|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | IDRC  Data Sharing Workshop |  | |  |  | |
| Exploring the opportunities and challenges of implementing open research strategies within development institutions.  9 – 10 March 2016, Ottawa, Canada | |

# Introduction to the Data Sharing Pilot

## Introduction

The aim of the IDRC Data Sharing Pilot is to refine guidelines for the implementation of development research funders’ open research data policies and to inform IDRC on the design and implementation of its Data Management and Sharing policy. The Pilot funded as an IDRC grant, will conduct open data case studies with eight (8) IDRC grantees to develop and implement open data management and sharing plans. The case studies will examine the scale of legal, ethical and technical challenges that might limit the sharing of data from IDRC projects including issues of:

* Privacy, personally identifiable information and protection of human subjects
* Protection of intellectual property generated from projects or potential for financial risks for projects or institutions
* Challenges in the local legal environment, including ownership of data
* Ethical issues in releasing or sharing of indigenous and community knowledge, and the relationship between project participants and investigators particularly in the context of historical expropriation of resources
* Local and global issues of capacity and expertise in the management and sharing of data

The Pilot commenced in October 2015 and will finish at the end of 2016. Case studies conducted with the eight pilot projects will run from March to late October. Each pilot project will be assisted in the process of conducting a data audit, a Data Management Plan (DMP) and the implementation of that DMP.

## Pilot Timetable

|  |  |
| --- | --- |
| October 2015 | Pilot Commenced |
| January 2016 | Pilot Projects Selected |
| March | Kick off workshop, Data Audits |
| April | Data Management Plans completed |
| May | Training for project participants |
| April - October | DMP Implementation |
| October | Final assessment of DMP implementation and Review Workshop |
| January 2017 | Final report and outputs due |

# Expectations

## What is expected from each project?

Your time and engagement! Including at the workshops.

1. Two documents:
   1. A data audit
   2. A data management plan
2. Engagement and work on implementing your DMP
3. Reporting on the experience of implementing your DMP
4. Contributing to the process of refining IDRC and funder policy and implementation

## How will the Pilot support each project?

1. The Pilot team is available for advice and support throughout the Pilot
2. An advisor will be assigned to support your project and provide advice and guidance
3. Training on good data management and repository deposition will be provided
4. The Pilot team will provide direct support to each project in the preparation of the Data Audit and the Data Management Plan, either via voice/video calls or email as preferred
5. The Pilot team will provide direct support to each project in the implementation of the DMP and data sharing practice, either via voice/video calls or email as preferred
6. The Pilot team will share with all projects experiences and insights that arise from each of the projects

# Workshop Program

## Wednesday 9th March

|  |  |  |
| --- | --- | --- |
| 09:00 | 09:15 | Arrive and Coffee |
| 09:15 | 09:30 | Welcome from IDRC |
| 09:30 | 10:30 | Introductions and welcome |
| 10:30 | 11:00 | Morning tea |
| 11:00 | 12:00 | Introduction to the projects |
| 12:00 | 12:30 | Exercise/breakout: What is data? Why share it? |
| 12:30 | 13:30 | Lunch |
| 13:30 | 14:00 | Introduction to Pilot |
| 14:00 | 14:30 | Exercise breakout: What data will the Pilot generate? |
| 14:30 | 15:00 | Report back |
| 15:00 | 15:30 | Afternoon tea (Pilot team to collate reports for DMP exercise) |
| 15:30 | 16:30 | Group Exercise: Preparing a DMP for the Pilot |
| 16:30 | 17:00 | Discussion, lessons learnt, close |
| 18:30 |  | Dinner at Métropolitain Brasserie |

## Thursday 10th March

|  |  |  |
| --- | --- | --- |
| 09:00 | 09:30 | Arrive, Coffee: Brief recap of previous day. |
| 09:30 | 10:30 | Breakout: Challenges with sharing |
| 10:30 | 11:00 | Report back and discussion |
| 11:00 | 12:00 | Project breakouts: Initial data audit |
| 12:00 | 12:30 | Present to partner groups and co-review |
| 12:30 | 13:30 | Lunch (further discussion and adjustment as necessary) |
| 13:30 | 14:30 | Report back on data audit and discussion |
| 14:30 | 15:00 | Breakouts: Refine data audit in response to feedback |
| 15:00 | 15:30 | Afternoon tea |
| 15:30 | 16:00 | Repeat the first exercise: What is data? Why share it? |
| 16:00 | 16:30 | Report back |
| 16:30 | 17:00 | Discussion, next steps for the Pilot, close |

# Aims of the Workshop

## Aims of the Workshop

1. Introductions and meeting Advisors, Program Officers, and Project Participants
2. Understand the aims of the project, timetable, and the roles and responsibilities of the various contributors to the Pilot
3. An understanding of the research data outputs of each project and first steps towards the Data Audit for each project
4. An initial discussion of the issues and challenges associated with data sharing in general and specifically for each project
5. An introduction to the DMP Assistant Online tool

# Exercise 1: Human Barometer

I will make a statement. You should consider how strongly you agree or disagree with the statement. I will indicate one end of the room for “agree” and another for “disagree”. Position yourself between the two ends based on the strength of your agreement or disagreement.

For example: The statement might be “Ottawa has the best weather in the world”. If you agree strongly go to one side of the room. If you disagree strongly go to the other.

Advisor’s Note:

It would be useful to have people take contrary positions for this exercise e.g. “Data sharing is easy…doing it effectively is hard”

Statement 1: Data sharing is easy

Statement 2: There is some data that should never be shared

Statement 3: I would like anyone to be able to use my data

# Introduction to the Projects

Notes

|  |  |  |
| --- | --- | --- |
| Title | Investigator | Program Officer |
| Crowd Sourcing Data to fight Social Crimes | Reem Wael | Raed Sharif |
| Human Resources for Health in Rural China: An assessment of the current situation and projection of future needs | Minquan Liu | Marie-Gloriose Ingabire |
| Strengthening the Economic Committee of the National Assembly in Vietnam | Nguyen Thang, Trung Dang Le | Edgard Rodriguez |
| The Impact of Copyright User Rights | Danae Tapia | Fernando Perini |
| Establishing a clearinghouse for tobacco economic data in Africa | Lynn Woolfrey | Natacha Lecours |
| Les problèmes négligés des systèmes de santé en Afrique : une incitation aux réformes | Aïssa Diarra | Sue Godt |
| Indigenous Knowledge in Climate Change | Cath Traynor | Ellie Osir |
| Virtual Herbarium as OCS Infrastructure | Dora Canhos | Elie Osir |

# Exercise 2: What is data? Why Share it?

One of the challenges of data sharing and management, particularly when it comes to developing and applying policies. In this exercise we will look at what the term “data” can be considered to cover and what the motivations for data sharing might be.

The OECD Principles and Guidelines for Access to Research Data from Public Funding[[1]](#footnote-1) states:

“research data are factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated”

Advisor Note:

The key purpose of this exercise is to

a) widen awareness of the things that might be considered data (i.e. just about anything) and start focusing attention on the choices that need to be made and

b) to illustrate that researchers, funders, and program officers may have different motivations to support data sharing.

Would you consider the following to be “data”?

* spreadsheets
* survey questions
* articles
* video of interviews
* database of grant proposals

What forms of data have you produced, or expect to produce, in your project? Try to make your scope as wide as possible.

What do you think are the benefits of sharing the data you generate? Consider the different perspectives of project researchers and program officers.

# Exercise 3: What Data Will the Pilot Generate?

In this exercise we will start to work through the process of considering the data a project will generate. We will start on this by looking at the data the *Pilot* will generate. That is, aside from the data the *projects* are generating, what will the Pilot itself create.

Advisor Note:

The purpose of this exercise is to continue the process of illustrating how wide the scope of data can be and to explicitly point out that the *Pilot* will generate data as well. This then helps us setup the following exercise where we will collectively prepare a DMP for the Pilot.

This exercise will be used to generate the data types that go into the DMP. Amongst other things the team has conducted interviews (transcripts and audio files as well as some video), conducted a small survey of the projects (survey questions and answers), and generated an initial report on data sharing policies.

We will also be using the data audits and DMPs as a source of data alongside interviews and monitoring of the projects and their data sharing and management.

We want to emphasise that the participants are in effect research subjects. Our current view is that the Pilot does not require IRB approval, however we will need to operate within the parameters of existing ethics frameworks and approvals for the projects.

*Will the Pilot generate research data?*

*What data has it already generated?*

*What data will it generate?*

*What are the subjects of this research?*

*Does this raise any issues (ethics, privacy, technical issues)?*

# Data Management Planning

As we have identified in the earlier sessions, planning for effective Data Management is an important aspect of reducing additional workload and ensuring data and metadata are collected, they are properly cared for, sharing is managed, and archiving is planned for. There are a range of tools to support Data Management Planning. We are going to use DMP Assistant which has been developed by the Portage Project as a bilingual version of the UK Digital Curation Centre DMP Online tool.

The tool provides a template, with a series of questions that need to be answered. In this exercise we will work through the process of preparing a DMP for the Pilot as a whole, based on the data that we identified in the previous exercise. We will introduce the tool and seek to answer any questions that you have. The website for the tool is at: <https://portagenetwork.ca> lease register for an account at <https://portagenetwork.ca>

Notes

# Discussion

Notes

# Dinner

Dinner at Métropolitain Brasserie,

700 Sussex Drive, Ottawa, ON K1N 1K4

Booked for 1830

We will be selecting off a menu. We will need to each pay out of our per diem (where appropriate) for the food. Drinks will need to be covered by yourself.

# Exercise 4: Challenges of Data Sharing

Data sharing is not straightforward. In this exercise we will consider the different types of challenge that make data sharing more difficult. There are a range of challenges including, technical, ethical, legal and workload issues to sharing.

*Discuss within the group the data you are generating and the issues with sharing it?*

Advisor Note:

Our aim is to have as wide-ranging discussion here as possible. We will be planning the groups so as to get diversity between them. Ideally aim to draw out:

1. the differences in the contexts and formats/schema of the data between groups and the issues they raise
2. the different classes of challenges (technical, ethical, standards etc)

In the preliminary survey relatively few groups noted the workload challenges of effective data sharing. This would be useful to bring out at this point.

*How different are the issues that the other research groups face?*

*Are these issues different to those raised by sharing data from the Pilot?*

# Data Audit Introduction

The first stage of the Pilot is for each project to do a Data Audit. The idea behind a Data Audit is to identify what data will be generated by a project, what forms it will be in and to identify issues with managing and handling it. This information will provide the basis for preparing a Data Management Plan.

These questions are similar to the ones we asked in the initial survey we sent out prior to the workshop.

*List all the different forms of research data you are expecting to produce*.

*What file formats will this data be collected in?*

*Considering the data you expect to produce what challenges or issues do you expect to arise in sharing that data with others?*

# Data Audit Form

The form on the next page is to guide and assist the process of considering the data outputs that a project will generate. Start with the stages or phases of the project, sometimes expressed as Work Packages or deliverables, and for each one consider what will be generated, what formats and file sizes are likely and the potential issues involved in sharing.

An example is given based on the first two phases of the Data Sharing Pilot.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Process/Work Package | Expected Data Outputs | Description | File Formats/Size | Potential Issues |
| Literature review  Prepare a review of state of the art of funder policies including interviews with relevant experts. | ⌧ Spreadsheets | Summary of funder policies | Google Doc, small | Formats for sharing?  Informed consent?  License to release under?  Appropriate repository? Internet Archive?  Who needs to sign off on release? IDRC owns copyright. |
| ⌧ Questionnaires or interview prompts | Outline questions for interview | Google Doc, very small |
| ⌧ Interview transcripts/forms | Typed transcripts of interviews | Google Doc, small |
| ⌧ Audio/Video recordings | Audio and video recordings of interviews | mp3, mp4 |
| 🞏 Databases |  |  |
| 🞏 Research records/notebooks |  |  |
| 🞏 Images |  |  |
| ⌧ Report and drafts | Report and draft versions of report | Google Doc, pdf, small |
| Kick-off workshop  Preparation and record of the kick-off workshop for the Pilot. | 🞏 Spreadsheets |  |  | Informed consent?  Image/recording rights?  Who needs to sign off on release? IDRC owns copyright |
| ⌧ Questionnaires or interview prompts | Initial questionnaire for project participants |  |
| ⌧ Interview transcripts/forms | Responses to questionnaire |  |
| ⌧ Audio/Video recordings | Audio recording of workshop |  |
| 🞏 Databases |  |  |
| 🞏 Research records/notebooks |  |  |
| ⌧ Images | Photos taken at workshop, including flip charts |  |
| ⌧ Documents for meeting | Workbook |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Process/Work Package | Expected Data Outputs | Description | File Formats/Size | Potential Issues |
|  | 🞏 Spreadsheets |  |  |  |
|  | 🞏 Questionnaires or interview prompts |  |  |  |
|  | 🞏 Interview transcripts/forms |  |  |  |
|  | 🞏 Audio/Video recordings |  |  |  |
|  | 🞏 Databases |  |  |  |
|  | 🞏 Research records/notebooks |  |  |  |
|  | 🞏 Images |  |  |  |
|  | 🞏 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |  |  |
|  | 🞏 Spreadsheets |  |  |  |
|  | 🞏 Questionnaires or interview prompts |  |  |  |
|  | 🞏 Interview transcripts/forms |  |  |  |
|  | 🞏 Audio/Video recordings |  |  |  |
|  | 🞏 Databases |  |  |  |
|  | 🞏 Research records/notebooks |  |  |  |
|  | 🞏 Images |  |  |  |
|  | 🞏 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |  |  |
|  | 🞏 Spreadsheets |  |  |  |
|  | 🞏 Questionnaires or interview prompts |  |  |  |
|  | 🞏 Interview transcripts/forms |  |  |  |
|  | 🞏 Audio/Video recordings |  |  |  |
|  | 🞏 Databases |  |  |  |
|  | 🞏 Research records/notebooks |  |  |  |
|  | 🞏 Images |  |  |  |
|  | 🞏 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |  |  |

# Exercise 5: Return to the beginning.

# What is data? Why share it?

Give three forms of data that you might not previously have considered as in scope

* Filler
* Filler
* filler

filler

Give two new reasons for sharing data that you’ve discovered at the workshop, one that motivates grantees and one that motivates program officers

Advisor Note:

It could also be worth probing for new reasons to *not* share data of course. The realization that this is complex and not black and white is part of our objective.

Also try to tease out the differences in motivation between researchers and funders.

* Filler
* Filler
* filler

Consider how the different perspectives between grantees and program officers might be a *benefit* in delivering on the IDRC policy

Advisor Note:

Our aim here is to get across how different perspectives are useful in identifying issues that one group might miss. As per the review, the idea is to promote the concept that data sharing should be a collaboration between researchers and funders to change culture.

# Colophon

Image Credits

Idaho National Library - <https://www.flickr.com/photos/inl/5097547405> - Used under a CC BY 2.0 License

Got Credit – Share - <https://www.flickr.com/photos/jakerust/16639834358> - Used under a CC BY 2.0 License

OXLAEY.COM – The Soliga - <https://www.flickr.com/photos/oxlaey/15780341237> - Used under a CC BY 2.0 License

## Authorship and copyright

Text and layout: Cameron Neylon

© International Development Research Center 2016

1. OECD’s Principles and Guidelines for Access to Research Data from Public Funding (2007),

   http://www.oecd.org/sti/sci-tech/38500813.pdf [↑](#footnote-ref-1)