

Building a Culture of Data Sharing: Policy Design and Implementation for Research Data Management in Development Research

Cameron Neylon [‡]

[‡] Curtin University, Perth, Australia

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

Reviewable v1

Received: 20 Oct 2017 | Published: 24 Oct 2017

Citation: Neylon C (2017) Building a Culture of Data Sharing: Policy Design and Implementation for Research Data Management in Development Research. Research Ideas and Outcomes 3: e21773.

<https://doi.org/10.3897/rio.3.e21773>

Abstract

A pilot project worked with seven existing projects funded by the International Development Research Center of Canada (IDRC) to investigate the implementation of data management and sharing requirements within development research projects. The seven projects, which were selected to achieve a diversity of project types, locations, host institutions and subject areas, demonstrated a broad range of existing capacities to work with data and access to technical expertise and infrastructures. The pilot project provided an introduction to data management and sharing concepts, helped projects develop a Data Management Plan, and then observed the implementation of that plan.

In examining the uptake of Data Management and Sharing practice amongst these seven groups the project came to question the underlying goals of funders in introducing data management and sharing requirements. It was established that the ultimate goal was a change in culture amongst grantees. The project therefore looked for evidence of how funder interventions might promote or hinder such cultural change.

The project had two core findings. First that the shift from an aim of changing behaviour, to changing culture, has both subtle and profound implications for policy design and implementation. A particular finding is that the single point of contact that many data

management and sharing policies create where a Data Management Plan is required at grant submission but then not further utilised is at best neutral and likely counter productive in supporting change in researcher culture.

As expected, there are significant bottlenecks within research institutions and for grantees in effectively sharing data including a lack of resources and expertise. However, a core finding is that many of the bottlenecks for change relate to structural issues at the funder level. Specifically, the expectation that policy initiatives are implemented, monitored, and evaluated by Program Officers who are the main point of contact for projects. The single most productive act to enhance policy implementation may be to empower and support Program Officers. This could be achieved through training and support of individual POs, through the creation of a group of internal experts who can support others, or via provision of external support, for instance by expanding the services provided by the pilot project into an ongoing support mechanism for both internal staff and grantees.

Other significant findings include: the importance of language barriers and the way in which assumptions of English language in materials, resources, services and systems permeate the entire system; that data infrastructures are poorly served by current funding arrangements and tools, particularly where they are obliged to seek continuing funding through project grants. There are also fundamental questions raised by the status of digital objects as "data". The concept of data is part of a western scientific discourse which may be both incompatible with other cultures, particularly indigenous knowledge systems. More importantly that discourse may be incompatible with values-based approaches that seek to respect indigenous knowledge through a commitment to retaining context.

With the possible exception of the last finding, none of these issues are exclusive to development research. The Development Research context surfaces them more strongly through its greater diversity of goals and contexts. In many ways this project illustrates not that Development Research has particular special needs, but that it is a site that surfaces issues in policy design and implementation deserving of more consideration across the research enterprise.

Keywords

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

Introduction

Goals of IDRC in developing a data sharing policy

IDRC along with many research funders is examining the opportunities for encouraging data sharing and improved data management. Data sharing, open data, and data

management policies are being implemented by many funders with a range of variation in their approaches (see Neylon (2017h), review for this project). Underpinning these differing policies are a range of motivations. As noted in the review these can be mapped to a range of discourses associated with Open Science (Neylon 2017h, Fecher and Friesike 2014) and there is often a mismatch between the explicit motivations, unstated or implicit motivations, and the design of the actual policy.

IDRC in developing a draft Open Data Policy in 2016 made the following statements on motivations:

"IDRC recognizes that the dissemination of research data can accelerate collaboration, scientific discovery, and even follow-up of research efforts. Open access to research data can be particularly important to researchers in developing countries as they may face additional institutional and financial barriers to access and archive data. IDRC is also committed to good stewardship of public funds and innovative use of knowledge for development."

Draft IDRC Open Data Policy and Guidelines FAQ - version of early 2016 (internal document)

As was noted previously (Neylon 2017h) this aligns with the "pragmatic" and "democratic" discourses identified by Fecher and Friesike (2014). The emphasis is on enhanced efficiency of research and "stewardship of public funds" alongside "access to research data... [for] researchers in developing countries".

In addition to these motivations the IDRC Strategic Plan for 2015-2020 (IDRC 2014) also engages with the Democratic and Public discourses identified by Fecher and Friesike. The Strategic Plan also has a strong focus on capacity building. This would appear as an implicit motivation for developing data sharing as a requirement for grantees. With both research and economic development increasingly dominated by discussions of "big data" globally improving the capacity to manage and mobilise data in developing and transitional countries, and indeed improve the capacity of IDRC staff to support this growth in capabilities is well aligned with the IDRC mission.

Overall the goals of IDRC in developing policy and guidance on research data management and data sharing are focussed firstly in achieving the greatest impact for funded research, supporting further research, and researchers, particularly in developing and transitional countries, and building capacity, both within IDRC and amongst grantees.

An implicit Theory of Change and the design of the Pilot

In its initial form, the program on data management and sharing was framed within IDRC as a policy question: what is the optimal form of a policy to encourage data management and sharing amongst grantees? In common with many other funders the initial draft policy is aspirational in form, while focussing on the provision of a Data Management Plan at the point of grant submission as the main point of leverage to change practice.

The implicit initial Theory of Change was therefore that by finding the "correct" policy design, alongside the provision of some support, the practice of grantees would be changed towards adoption of data sharing. This implicit Theory of Change is common amongst funders who develop and implement policies. The political realities of policy implementation then lead to an initial development of aspirational policies that are, in ideal circumstances, strengthened over time, as was noted in the initial review.

The IDRC approach was different, however in taking a research project based approach to testing and designing a policy and implementation. The current project was designed as a pilot, in which the draft policy would be implemented within a small number of active IDRC-funded projects to examine the issues that policy implementation might create. Nonetheless the first specific objective of the project (Neylon and Chan 2016) was to "Test and refine implementation guidelines for development researcher funders' open research data policies".

A shift from targeting policy-mediated change to a focus on culture change

As part of the initial review (Neylon 2017h) for the project a concern with this conventional approach to encouraging data sharing was raised. The review entitled "Compliance Culture or Culture Change?" noted that the focus on the generation of Data Management Plans at grant submission time was the subject of strongly differing opinions amongst interviewed experts. In particular the use of an administrative requirement was viewed by some as leading to data management and sharing being viewed as just one "tick-box" amongst many requirements imposed by funders.

The question was raised whether such approaches risked creating a "compliance culture" where data management was viewed as merely an administrative requirement could actually damage the goal of supporting a culture where sharing and management of data were part of standard practice. At the same time the interviews identified a strong view amongst many experts that it was necessary to push researchers to think about data management and sharing, and that without a formal requirement at some point in the grant cycle this would not happen. While strongly supporting the idea that data management and sharing should be considered throughout the project lifecycle there was agreement that grant submission remains the most obvious place to insert a compliance requirement.

Alongside this the review also noted critical issues of capacity. Capacity issues amongst researchers, particularly in developing countries were expected. What was less expected was the importance of capacity and time availability within funders. In particular it was noted that there is a systemic structural problem with policy implementation for most funders. The majority of funders organise their programs with Program Officers having responsibility for management of, communication with, and in some cases soliciting proposals for, grants. With communication flowing through a single person, any policy change or implementation effort requires each Program Officer to gain sufficient expertise in the new area to offer each of their grantees support and advice. Frequently responsibility for tracking policy adoption and compliance also rests with individual Program Officers.

The need to examine and improve capacity at the funder level, alongside addressing capacity building amongst grantees aligns well with an agenda for culture change. Capacity building on both sides for collaboration offers opportunities to develop a common narrative, and therefore a common route to culture change. In response to this we adapted the case study design of the pilot to examine experience and capacity amongst grantees and to identify what forms of support are necessary on the funder site to address these. In addition we frame the issues faced in implementing and adopting data management and sharing practice in terms of the culture in which the research teams, and program officers, work.

Pilot design and analysis

The Pilot was designed around seven case studies of existing IDRC funded projects (Neylon 2017a, Neylon 2017b, Neylon 2017c, Neylon 2017d, Neylon 2017e, Neylon 2017f, Neylon 2017g). These projects were selected to span a range of geographies, scales and subjects. The selection deliberately included social, economic, and bioscience and health related data, as well as projects relating to Indigenous Knowledge and one project involving a francophone team.

The basic model of the pilot was to work with each project to develop an inventory of the data they were creating (or had created) followed by the development of a Data Management Plan. At the conclusion of the Pilot the progress of each project in delivering on the DMP would be examined. This was supported through two workshops, one at the beginning of the project and one near the end. Through this process we sought to address the following questions for each project:

1. What was their awareness of and thinking around "data" as a concept? What data did they expect the project to generate? How complete was this view?
2. What challenges did the project face in developing a Data Management Plan? How did the nature of the project and its outputs affect this.
3. Are existing tools and systems for data management planning fit for purpose for:
 1. Researchers in developing and transitional countries
 2. Researchers working on development research projects
4. What challenges did the project face in implementing their Data Management and Data Sharing Plans?
5. Did the process of participating in the Pilot lead to changes in practice or changes in culture that could be documented?

Methodology

Project selection

Projects to participate in the Pilot were solicited through contact with IDRC Program Officers (Table 1). A submission form was provided and either Program Officers or project investigators filled out the form. Ten projects were submitted, including two which were sub-

projects of the [Open and Collaborative Science in Development Network](#) (Chan et al. 2015). Of the ten, eight were selected, one of which did not ultimately proceed as the host institution was unable to satisfy IDRC requirements for signing up.

Name and abbreviation	Host	Country	Discipline	Notes	Case Study
Crowd Sourcing Data to fight Social Crimes (HMP)	Harassmap (NGO)	Egypt	Social Sciences	Small NGO with focus on web presence and advocacy	Neylon 2017e
The Brazilian Virtual Herbarium (BVH)	CRIA (National Infrastructure)	Brazil	Biosciences and Ecology	Substantial infrastructure drawing on upstream data sources	Neylon 2017d
Strengthening the Economic Committee of the National Assembly in Vietnam (ECV)	Centre for Analysis & Forecasting (CAF) (research institution)	Vietnam	Economics	Government department gathering data for direct policy support	Neylon 2017b
The Impact of Copyright User Rights (DED)	Derechos Digitales (NGO)	Columbia	Law	Distributed project	Neylon 2017a
Establishing a clearinghouse for tobacco economic data in Africa (TED)	DataFirst (project within UCT)	South Africa	Economics and Health	Infrastructure and issues with upstream usage rights	Neylon 2017f
Les problemes negliges des systemes de sante en Afrique : une incitation aux reformes (NDF)	LASDEL (Research Unit)	Niger	Health Policy	Francophone project and project team	Neylon 2017c
Indigenous Knowledge in Climate Change (IKC)	Natural Justice (NGO)	South Africa	Law and Culture	Focus on Indigenous Knowledge	Neylon 2017g

The selection of projects was based on seeking broad geographic representation, a variety of subject areas for the research and variation in the nature of the projects and project teams. The final set of seven projects included two from Latin America (one Brazil, one based in Columbia), four from the African continent (two from South Africa, one from Niger, and one from Egypt) and one from Vietnam. The projects included one francophone project and one dealing with Indigenous Knowledge. The institutional hosts for the projects varied from NGOs to Research Institutes, Universities and Government Departments.

The Data Management Plan for the project as a whole is published (Neylon 2017j) as is a Data Package containing all publicly released data and materials (Neylon 2017i).

Initial workshop

An initial workshop over two days brought together project leads, program officers, as well as expert advisors to initiate the project. Each project was presented and these presentations are available in the project data package (Neylon 2017i, Introductory Workshop Presentations directory). The main focus of the workshop was to initiate thinking for the projects on what they would consider as data, what challenges they would face in managing and sharing that data, and to introduce the tools for data management planning, specifically the Portage developed [DMP Assistant](#), a multi-lingual version of the UK Digital Curation Centre [DMP Online tool](#).

The workshop materials and schedule are available in the project data package (Neylon 2017i, Introductory Workshop Materials directory). Materials were produced in English and translated to French. Specific prompts were provided for advisors and experts to help prompt and guide discussion. The activities developed a discussion of what could be considered data, with the aim of showing that the scope can be expansive. Following this an initial Data Inventory was developed (see Data Management Planning > Data Inventories in the data package).

The draft Data Inventory was then used as a means of focussing discussion on the different types of data being generated by each project, and the challenges involved in managing and sharing that data. The goal of discussion was to have projects communicate with, and critique each other's draft Inventory. The intent was to develop an appreciation of how data resources were different across the projects and how attitudes across the projects and the issues associated with data sharing were challenging.

Throughout the first workshop there were challenges in translation. As only one group was operating in French and there were no other French speakers amongst the contributing project participants attending this meant that conversation between the francophone and non-French speaking participants was limited. While a number of bilingual participants were at the meeting, amongst advisors and program officers to follow the program, the degree of interaction was more limited. Simultaneous translation was valuable in addressing this, although expensive. If this workshop design were to be used in future ensuring that any language for working in was represented by at least two projects would be important.

Data inventory

The data inventory was based on a simple form that prompted the participants to identify data resources and to identify their formats, approximate size, and potential issues with sharing. A draft version was prepared at the first workshop and this was refined and finalised following the workshop.

The aim was to focus attention on the process of surfacing the full set of objects generated by projects that might be considered data. The process starts from each phase or Work Package in a project and then offers a set of possible data products that might be

generated from each phase as prompts. These products are then named and described, and then analysed in terms of their format and issues that arise from management or sharing of these data.

The form was provided in physical form although several participants subsequently requested a digital version. The instrumental and guided approach to surfacing data objects was seen as valuable and helped to make the process of considering sharing and management more concrete. The blank form and filled out version for most projects is available in the project data package (Neylon 2017i see Data Management Planning Directory > Data Inventories).

Data management planning

The DMP Assistant tool provided by Portage was recommended as the basis for each participating project to develop a Data Management Plan. The tool was introduced at the first workshop and used in default template mode. For future efforts using DMP Assistant it would be advisable to develop or adopt a simplified template designed for IDRC or the specific user group.

Participating projects mostly used the online tool. The exceptions were one case where a login problem led the user to use a different online tool ([DCC DMP Online](#)) and where concerns over network access led to the use of a blank plan downloaded as a Microsoft Word document (NDF). The participants were prompted several times to complete the DMPs and plans were developed with differing degrees of specificity (see data package, Data Management Planning > Data Management Plans) and in some cases a finalized version was formally published (Canhos 2017, Traynor 2017, Wael 2017, Woolfrey 2017).

Follow-up, interviews and analysis of data sharing and management

Pilot Project participants were interviewed after the DMPs were supposed to be due. The interviews focussed on a reflection on the process of Data Inventory and Data Management Plan production, specifically what parts of each process were useful, what not so, and how this had changed practice within the research project or group. The interview rubric and notes for each interview are available (see data package Data Management Planning > Data Management Interviews). In several cases the DMPs had not been completed at this stage and this limited the ability to focus on concrete aspects of the DMP process.

Self-reported performance against the DMP was collected at a final workshop and where feasible the actual sites for data sharing were examined. In most cases the projects self-reported that actual data posting was still in process so no final audit of performance against the DMP was possible. Ideally this could be assessed after a further six and twelve months to identify progress. The forms used to prompt the self-audit are available (see data package Data Management Planning > Final Self Audit > Templates for Self Audit) as are the filled out forms (Data Management Planning > Final Self Audit).

Case study analysis

The individual case studies are published separately (Neylon 2017a, Neylon 2017b, Neylon 2017c, Neylon 2017d, Neylon 2017e, Neylon 2017f, Neylon 2017g), and the underlying data is available in the project data package on Zenodo (Neylon 2017i).

Awareness and pre-existing capacity for managing and examining data

The project showed substantial variation in previous thinking about data management and capacity to manage and share data. The projects ranged from sites who were interested but had no experience in research data management, to professional data managers and infrastructure providers with decades of experience.

With respect to the less experienced groups, while many of the participating projects had experience in creating, collecting and managing digital objects within the majority there was little previous experience of their management and handling. In the initial exercises at the first workshop and within the Data Inventory it was clear that diffuse but limited concepts of data became rapidly expanded. This led to a concern about the scope of data management, and in most cases to a realisation that management had to be focussed on a subset of objects. This appeared to be a very productive discussion for thinking pragmatically about what was valuable and achievable.

The more experienced groups (TED and BVH) had preexisting frameworks and ways of thinking about data that were different to each other and not always a good fit with the general framing used in the workshop. This was in part due to their nature as infrastructures more than projects but also due to their longer history working with data. Nonetheless the change in perspective still was seen as having some value and the experienced contributors were able to probe and support other groups, as well as provide examples for their experience, in the workshop setting.

Across the projects there was also a substantial diversity of data types and collection modes. Defining any general guidance for best practice in the research practice around recording data would have been extremely challenging. Beyond good practice for local archiving and backup there was little in common in terms of potential training needs that could be provided centrally. The initial workshop, and particularly the first day where questions of what data are and the different issues that arise with sharing and management, was well received. Having a shared framing was useful for all participants.

The development of data inventories

The development of the data inventories within the initial workshop proceeded quite well and in most cases a completed inventory was produced on schedule. The inventories focussed on a step by step articulation of the parts of the project, the data products to be produced, their expected format, size and then issues relating to management and sharing.

The Data Inventories indicated that data formats were largely chosen *ad hoc* based on the software or systems being used. The importance of using open and accessible formats for archiving was emphasised but in practice there was little evidence of uptake on this. This is consistent with the common finding that there is limited motivation to undertake extra work - in this case changes in format - to enable data sharing and management.

The potential issues with data sharing described in the inventories were varied. Most focussed on ethical issues, and questions of control and permissions. Some technical issues were raised but at a surface level in most cases. The more experienced groups unsurprisingly identified more data objects and a wider range of technological issues.

The inventories were seen as a useful exercise, in part because they were highly concrete, and were prepared initially in a supported context, as part of the flow of the workshop. In the context of the first workshop the intended process of reflection and refinement was less successful and it would have been more productive had the inventories been completely finalised in the context of that workshop. Overall this was a useful exercise and there was evidence within the Data Management Plans and in eventual practice that the issues raised had continued to be considered.

Data management planning

Data Management Planning using the DMP Assistant Tool was carried out by all the groups, with differing levels of support and success. Variation in the level of details in Data Management Plans was substantial and the process was generally seen as less connected to the project than the Data Inventories. A common criticism was that the Inventory was not easily incorporated into the planning workflow or into the tool itself.

The DMP Assistant tool is a form-based online tool. The project used the default template which provided generic questions. For a future project the preparation of a template that maps onto existing funder processes and language as well as connects more directly to the preparation and support that the Inventory provided would be valuable. Technical issues are discussed in more detail below (Tools and systems > Technical issues).

In two cases (NDF, IKC) the submitted DMP was revised at a later date in the project (see data package Neylon 2017i and published DMP for IKC project Traynor 2017). In general, while there was a shared sense that the process of planning was useful, the DMP itself did not seem to be of significant further use to the contributing projects. It was useful to the Pilot Project as a whole as a means of auditing performance against the plan but in terms of supporting culture change and deeper evaluation the DMP itself was of limited value after being prepared. This is broadly consistent with the findings from the initial review and aligns with the fears expressed by some expert interviewees for that review.

Most contributing projects regarded the process as useful and shared the view that earlier planning would have been valuable. In a few cases (HMP, DED, ECV) the view was expressed that DMP preparation at grant submission stage would not have been of use because the forms of data resources to be produced were not yet clear. In the case of the

Indigenous Knowledge project there were complex interactions between obligations to multiple institutional review boards for ethical oversight (Traynor et al. 2015, Traynor 2017, Neylon 2017g) and it is not clear when in this process Data Management Planning would have been most useful and efficient.

Tools and systems: Experience of use in a development context

There were broadly three classes of issues relating tools and systems utilised by the project. These were technical issues, language issues, and more general work flow and intelligibility issues.

Technical issues

The technical issues ranged in severity from a temporary glitch in login authentication through to the online system being practically inaccessible due to unreliable network access in Niger. Several contributing projects preferred to use a downloaded version of the DMP questions as a Microsoft Word document than to use the system online for similar reasons.

Unsurprisingly, in developing and transitional nations network access can be limited, but more importantly unreliable. Web-based systems that rely on continuous internet connectivity are particularly poor under these circumstances as it can be unclear what has been saved back to the server and loss of connectivity can mean the loss of work. Continuously ensuring that the system has saved data is irritating and distracting. Of the contributing projects four (BVH, ECV, DED, NDF) preferred to complete the DMP offline. Where possible developers of web-based systems should consider their use in an off-line mode and local caching of data.

Perhaps more surprisingly there was limited evidence of any real benefit from using an online system. With the exception of one project (HMP) there was no effort to utilise the collaborative features of the DMP Assistant Tool and no evidence of sharing the content of the DMP through the system. Two projects offered access to their developing DMP to the Pilot Project (IKC, HMP). However all those who used the system ultimately submitted the DMP as a PDF. This reinforces the notion that the DMP is viewed as a document, to be prepared in a finalised form, rather than a tool for ongoing use.

Language issues

The Pilot Project made the deliberate choice to include at least one non-English speaking project amongst the contributing projects. The choice of using DMP Assistant was made in large part because it had included multilingual support, specifically at this point for French. This was helpful, however it also emphasised the challenges arising for non-Anglophone researchers in a highly English-centric world.

For the francophone project (NDF) substantial support was provided throughout by French speaking advisors and program officers. The core resources provided by the project were translated but the loss of efficiency in communication and challenges in translation were

evident throughout. For the other contributing projects where English was not the preferred language there were challenges in some cases in communication and comprehensibility, particularly with respect to the DMP template questions.

Language issues also appeared in the utilisation of data platforms. Specifically the NDF group selected the DataFirst platform to archive and make available data. However the metadata requirements for DataFirst include a requirement for English. This both creates the immediate issue of the resources required for translations but also a deeper issue of whether the translation of the locally contextualised metadata schema is appropriate. Is there a risk of loss of context, particularly given the intent to archive the collected data itself. Is it appropriate to enforce metadata standards designed in English, perhaps by people with no experience of other languages, on researchers operating in the vernacular in their local context?

The problems with issues of language are well rehearsed, but they are pervasive and reach deep into every area of communication, data management, and data sharing practice. The Pilot Project was not well placed to develop strong findings on the downstream effects of language on data reuse and discovery although at least one project (HMP) noted challenges arising from data discovery due to having data - in their case reports - in two languages, English and Arabic (Neylon 2017e).

Workflow and comprehension

The default assumption for Data Management Planning requirements is that funders will make these a required part of grant submission process. As noted in the initial review, there is both support for this and disagreement amongst experts (Neylon 2017h). Supporters note that this is one of the few points in the grant cycle where a requirement can be made absolute, with minimal monitoring requirements. It is therefore the best point to ensure that grantees consider Data Management Planning, and a robust means of signposting the importance of the issue.

Dissenters to this view note that detailed planning is difficult so early in the life cycle of a project and that making a DMP a documentary requirement tends to mean that the process of planning is not highlighted, rather it is seen as an additional administrative requirement for grant submission. In the initial review (Neylon 2017h) it was proposed that the means of addressing both the opportunities and concerns was to require some planning at submission stage, but to develop tools and systems by which the planning process would become a collaborative and ongoing exercise that engaged both program officers and grantees.

In practice Data Management Planning was a drawn out and in most cases arduous task for the contributing projects. The level of detail prompted by the template questions was confusing for many of the contributing projects. For a funder seeking to increase capacity, and therefore presumably the diversity of those proposing projects, such a technical requirement may impose a substantial barrier for less experienced project proposers.

Ideally a very lightweight requirement at project submission stage, perhaps similar in spirit to the Data Inventories, could provide a useful start. Full scale Data Management Planning required a level of support that would be difficult to provide across a full grant portfolio, and the quality of the process would therefore be likely to suffer. As noted above the lack of obvious flow and reuse of content from the Data Inventory (seen as a useful exercise and document) to the Data Management Plans was a concern across the groups. The design of future DMP systems should consider how best to guide users through the process of planning and recording at the various stages of the project lifecycle.

More specifically almost all contributing projects reported some confusion around the specific questions and guidance provided by the templates. There was some evidence that some of this may have been due to language difficulties. However in most cases there were also issues raised by clarity of the questions themselves. In contrast to the very concrete nature of the Inventory, focussed on specific data objects and their characteristics, the questions in the DMP template appeared very abstract. Unsurprisingly the contributing project with the most familiarity with Data Management Planning (TED) was comfortable both with the questions, and reported that the flexibility gained by their abstract nature was an advantage.

Challenges of implementation and data sharing

Going into the project there were a series of expected challenges that would be faced by the contributing groups including local expertise and capacity, access to digital networks and technical platforms, ethical issues of sharing, as well as issues of language. All of these were observed in the Pilot Project as contributing to the difficulties of managing and sharing data. The severity of these issues varied, in some cases being deeper and more challenging than expected, and in others being less significant.

As noted in the introduction the original Pilot Project design had focussed on the issues amongst grantees. However it was seen early on that similar issues at the level of the funder were also a substantial issue for implementing Data Management Planning and Sharing. Those issues are addressed more completely below (see Challenges at the funder level). In this section the focus is on challenges faced by the contributing projects.

Alongside the expected issues there were two issues that arose that had a qualitatively greater degree of importance than was expected; language and the challenges of infrastructure. Alongside this an unanticipated issue was raised by the Indigenous Knowledge contributing project (IKC), which prompted an epistemological question of how data can or should be dealt with within projects that are constructed around a value system at odds with common assumptions of what data is.

Local technical capacities and expertise

As was expected a lack of local technical knowledge and previous experience of Data Management was a substantial barrier in several cases. Three groups (HMP, DED, ECV) sought to build a local data infrastructure for sharing (Neylon 2017e, Neylon 2017a, Neylon

2017b, Wael 2017). To achieve this required bringing in external support and the internal expertise to decide between options was not consistently available. In each of these cases a question remains whether local provision is the best choice in terms of both long term sustainability and ensuring security. In the case of the Vietnam project long term government support may be expected but for both Derechos Digitales and HarassMap long term sustainability will depend on ongoing funding and interest.

For the LASDEL group (NDF) an external infrastructure was sought for data availability (Neylon 2017c). In contrast to other groups it was clear that the lack of network reliability in Niger would have made local provision of data sharing infrastructure unviable. Challenges were experienced in part due to a lack of knowledge of options and the criteria to distinguish between them. The decision to use DataFirst as the platform was driven in part by participation in the Pilot Project. The process for data deposition which is still ongoing required substantial support from a project advisor to move forward.

Both contributing projects that provide data infrastructures (BVH, TED) were well supported by a surrounding technical capacity, network infrastructure, and local expertise (Neylon 2017d, Neylon 2017f). Overall, all the projects were able to access expertise and technical capacity to deliver a wide range of options for data availability. Network access, and probably more importantly reliability, remains an issue in some areas (NDF), although this is improving. In terms of capacity building difficult decisions need to be made on supporting regional and global provision vs local capacity. There are good data resources providing access to data of global importance in most of the regions of interest, however it is those regions with the most limited access that have the least local provision and consequently the least capacity to build up that provision.

Ethical and cultural issues in data sharing

Issues of ethics and informed consent for data sharing were expected across many of the projects (DED, NDF, ICK, ECV, HMP) and these did surface in practice. Despite some evidence of a past lack of detailed consideration of informed consent issue in a few projects and opportunities to improve practice, anonymity and consent were generally well handled. Contributing projects took privacy and consent seriously and there was evidence throughout of careful thought being applied to principles of operation, even where there was some lack of adherence to best practice.

Contributing projects showed a desire to consider what could be shared and how to enable this with due consideration of ethical issues. For instance the LASDEL group (NDF) worked to identify aggregate data that could be shared from datasets with significant issues of privacy and potential for harm to participants. Derechos Digitales developed a characterisation of data resources that included public, available on request to specific groups, and secure. In the case of Natural Justice where the decision was taken that none of the primary material could be shared there was still an effort made to identify objects (including materials generated by the research teams themselves) that might be shareable.

In the case of the Natural Justice project involving Indigenous Knowledge a substantial barrier to sharing were the inconsistent and overlapping obligations of the various project contributors to many different players (Neylon 2017g). Firstly there were two institutions involved, each with their own IRB and informed consent process, with a bias towards retaining data securely. The funding arrangements contained further obligations with a contrary bias towards data sharing. Finally the project itself was focussed in part on developing community contracts intended to define the relationship between the project and the participants, who were framed as partners in the research program (Traynor et al. 2015). The contradictions between these overlapping obligations and ethical frameworks have the potential to create substantial tensions including, as the project contributors noted, serious conflicts of interest for researchers themselves. Harmonising policy and process requirements across stakeholders is a serious issue currently mitigated only by the fact that few projects face the level of complexity of the Natural Justice Project.

A broader cultural issue which was consistent across the projects and is consistent with other reports (Borgmann 2014, Fecher et al. 2017) was the sense of ownership and desire for control over data. This arose both in the decision to develop local data sharing infrastructures in three cases but also more generally. Concerns raised with giving up control over access ranged over several issues including concerns for security, mis-use of the data, and an ability to report on details of usage.

An explicit desire for credit was not raised in contrast to concerns over control. For instance the hypothesis that formal publication of the DMPs would encourage preparation was not supported by the experience of the project. This differs slightly from other recent surveys of concerns and motivations for data sharing but might be due to the prevalence of non-university groups within the contributing projects. A perceived need to be able to report on details of usage to funders (raised explicitly by TED and HMP) could be seen as relating to credit.

Permissioning and deeper issues: Can indigenous knowledge be "data"?

The process of Data Management Planning surfaced a range of permissions issues. In the DataFirst project explicitly raising questions of rights led to the realisation that for several of the datasets that had been offered to the project the data holders either did not have explicit rights to make them publicly available or had no clear rights at all (Neylon 2017f). For Derechos Digitales the process led to a realisation that informed consent processes for interviews and surveys could be strengthened (Neylon 2017a).

The Natural Justice project raised a much more fundamental set of questions. This project was explicitly included to probe issues relating to the ethics of sharing data that arises from engagement with indigenous knowledge, in this case knowledge of managing the effects of climate change within indigenous South African communities (Foster 2014, Traynor et al. 2015).

As noted above this project raised issues of tensions between institutional ethical and policy requirements. Running beneath these issues of procedural obligations are more

fundamental questions of ensuring justice and respect for a historically disenfranchised community. This community holds information that is seen as potentially valuable assets by the western knowledge system. In thoughtfully addressing the tensions in play here the contributing project team have made a commitment to maintaining the knowledge resources that they hold in their original context.

The project team in assessing their existing obligations to the project participants decided that it was not feasible to make primary digital objects publicly available (Traynor 2017). However they noted that this was not inevitable, and that earlier planning might have made it possible to negotiate the public release of some objects. It is therefore useful to consider how this "data" might have been released in a way that preserves its full local context of collection.

Data is not a well defined concept, and the Pilot Project has not sought to provide a definition. Indeed the initial exercises in the first workshop deliberately sought to expand the participants consideration of what they might think of as data. Nonetheless the rhetoric around data management and sharing policies has a consistent thread. The arguments for data management and sharing focus on either re-use of data in new contexts or in their use to replicate or validate claims that are supposed to be supported by them. That is, the value of data is in its use outside its initial context.

Indigenous Knowledge Systems will in many cases be incommensurate with Western (Scientific) Knowledge Systems. This already raises questions of how "data" derived from interrogation of indigenous knowledge systems can be used in scientific knowledge systems. Such digital objects will be boundary objects, with the concomitant lack of understanding on both sides of the boundary as to their implications in the two knowledge systems.

This may or may not be a problem in itself for knowledge production. The question needs to be raised as to whether the concept of data makes sense within a specific Indigenous Knowledge System. Western Scientific Knowledge Systems reify a concept of data rooted in the notion of "matters of fact" developed by Boyle and others in the 17th century (Shapin and Schaffer 1985). Even within modern models of science as a socially constructed process the status of data as shared objects which are intended to approach objectivity and generality is central. Again, the status of the data object lies precisely in its capacity to retain utility when decontextualised (see for instance Leonelli 2011 on the role of decontextualisation and recontextualisation in scientific data re-use).

However, the in principle problem of communicating and understanding implications of these boundary objects in two incommensurate knowledge systems raises profound ethical questions, not least on the possibility of informed consent. Informed consent may simply not be feasible where communication is limited. Current solutions for dealing with situations where informed consent is not feasible are generally both paternalistic and rooted in western conceptions of knowledge and ethics. Fundamentally the concept of an abstracted data object, and the set of digital affordances designed for its transmission and sharing, is incompatible with the values expressed by the IKC project, who regard an important part of

their role as protecting the context of the artifacts they have "collected" or "produced" (Traynor et al. 2015, Traynor 2017, Neylon 2017g). The imposition of the concept of "data" in its western sense is potentially unethical in and of itself.

As the contributing project notes in its own work (Foster 2014, Traynor et al. 2015), Indigenous Knowledge is handled inconsistently in international law and in local implementation of international treaties into national law. The Pilot Project has also surfaced issues with how indigenous knowledge systems can be treated within the context of western-derived data practices and policies. More research into the underlying epistemological and ethical issues and how they are aligned or in tension is merited and may have substantial value in helping to resolve how the interaction of these knowledge systems can be managed in a principled fashion.

Changing researcher culture?

The underlying theory of change for the project led to the view that the ultimate goal for achieving effective data sharing and management behaviour involves cultural change amongst grantees supported by appropriate policy, technical frameworks, and support systems. Overall the project was too short to provide evidence for sustained culture change. It is however possible to look at pre-existing culture and how funder actions and other environmental factors may contribute to the targeted change.

In all cases it was possible to observe an existing sympathy for and motivation to share data in principle. While within the Pilot Project this may be a consequence of a deliberately biased sampling process this is consistent with other studies that show in-principle support for data sharing and good management practice is very high (Fecher et al. 2017). Other studies show that in practice it falls short, with issues of credit assignment, workload, ethical obligations, and lack of benefit given as reasons for the gap between theory and practice (Borgmann 2014). Therefore our focus is on how actions of the funder might strategically drive cultural change that closes these gaps.

Contributing projects noted that the funder interest, combined with varying degrees of pre-existing motivation, contributed to their willingness to engage with the issues of data management and sharing. Signalling from a funder that it values improved practice in data management and sharing, alongside an expressed aspiration to achieve stated goals through that enhanced data availability supported awareness amongst contributing projects.

Throughout the project there was limited interest in the details of any policy. Reactions to policy details were always specific to a defined situation, in most cases a challenge to a choice made about the mode of data sharing (limiting of access (TED), requiring login, choice of licenses (HMP)). Rhetoric relating to policy was therefore generally either abstract and positive, or concrete and negative. In addition in an exercise at the final workshop posing the question of what an optimal policy design would look like there were few novel suggestions for policy elements. The consensus amongst the group was for

requiring sharing and planning while allowing some exceptions. However, there was little evidence of this view being related back to their experience and in the context of the rest of the report we do not recommend this as an immediate policy.

By contrast when asked what additional support the projects could use there were a range of requests that clearly derived from the contributing project's experience. Limitations to the support available consistently provided a barrier to progressing the agenda. These were generally small issues but sufficient to deprioritise action. This suggests that providing capacity and support may be more important than the details of policy design.

Following the process of developing the Data Inventory and Data Management Plans most contributing projects reported that they would apply similar processes in future. Particularly across the projects with limited experience of data management there was a substantial improvement in self-reported quality of backup and archival practice for data. Whether this improvement continues may be worth following. Nonetheless the process was linked to a narrative of good practice and to a process that had generated some value. It should be noted that this was in the context of a supported process and not an administrative requirement prior to grant submission. Nonetheless it would be reasonable to expect that if the contributing projects encounter a requirement for DMPs in the future they will now be better prepared.

Overall there was evidence of both motivation for and sympathy with an aspiration for greater data sharing. This was expressed in the context of interactions, with the Pilot Project, standing in some ways as a proxy for the funder. This aligns with the findings of the initial review in which the use of ongoing Data Management Planning as a means of contact and communication between grantee and funder. Through continued interaction and support the funder both signals the value placed on data sharing and continues the interactions that strengthen the internal narrative leading to cultural change.

There remains further work to be done comparing interventions in terms of success in promoting culture change. The current study did not compare interventions. However the fundamental notion that funder actions can have an impact on culture change, is at least validated in principle. The central finding is that interactive and practical interventions relating to practice had a greater identifiable effect amongst this group than abstract and policy based interventions. However the signalling that policy statements provides provide a valuable prompt that is consistent with the pre-existing motivation towards data sharing in principle.

Summary of case studies

The case studies showed a high level of diversity amongst the contributing projects. Those things that were expected to be barriers to data sharing and good management practice - network access and reliability, existing knowledge and capacity, language issues, and ethical constraints - were all contributing factors to limited data sharing in practice. Overall the timeline was not long enough to observe substantial data sharing from the projects.

However reported quality of practice on backup and archiving of data improved across the projects, particularly those with less experience at the beginning.

Language issues for projects not operating in English was a much more substantial problem than was expected. While mitigation efforts including translation of key materials, and multi-lingual support platforms were valuable the assumption of English language operation permeates through every layer of the systems and services relevant to data management and sharing.

Issues arising from the handling of indigenous knowledge were expected, however the challenges run much deeper. Further work considering how issues of knowledge systems, digital objects, concerns for social justice, and informed consent may be in tension. It may be inappropriate to use the term "data" with respect to indigenous knowledge until its implications can be disentangled.

Finally with respect to culture change, within the contributing projects there seemed to be a greater effect that could be traced to interactions and support than could be traced to abstract policy. Nonetheless the statement of intent and aspiration is also important. Demonstrating that the issue is important to the funder seems as effective as any specific policy decision.

Challenges at the funder level

Capacity to monitor and support policy implementation is a systemic structural issue

An early finding from the initial review (Neylon 2017h) was that the structure of funding organisations creates a systemic issue with policy implementation. Most funders have an organizational structure where specific individual Program Officers are the main point of contact for specific grantees. Depending on the scale of the organization these same Program Officers may also have responsibility for developing grants in collaboration with grantees, designing programs, and soliciting proposals, as well as monitoring and evaluating performance both within and post-grant.

Larger funders do often have separate evaluation and policy units but most retain the single main point of contact with grantees. In most cases therefore all policy change, policy implementation, and the evaluation of policy compliance is handled by individuals who are not specialists in the policy area and are already responsible for tracking compliance with existing policies for grants. New policy, whether related to Data Management or other topics requires Program Officers to gain new expertise so as to offer guidance, support and monitor implementation.

Large funding organisations have the capability to put additional support in place. The most successful policy interventions in this area, those requiring Open Access through deposit in Pubmed Central and Europe Pubmed Central by biomedical funders, have combined an infrastructure for implementation with infrastructures to support monitoring. Large funding

organisations are also capable of interacting directly with Research Performing Organisations, as well as through individual grantees. The policy approach of the UK Engineering and Physical Sciences Research Council on data management has created a response from universities because of that level of direct engagement.

Smaller funders, where interactions are focussed on individual Program Officers, face challenges in delivering this level of infrastructure. In the case of IDRC the burden of policy implementation falls almost entirely on individuals. There is some evidence of specialisation within program groups, with individual Program Officers focussing on particular policy areas, but this is not institutionally supported and there appears to be little strategic coordination in managing capacity and workload.

In discussions with Program Officers there were striking differences in both the perception of how workload was intended to be distributed and what different activities might be categorised under. In particular the way in which workload was discussed was usually divided into parts of the grant life-cycle (program design, project solicitation and selection, monitoring, and project/program evaluation). The way in which activities, such as policy implementation, that are relevant throughout the grant lifecycle contribute to workload did not appear to be a subject that had been considered at a strategic level.

Leading culture change amongst grantees requires funder capacity

The goal, and theory of change, has been articulated as involving culture change amongst two groups of actors in differing institutional contexts: grantees within their research practicing organisations, and staff within research funding organisations. In theory, the introduction of the right policy framework, alongside appropriate support and infrastructure, will lead to this change in culture, and therefore of practice. The Pilot Project was designed to focus on grantees so as to understanding institutional, logistical and cultural barriers to this change. However this work has consistently surfaced a parallel need for a focus on the culture, institutions and capacities within the funder (Neylon 2017h). As a referee of the original version of the project literature review noted (O'Donnell 2017), the various cultures found within a funder and other interacting actors including government bodies also need to be considered.

Intentional design of cultural change is challenging and can be approached from a number of angles. The Institutional Analysis and Development (IAD) framework (Ostrom, 2005) Fig. 1 provides an institutional focus framing of this, which is a useful contrast to the narrative and discourse focus that has been applied thus far.

The IAD framework focusses on two spaces which drive actions and practice. The first is the environment, made up of i) "bio-physical characteristics" which in our case includes available technical infrastructures, ii) attributes of communities, and iii) rules in use. The latter two both contribute to what has been more loosely described as "culture" within the current project. It is convenient for our purposes to modify the distinction slightly to designate "formal institutions", which includes policy and regulations, and "culture" or

informal and implicit institutions. Part of our argument throughout has been that these can, and in some cases do, act in opposition to each other.

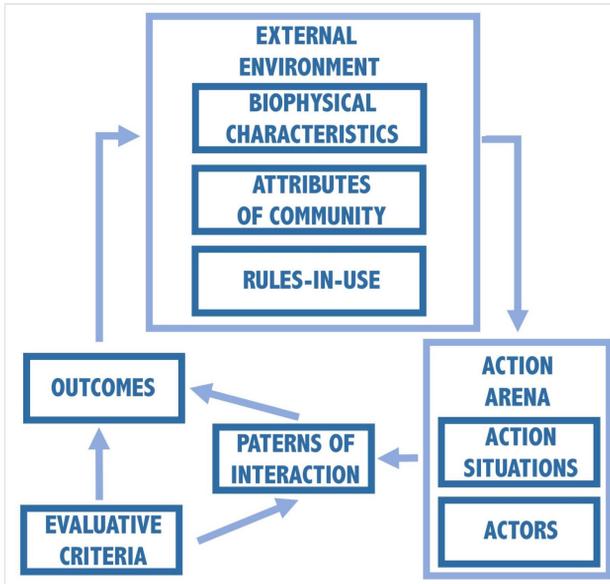


Figure 1. [doi](#)

The Institutional Analysis and Design framework adapted from (Ostrom 2005).

The second main space in the IAD framework is the "action arena". This is the space in which actors interact in specific "action situations". The action arena is where practice occurs. In the IAD framework what happens in the action arena is analysed in terms of the patterns of interaction. Patterns lead to outcomes or may be further subject to evaluative criteria that feed into outcomes.

The value of the IAD framework is in helping integrate the way that differing changes shift outcomes within the system. It can also help us to identify what is needed to make intentional change. In particular if our aim is to change culture, that is the informal and implicit institutions within grantee and funder communities, then changing rules in use, the formal institutions, is likely insufficient. Changes in outcomes are required to drive changes in culture. Changes in outcomes will be delivered through changes in evaluative criteria. However this requires additional capacity for that evaluation above and beyond what is already institutionalised.

An alternative framing using the group-level model of Cultural Science (Hartley and Potts 2014) reaches the same conclusion. In the Cultural Science framing culture makes groups, not the other way around. The group in turn enacts the culture in a reinforcing cycle (Fig. 2). Therefore we begin by positing a "target culture" of data sharing and ask how it might grow through attracting new group members or in competition with a traditional "non-sharing" culture. The cycle of culture-made group and group-making culture exists in a

complex environment. Therefore intentional change requires consistent evaluation and investment, again requiring additional evaluative capacity.

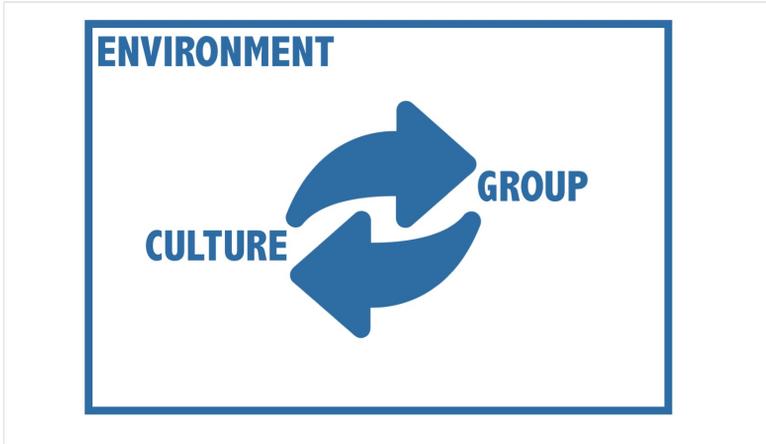


Figure 2. [doi](#)

The Cultural Science model of Hartley and Potts (2014). The co-creation of culture and group in the context of an external environment.

More concretely, as noted above, it is possible to point to specific successful cases of encouraging open practice where funders have invested in embedded evaluative capacity as part of the process. Examples include the NIH and Wellcome Trust Open Access policies (where monitoring is implemented through the designated repository). It is also the case that unsuccessful and weakly successful policy implementation by funders is generally associated with a lack of consistent and coherent evaluation. This is consistent with Ostrom's (Ostrom 1991) finding that successful solutions to collective action problems usually involve a monitoring system that functions as a side effect of implementing the desired behaviours and practice.

For Data Management and Sharing, where the diversity and sites of deposition of objects makes monitoring substantially more challenging this evaluation is clearly challenging. The ideal is probably a dynamic Data Management Planning system, which provides a space shared between funder and grantee and supports the ongoing management and monitoring of data production and deposition. Such systems are not currently available, except in specific purpose designed cases. In the shorter term the most productive process is likely to be an increased expectation of reporting on data availability and deposition and ongoing informal engagement by Program Officers on the issue of data management throughout a project.

Culture change is a collective action problem that starts with the funder

Regardless of the framing used to discuss culture it is a collective property of the group, affected by both external (environmental) and internal dynamics. Thus intentional culture

change is a collective action problem. An implicit aspect of the funder narrative is the goal of working "with the community" to guide and support change. This implies a narrative in which the funder, and by extension Program Officers themselves are part of the community.

The findings from the case studies suggest that coherent signalling from the funder is important in motivating both immediate action *and* longer term change amongst grantees. The lack of engagement with the specifics of policy, which is not limited to the current project, suggests that this signalling supports an overall motivation towards culture change but does not necessarily directly drive changes in practice. Evaluative and administrative requirements can drive specific changes in practice but do not necessarily act in synergy with a motivation towards culture change. In particular interventions at a specific point in the project life cycle that are not coupled to ongoing interactions do not lead to changes in culture. Both the IAD and Cultural Science frameworks predict this.

Ongoing interactions relating to an aspirational policy change are resource intensive and require change in practice at the funder level (as well as ideally amongst other stakeholders). The question arises therefore how to support practice amongst Program Officers that incorporates efficient and ongoing interactions relating to data management and sharing (or indeed other policy initiatives). Such a substantial change in practice requires that it be embedded in the day to day activity and narratives underpinning the work of Program Officers. That is, it requires culture change within the funder.

Alongside the systemic structural issue of expecting Program Officers to act as change-makers across the portfolio, this argument suggests that policy design needs to consider how it articulates and interacts with the culture within the funder as much as with that of grantees. The question of how policy design affects Program Officers has not been examined in any detail, alongside the related question of how strong and coherent culture can be developed within a funder and its staff.

Findings and recommendations

Diversity of the grant portfolio and community is a challenge

There is a huge diversity amongst projects, grantees and host organisations within this project and amongst funders more generally. This makes drawing general conclusions very challenging. There is variety not just in technical capacities, network access, and access to IT resources, but also in research models, underpinning values, and ways of thinking about research. Any strict policy requirements need to reflect this diversity and provide the diverse support and resources that can help grantees to deliver on those requirements.

Developing a theory of change for policy implementation

The main shift over the course of the project was from an implicit theory of change in which policy implementation lead to behavioral change to the idea that the goal was to achieve culture change within grantee communities. The articulated theory of change for this report

was therefore that culture change towards a practice of improved data management and sharing could be supported by policy and support interventions. Following the analysis we can articulate a more detailed theory of change which can be used to underpin policy design.

A necessary but not sufficient condition for culture change is an existing sympathy within the individuals that make up the group and a sense of shared purpose for change with the policy maker. Existing work and the results of this project suggest both of these are present amongst most researcher groups, although the articulation of the aspirations of policy-maker can often be unclear. This provides the "field" of motivation for change. Both intrinsic and extrinsic motivation are important.

Intentional culture change is a cyclical process involving reinforcement and interaction over multiple cycles of grants and projects. It is in that sense an evolutionary process (Wilson et al. 2014). Behaviour can be modified through stimulus response processes, and this leads to instrumental responses, in the current case the "compliance culture" discussed as an undesirable outcome in the initial review (Neylon 2017h). For interactions to engage with positive culture change they need to be explicitly connected with overarching shared narrative for change.

Maintenance within the cycle is an active process. A group may exit the cycle because it no longer interacts with the funder. Such groups cannot be a target for cultural change. They may exit because they no longer feel a shared sense of purpose with the funder, which might be the result of an active intervention (being required to do something that does not align with their values) or a lack of intervention. In these cases the message is sent that the purported interest of the funder is not real. Finally they may exit due to a lack of interaction and reinforcement over time.

Interactions and reinforcement may be antagonistic or even sites of conflict between funder and grantee or community. However if that conflict is resolved productively in a way which is seen to align with the shared purpose its end result may be positive in the long term. Identifying and negotiating around sites of conflict and contestation may be a productive way of surfacing and dealing with issues of implementation. This suggests that it may be more important that interventions align with perceived shared purpose and reinforce throughout and across project life cycles than that they actually provoke the final target behaviour. The funder is also in a position to modify the environment both to lower the costs of behaviour change (provide resources, support that makes it easier) and to enhance the outcomes of evaluation to the benefit of those engaging.

This model has the potential to be expanded to include other relevant stakeholders that were not examined as part of this project. It additionally implies that mediated culture change is not feasible unless an alignment on aspiration for change can be found. It also emphasises the active nature of the process, and argues against the feasibility of "set and forget" policy interventions. It also suggests a focus on groups that continue to interact with the funder as opposed to those with less interaction.

Platforms vs projects

Amongst the means available to funders to shift the environment is the provision of support platforms. These may range from internal support and training within a funder to third party technical platforms, and include support in the form of training, expertise, technical provision or infrastructure. Two of the contributing projects provide such support platforms in the form of technical infrastructures. Their interaction with the Pilot Project relates to two specifically funded projects, and this is a common mode for seeking sustainability for such platforms.

A project-based model is not well suited to sustaining infrastructures and platforms. Project-based assumptions in policy implementation are also not always a good fit. The IDRC has other models including that of supporting the [DECI-2 project](#) so as to provide support for grantees seeking advice on maximising impact and engagement. IDRC also provides some platforms internally (such as the Digital Library) and could coordinate with other funders to support third-party platforms.

A potential risk is the mixing of models. Where a platform is intended to support a policy objective of culture change, it must necessarily be able to provide assurances of long term sustainability. Where the goal is therefore to change the environment new funding models will be necessary.

Language and conceptual framing remain serious issues

Throughout the Pilot Project the issue of language barriers was deliberately engaged and interrogated. It was expected that this would be a significant issue. What was evident was how pervasive the English-language bias is within resources, systems and tools relating to data management and sharing.

This language bias also stands as an exemplar of a deeper challenge relating to the centrality of concepts of data to a very particular form of Western Scientific Knowledge System. Contradictions relating to Indigenous Knowledge were expected. However the fundamental questions of incompatibilities within the assumptions, policies, tools and language of a western and English-language framing of what data is, against a background of seeking not just equitable engagement with research participants but also to mitigate a history of expropriation is profound.

Practical policy development and implementation implications

The details of the implications of this study for policy design are discussed in detail below. The key finding is that a greater emphasis on the motivation behind policy change is merited and that this motivation should speak to internal as well as external stakeholders. Alongside this is the importance of ensuring the policy interventions and implementation actually align with the expressed motivations behind the policy.

It is important that requirements are backed with appropriate resources and support and crucially that this support is deployed in a way that underlines the funder commitment to the shared policy goals. Requirements should be structured so that they support interactions throughout the project life cycle.

Concretely this suggests that the traditional approach to data sharing policy, requiring a Data Management Plan at the point of grant submission, with little or no further follow up is inappropriate. DMP requirements need to be linked to policy motivations and support ongoing interactions between the funder and grantees. Requirements imposed on data sharing, such as the requirement for data associated with formal publications to be made available, or for all formal articles to have data sharing statements, must be auditable and audited.

Structurally there is an argument for separating the motivational parts of a policy statement, that are necessarily aspirational, from the implementation details. This enables flexibility and also enables a process of change for implementation, in line with the idea that culture change is a gradual process for both funder and grantee communities. Within all this the funder has a role to act where it can to change the environment to make compliance with policy requirements, and the development of good practice, easier and more productive.

Implications for policy design and implementation

This project has raised fundamental questions for policy design and implementation. If the ultimate target of cultural change is correct then many existing policy and implementation programs come into question. These programs and systems around them have focussed on changes in practice and behaviour (i.e. at the individual level). The apparent divide identified in this project's initial review between those in favour of traditional DMP requirements at grant submission and those concerned about the potential for creating a compliance culture can also be understood in terms of the distinction between individual behaviour and group level culture.

In addition the focus on culture change, as opposed to behaviour, also highlights the question of change amongst groups beyond grantees. Specifically it highlights how change within the funder (and by implication amongst other stakeholders) is an important aspect of achieving culture change amongst researcher communities. The focus on culture change also acts to highlight the need for capacity. Various framings of the problem of change emphasise how creating capacities that support changed practice (characterised for example as Biophysical Characteristics in the IAD framework) needs to be coupled to changes in policy (Rules in Use within IAD). IAD and Cultural Science both also emphasise evaluative capacity, linked to changes in outcomes, as the means of coupling capacity and policy to cultural change (Attributes of Community).

Key points for policy design

1. Policies serve at least two distinct functions. They serve as a) signals that a specific issue, in this case improvements in data management and sharing, are an issue that the

policy maker takes seriously and b) as a means of creating interventions in the behaviour of those subject to the policy. These two aspects of policy intent can be in tension with each other.

Policies in general have been designed with the intent of requiring and driving behaviour change. The aspect of signalling, and its role in supporting culture change has been taken less seriously. In the initial review it was noted that the aspirational signalling aspects of some policies were at odds with the content of their interventions.

2. The signalling function of a policy is important for culture change. Researchers and funder staff take note of the direction of travel. Articulation of aspirations, particularly when they align with existing sympathies and narratives is important. The details of policy design seem less important and can be antagonistic to cultural change where the aspirations are not matched by capacities provided by the policy maker.

3. Internal audiences are at least as important as external. Policy needs to be designed with a view as to how it can effect the desired cultural change *within* a funder as much as among researcher communities. How does the policy articulate the importance of its aspirational goals, does it give permission for individuals to act, how does it help create a strong culture that aligns with the policy goals? Perhaps most importantly does it provide the necessary internal levers to ensure that those charged with implementing policy can access the necessary resources, infrastructure and expertise?

4. The primary failure mode for policy is overreach. The most common issue with policy implementation is where demands are made that cannot be met or evaluated. Generally the concern is whether researchers have the capacities, resources or expertise to deliver on policy requirements. However the risks can be substantially higher where the shortfall is at the funder.

The centre of our developing theory of change is that the core of culture change is a sense that the aspirations underpinning the change are both important and shared amongst stakeholders. Where a grantee falls short, allowances can be made in a way that reinforces the importance of the direction of travel to the grantee. However when a funder fails to provide the support and infrastructures - including Program Officer time, skills and expertise - necessary for policy implementation, and most particularly when implementation is not evaluated, the message is sent that the issue is not in fact important. This in turn leads to instrumental behaviour and in turn a lack of, or even negative, cultural change.

5. The interventions required by policy must be properly resourced, continuous, and self-reinforcing. The second aspect of policy remains important. Interventions at the level of requirements and practice have a role to play, both as opportunities to engage - and therefore transmit culture - and as a means of driving best practice and therefore achieving successes for grantees from their behaviour change. For this goal to be achieved the interventions need to be embedded in an ongoing and reinforced narrative that engages both researchers and funder staff.

Data Management Planning as a worked example

The specific example of Data Management Planning is illustrative. The DMP process is applied within policy implementation as an intervention, albeit with the intention of "forcing researchers to consider the issues" i.e. with the goal of culture change. However benefits of good practice and management are rarely articulated as part of the goal of the policy. Signalling does not align with intervention (points 1,2).

The introduction of Data Management Planning as a requirement is rarely coordinated with internal staff capacities and aspirations (point 3). The role of the DMP as a potential means of grantee engagement could be, but rarely is, explored. In practice, Program Officers are left with yet another object to assess which does not appear to fit with their existing workflows or support their existing evaluation requirements. In many cases the DMP requirement is implemented within a grant submission system with which grantee-facing Program Officers may not even interact. The introduction of a DMP requirement needs to be coupled with a shared understanding amongst Program Officers of how it serves their aspirations and narratives as well as helping them do their work in a practical sense.

In the interviews for the initial review, those experts concerned with the risks of DMP requirements most commonly raised the concern that they were not well assessed, and were not used after the grant proposal process. Good implementations generally involve significant support and infrastructure development (point 4), interrogation and on-going use of the DMP, or evaluation of performance against the stated goals over the course of the project (point 5). Poor implementations involve a requirement for a badly specified DMP that is prepared too early in the project lifecycle to match reality, is frequently not assessed, and is not used either by the funder or the grantee at any later stage.

Overall the findings of this project emphasise that for DMP requirements to be supportive of culture change they need to be well supported with expertise, systems and guidance in place. They additionally need to be clearly aligned with the funder narrative that underpins the policy goals. Finally, and consistent with the findings of the initial review, they should provide the basis for continuing interactions between funders and grantees, and provide additional value for the grantees themselves throughout the project life cycle.

This means that the capacities of funder place clear limits on the scope and speed of implementing a data management policy. Strategic targeting of implementation is likely to be necessary and this is one reason for separating the aspirational narrative from the details of implementation (point 1,2). Aligning implementation details, particularly requirements, with the capacity for monitoring progress will be crucial, whether this is formal through infrastructural systems or informal, through interactions with POs.

Recommendations for (development) funders

Much of the central recommendations for funders of research generally and development research specifically are already embodied in the policy design principles described above. At a high level the major recommendations can be described as follows.

1. Centre policy design in clearly articulated narrative of the motivations and goals of change. Adopt the principles for policy design discussed above.
2. Separate policy motivation and implementation documents so as to both be clear on direction of travel and allow flexibility and gradual strengthening of requirements.
3. Ensure that sufficient support is in place so that any requirements can be delivered, monitored, evaluated, and celebrated before policy requirements are imposed. Capacity and support is more important than the details of policy design.
4. Recognise the importance of internal stakeholders as an audience for policy design and internal culture and practice change as an important part of implementation
5. Support critical enquiry into the origin of discourses around data and how appropriate they are, with a particular focus on those projects engaging indigenous knowledge.
6. Contribute to the development of DMP platforms that will support an ongoing interaction between grantees and funder and that provide ongoing value for the researcher.

Specific recommendations for IDRC

Specific recommendations for IDRC are focussed on the immediate follow-up to the current project and to the project goals as specified in the original proposal. In particular they address next steps for developing an IDRC Policy and an Implementation Plan. The most important question to address is that of what capacity is available or can be made available, and what limitations that places on the pace and scope of implementation.

1. Clarify and develop a shared view of the core goals and aspirations for IDRC in developing and implementing Data Management and Sharing expectations.
2. Audit the capacity for POs and other staff to support data management and sharing practice with respect to targeted and global implementations. Identify gaps.
3. Continue to work with partners to develop shared capacities for data sharing and data management infrastructures, including training and support capacities for grantees and internal staff.
4. Consider the value of supporting a project that expands on the work of the Pilot Project in providing support to grantees and POs in implementing Data Management and Sharing arrangements.
5. Follow-up with the contributing projects to qualify progress on the self-reported improvements in practice and the possible change in culture identified through the Pilot.

Acknowledgements

A very large number of people have contributed to this project. Leslie Chan as co-PI was a major player in defining the project plan and supporting the work. Staff at IDRC including (but not limited to) Sue Godt, Edgard Rodriguez, Fernando Perini, Marie-Gloriose Ingabire, Raed Sharif, and Natacha Lecours, generously gave time and advice in developing, prosecuting and analysing the project as well as giving their perspective on recommendations and practical ways forward.

The project benefited from expert advisory support from Simon Hodson, David Carr, Pascal Aventurier, Carly Strasser and Chuck Humphrey. Other participants at the workshops who gave valuable input included Stephanie Simms, Jeremy Geelen, Richard Akerman, Suzanne Board, Matthew Lucas, Philippe-Olivier Giroux, and Arash Shahsavarani.

The Pilot project would have been impossible with the generous contribution of time and significant efforts of the contributing projects, with substantial input from Aissa Diarra, Cath Traynor, Laura Foster, Danae Tapia, Constanza Figueroa, Dora Canhos, Le Dang Trung, Lynn Woolfrey, and Reem Wael. Finally the project was driven in large part by the energies of Barbara Porrett, who maintained the connections, pushed as necessary and made progress possible throughout the entire project. All the errors and infelicities in this final report are mine, most of those that have been expunged are due to Barbara's support in editing.

Grant title

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

References

- Borgmann C (2014) *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, Cambridge, Massachusetts. [ISBN 9780262327855]
- Canhos DAL (2017) *Data Management Plan: Brazil's Virtual Herbarium*. Research Ideas and Outcomes 3: e14675. <https://doi.org/10.3897/rio.3.e14675>
- Chan L, Okune A, Sambuli N (2015) What is open and collaborative science and what roles could it play in development? In: Albagli S, Maciel M, Abdo AH (Eds) *Open Science, open issues*. IBICT, Rio de Janeiro. URL: <http://livroaberto.ibict.br/handle/1/1061> [ISBN 978-85-7013-111-9].
- Fecher B, Friesike S (2014) *Open Science: One Term, Five Schools of Thought*. In: Barling S, Friesike S (Eds) *Opening Science*. Springer, Heidelberg. [ISBN 978-3-319-00025-1]. https://doi.org/10.1007/978-3-319-00026-8_2

- Fecher B, Friesike S, Hebing M, Linek S (2017) A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications* 3: 17051. <https://doi.org/10.1057/palcomms.2017.51>
- Foster (2014) Critical Cultural Translation: A Socio-Legal Framework for Regulatory Orders. *Indiana Journal of Global Legal Studies* 21 (1): 79. <https://doi.org/10.2979/indjglolegstu.21.1.79>
- Hartley J, Potts J (2014) *Cultural Science: A natural history of stories, demes, knowledge and innovation*. Bloomsbury 3PL, London. [ISBN 9781849666022]
- IDRC (2014) IDRC Strategic Plan. <https://www.idrc.ca/en/about-idrc/governance/our-strategic-plan>. Accessed on: 2017-10-20.
- Leonelli S (2011) Packaging Data for Re-Use: Databases in Model Organism Biology. In: Howlett P, Morgan MS (Eds) *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*. Cambridge University Press, Cambridge. [ISBN 978-0-521-19654-3].
- Neylon C (2017a) Case Study: Derechos Digitales. *Research Ideas and Outcomes* 3: e21698. <https://doi.org/10.3897/rio.3.e21698>
- Neylon C (2017b) Case Study: Strengthening the Economic Committee of the National Assembly in Vietnam. *Research Ideas and Outcomes* 3: e21699. <https://doi.org/10.3897/rio.3.e21699>
- Neylon C (2017c) Case Study: Neglected Health Issues in Niger. *Research Ideas and Outcomes* 3: e21700. <https://doi.org/10.3897/rio.3.e21700>
- Neylon C (2017d) Case Study: Brazilian Virtual Herbarium. *Research Ideas and Outcomes* 3: e21701. <https://doi.org/10.3897/rio.3.e21701>
- Neylon C (2017e) Case Study: HarassMap. *Research Ideas and Outcomes* 3: e21702. <https://doi.org/10.3897/rio.3.e21702>
- Neylon C (2017f) Case Study: Tobacco Economics Control Project. *Research Ideas and Outcomes* 3: e21703. <https://doi.org/10.3897/rio.3.e21703>
- Neylon C (2017g) Case Study: Indigenous Knowledge and Data Sharing. *Research Ideas and Outcomes* 3: e21704. <https://doi.org/10.3897/rio.3.e21704>
- Neylon C (2017h) Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes* 3: e21705. <https://doi.org/10.3897/rio.3.e21705>
- Neylon C (2017i) Dataset For Idrc Project: Exploring The Opportunities And Challenges Of Implementing Open Research Strategies Within Development Institutions. International Development Research Center. Zenodo <https://doi.org/10.5281/ZENODO.844394>
- Neylon C (2017j) Data Management Plan: IDRC Data Sharing Pilot Project. *Research Ideas and Outcomes* 3: e14672. <https://doi.org/10.3897/rio.3.e14672>
- 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>
- O'Donnell J (2017) Review of: 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.r57782>
- Ostrom E (1991) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge. [ISBN 978-0-521-40599-7]

- Ostrom E (2005) *Understanding Institutional Diversity*. Princeton University Press, Princeton. [ISBN 9781400831739]
- Shapin S, Schaffer S (1985) *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press, Princeton. [ISBN 9781400838493]
- Traynor C (2017) Data Management Plan: Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights. *Research Ideas and Outcomes* 3: e15111. <https://doi.org/10.3897/rio.3.e15111>
- Traynor C, Foster L, Schonwetter T (2015) Multi-Institutional Research Collaboration: Regulation, Ethics and Power. <https://ocsdnet.org/multi-institutional-research-collaboration-regulations-ethics-and-power/>. Accessed on: 2017-10-13.
- Wael R (2017) Data Management Plan: HarassMap. *Research Ideas and Outcomes* 3: e15133. <https://doi.org/10.3897/rio.3.e15133>
- Wilson DS, Hayes S, Biglan A, Embry D (2014) Evolving the future: Toward a science of intentional change. *Behavioral and Brain Sciences* 37 (04): 395-416. <https://doi.org/10.1017/s0140525x13001593>
- Woolfrey L (2017) Data Management Plan: Opening access to economic data to prevent tobacco related diseases in Africa. *Research Ideas and Outcomes* 3: e14837. <https://doi.org/10.3897/rio.3.e14837>