaging to the process of cultural change amongst researchers towards data sharing as an element of research practice.

As noted, Data Management Plans, sat at the centre of this discussion, either as a successful example of interventions that encouraged thinking about data planning, or as an example of an exercise seen by researchers as unhelpful and purely administrative in nature. Bringing these two views together identifies an opportunity to improve Data Management Planning as a process and to place it at the centre of a collaboration between funder, researcher, publishers, data stewards and downstream users.

In an ideal world it would be feasible to develop systems in which Data Management Planning could be part of an initial proposal but would be further developed as a project was approved and commenced. Systems supporting its production would enable collaborative authoring and structured metadata on expected outputs could be captured as part of this process. In turn the document could serve as a developing manifest for data outputs as the project proceeds, as a checklist and basis for discussion between funder and research group, and as the source of information on data location and access arrangements to be passed to publishers when relevant articles are published. Finally it could form a record, manifest and index for data products when they are archived following the project completion.

With limited resources it will be crucial to develop tools and design the scope of policies and expectations on researchers so as to best align researcher and funder motivations. Simply requiring data management planning at proposal stage without providing support will likely lead to the production of documents that are at best ignored. Providing ongoing support will be resource intensive. Identifying the best point at which to apply available resources, and the scope of data the funder is most concerned with, will therefore be important. Where do the concerns of the funder for telling success stories and achieving impact align with the needs of researchers to see their research taken up by other colleagues? How can data products of funded research be promoted effectively so as to maximise use that leads both to wider impact and academic citations, and how can examples of such use be captured?

Funders are of course highly diverse as are the communities they support. Answering these questions is therefore likely to be dependent on the funder and the research communities they fund. At the same time the tools and systems being used to support data sharing and planning are more general or at least are likely to require resourcing by a wide

range of funders. The tension between generally applicable and locally useful is not limited to data of course. The key questions for a specific funder will be how to match the aspiration for data sharing, and the desire for greater impact, with the available resources and expertise.

### **Preliminary recommendations for funders**

The challenge for funders in developing policy is balancing aspiration, resources, external political pressures and the ability of the research community to act. Hodson and Molloy (2015) provide an excellent template for identifying the key elements a policy on research data management and sharing should address. This should form the basis for policy design efforts.

Around any policy development a funder should consider its overall program for supporting Research Data Management and Research Data Sharing. We recommend framing such a program as one of supporting cultural changes. Crucially these changes need to include both the cultures of the disparate research communities the funder supports *and the funder themselves*. Providing clarity on the funder's motivations and how these align, and do not, with those of the community will be valuable. Understanding the capacities, capabilities and interests of those within the funder, and how this might need to change is also important.

While this is a complex process, and arguably one for which there is, as yet little support the following concrete actions can be recommended as a core of activities in any funder program to support RDM and RDS.

- Clearly define the aspirations, goals, and motivations of the funder in developing an RDM/RDS program for funded research.
- Develop an in depth understanding of the current state of the funded research communities including capacity to support RDM/RDS, readiness to engage, interest in the issues and motivations, and critically the degree of heterogeneity across those communities.
- Audit and consider the capacity of funder staff in terms of their expertise and existing workload to support the RDM/RDS program in general and their ability to monitor and engage on a continuous basis with funded research projects on these issues.
- Identify and consider externally available tools and resources, that might enhance internal capacity or cover gaps in capacity. Seek to partner with other like-minded organisations to provide support where possible.
- Develop policy statements so as to align with the funder's capacity to monitor and support. Ensure that the scope of policy requirements with respect to data types and classes matches funder capacity to monitor and support.

- Develop policy implementation processes in collaboration with research communities and identify champions (and friendly critics) to enhance prospect of cultural change within communities.

### **General recommendations for further investigation and development**

This review has highlighted issues that arise when funder policy meets community cultures at the point of implementation. The biggest issue that arises is the way that “culture” is used as broad brush term to describe aspects of community interest and readiness, capacity, as well as problems in implementation. “Culture” acts here as a word that places responsibilities elsewhere or otherwise externalises issues. The lack of direct engagement with culture itself and how it might change, indeed the lack of any reference to frameworks within which culture and culture change might be interrogated may be directly contributing to some of the difficulties facing RDM/RDS implementation.

This particularly emerges in the disagreement over what it is that Data Management Planning and requirements for Data Management Plans are *doing*. Are they an important part of introducing these concepts to a community (and therefore to its culture) or are they actually slowing the desired cultural changes within research communities by lumping in RDM/RDS with other purely administrative activities that are imposed on, and therefore irrelevant to, the story those communities tell themselves about what is important? Without a framework for understanding where communities are, their internal cultures of sharing (or not), and an understanding of how interventions effect cultural change (or do not) this disagreement can not be readily resolved.

With such an understanding we might be able, as a community, to design the tools and support structures so as to best support change. Data Management Plans are likely to sit at the heart of any such process as a common, if not universal, intervention that provokes a range of responses. Identifying how data management planning and plans might be adapted so as to become more fully integrated into research practice, and therefore part of research cultures would be valuable and provide a test bed for probing cultures more directly. As part of examining the interaction of DMPs with different communities it will be valuable to develop tools that help to understand how the underlying culture of those communities affect their attitude to data sharing. This might be presented as a typology of cultures or a readiness assessment for communities that would identify key underlying factors.

In summary there are three specific general recommendations for further development by the community.

- Develop tools and frameworks that support an assessment of research communities for their capacity, readiness and interest in data management and data sharing.

- Identify how DMP tools might be adapted to become more fully integrated into research practice and how they might address the different cultural issues that a readiness audit might raise for specific communities.
- Focus attention on the challenges of encouraging improved practice in data management and sharing as issues of cultures and how they interact and change. Use the lens of culture to examine how research communities that support effective data sharing differ from those that do not and develop frameworks that utilise and support information that technical and policy interventions can provide.

## Grant title

Exploring the opportunities and challenges of implementing open research strategies within development institutions (Neylon and Chan 2016).

## References

- Borgman C (2012) The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63 (6): 1059-1078. <https://doi.org/10.1002/asi.22634>
- DFID (2013) DFID Research Open and Enhanced Access Policy v1.1. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf). Accessed on: 2016-1-18.
- Digital Curation Centre (2012) Overview of funders' data policies. <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>. Accessed on: 2015-8-16.
- European Commission (2015a) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf). Accessed on: 2016-1-28.
- European Commission (2015b) Guidelines on Data Management in Horizon 2020 Version 2.0. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). Accessed on: 2016-1-28.
- Fecher B, Friesike S (2013) Open Science: One Term, Five Schools of Thought. SSRN 2013: 2272036. URL: <http://papers.ssrn.com/abstract=2272036>
- Hodson S, Molloy L (2015) Current Best Practice for Research Data Management Policies. Zenodo <https://doi.org/10.5281/zenodo.27872>
- Leonelli S (2013) Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology. *Bulletin of Science, Technology & Society* 33: 6-11. <https://doi.org/10.1177/0270467613496768>
- National Health and Medical Research Council, Australia, Australian Research Council, Universities Australia (2007) Australian Code for the Responsible Conduct of Research. <http://www.nhmrc.gov.au/files/nhmrc/file/research/research-integrity/>

[r39\\_australian\\_code\\_responsible\\_conduct\\_research\\_150811.pdf](#). Accessed on: 2016-1-31.

- National Institutes of Health, USA (2015) NIH Data Sharing Policies. [https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_policies.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html). Accessed on: 2016-4-01.
- Neylon C, Chan L (2016) Exploring the opportunities and challenges of implementing open research strategies within development institutions. *Research Ideas and Outcomes* 2: e8880. <https://doi.org/10.3897/rio.2.e8880>
- OECD (2007) Open Science. <http://www.oecd.org/sti/outlook/e-outlook/stipolicyprofiles/interactionsforinnovation/openscience.htm>. Accessed on: 2015-9-17.
- OpenAccess\_EC (2016) Open Access EC, Twitter Status. <https://twitter.com/OpenAccessEC/status/692696388174528512>. Accessed on: 2016-1-28.
- Royal Society (2012) Science as an Open Enterprise. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>. Accessed on: 2016-1-27.
- Shearer K (2015) Comprehensive Brief on Research Data Management Policies. <http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1>. Accessed on: 2016-4-01.
- Spichtinger D (2015) Open Access in a European Policy Context. <http://de.slideshare.net/RightToResearch/daniel-spichtinger-open-access-in-a-european-policy-context-opencon>. Accessed on: 2016-1-28.
- Tananbaum G (2014) Data Policy: A primer for research funders. <http://sparcopen.org/wp-content/uploads/2016/01/sparc-open-data-primer-final.pdf>. Accessed on: 2016-4-16.
- UK Open Data Research Forum (2015) Concordat on Open Research Data. <http://www.rcuk.ac.uk/research/opendata/>. Accessed on: 2016-1-29.
- United Nations (2013) Guidelines on Open Government Data for Citizen Engagement. <http://workspace.unpan.org/sites/Internet/Documents/Guidelines%20on%20OGDCE%20May17%202013.pdf>. Accessed on: 2016-4-18.
- White House Office of Science and Technology Policy (2013) Increasing Access to the Results of Federally Funded Scientific Research. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf). Accessed on: 2016-1-29.
- Whitmire A, Briney K, Nurnberger A, Henderson M, Atwood T, Janz M, Kozlowski W, Lake S, Vandegriff M, Zilinski L (2015) A table summarizing the Federal public access policies resulting from the US Office of Science and Technology Policy memorandum of February 2013. *Figshare* 2015: 0. <https://doi.org/10.6084/M9.FIGSHARE.1372041>
- World Bank (2012) World Bank Open Access Policy for Formal Publications (English). <http://documents.worldbank.org/curated/en/2012/04/16200740/world-bank-open-access-policy-formal-publications>. Accessed on: 2016-1-28.