Designing a Data Management Plan: the GLOBAL-RURAL experience

PROFESSOR MICHAEL WOODS
ABERYSTWYTH UNIVERSITY
m.woods@aber.ac.uk
GLOBAL-RURAL

ERC Advanced Grant
2014-2019

Examining how globalization is reproduced through rural localities and how rural agents respond to opportunities and challenges

9-strong team led by Professor Michael Woods at Aberystwyth University, UK

Research across 4 work packages and case studies in 12 countries

Mixed methods approach

Website and blog: www.globalruralproject.wordpress.com
Types of Data

- Publicly available secondary quantitative data (e.g. census statistics)
- Commercially obtained secondary quantitative data
- Newly generated quantitative data (e.g. survey data)
- Interview data (recorded and transcribed)
- Interview data (not recorded)
- Focus group data
- Observational / ethnographic notes
- Archival data (historical documents, press articles etc)
- Visual data (photographs, video)
- Participant generated data (written, online etc)
Data Sensitivities

<table>
<thead>
<tr>
<th>Individual lifestyle data</th>
<th>Commercially purchased data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual life course data</td>
<td>Data from semi-public spaces and forums</td>
</tr>
<tr>
<td>Business data</td>
<td>Third party data</td>
</tr>
<tr>
<td>Information/opinion on controversial issues</td>
<td>Moving data internationally</td>
</tr>
<tr>
<td>Georeferenced data</td>
<td>Data collected in non-democratic states</td>
</tr>
</tbody>
</table>
Context

Data protection and privacy laws in UK and case study countries
ERC policies
University policies
Research institute protocols
Gatekeeper/partners’ policies
Guarantees made to research participants

Advice from ERC Ethics Advisor, University Ethics Advisor, University Data Protection Officer, Independent Ethics Advisor, colleagues and fellow researchers
Context

Start of grant coincided with development of new Research Data Management policy by university
Research Data Management (RDM)

What?
What is RDM?
What is research data?

Why?
Why should you manage your research data?

Who?
Who requires RDM?
Who can help you with RDM?

When?
When should you manage your research data?

How?
How to manage your research data
How to share your research data

Where?
Where can you store your data?
Where to find data produced by Aberystwyth University
Resources

https://dmponline.dcc.ac.uk/
Resources

Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, select here to write a generic DMP based on the most common themes.

- **If applying for funding, select your research funder.**
  - European Commission (Horizon 2020)
  - Otherwise leave blank.

- **To see institutional questions and/or guidance, select your organisation.**
  - Aberystwyth University
  - You may leave blank or select a different organisation to your own.

- **Tick to select any other sources of guidance you wish to see.**
  - DCC guidance
**Resources**

An initial DMP should be completed within 6 months of starting the project.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

The DMP is not a fixed document, but evolves during the lifespan of the project.

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

<table>
<thead>
<tr>
<th>Sections</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>For each data set specify the following:</td>
<td>- Data set reference and name</td>
</tr>
<tr>
<td></td>
<td>- Data set description</td>
</tr>
<tr>
<td></td>
<td>- Standards and metadata</td>
</tr>
<tr>
<td></td>
<td>- Data sharing</td>
</tr>
<tr>
<td></td>
<td>- Archiving and preservation (including storage and backup)</td>
</tr>
</tbody>
</table>
Resources

http://www.dcc.ac.uk/resources/data-management-plans
# Resources

http://www.dcc.ac.uk/resources/data-management-plans

<table>
<thead>
<tr>
<th>Questions to consider:</th>
<th>Responsibilities and Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the risks to data security and how will these be managed?</td>
<td>Who will be responsible for implementing the DMP, and ensuring it is reviewed and revised?</td>
</tr>
<tr>
<td>How will you control access to keep the data secure?</td>
<td>- Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?</td>
</tr>
<tr>
<td>How will you ensure that collaborators can access your data securely?</td>
<td>- Who will be responsible for each data management activity?</td>
</tr>
<tr>
<td>If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?</td>
<td>- How will responsibilities be split across partner sites in collaborative research projects?</td>
</tr>
<tr>
<td>Guidance:</td>
<td>- Will data ownership and responsibilities for REDCap be part of any consortium agreement or contract agreed between partners?</td>
</tr>
<tr>
<td>If your data is confidential (e.g. personal data not already in the public domain; confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with (e.g. ISO 37001).</td>
<td>Guidance:</td>
</tr>
<tr>
<td>- Consider how the data may be used e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for sharing and preservation.</td>
<td>- Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving &amp; data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions to consider:</th>
<th>What resources will you require to deliver your plan?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What data must be retained/destroyed for contractual, legal, or regulatory purposes?</td>
<td>- In addition to specialist expertise (or training for existing staff) required?</td>
</tr>
<tr>
<td>How will you decide what other data to keep?</td>
<td>- Do you require hardware or software which is additional to or exceptional to existing institutional provision?</td>
</tr>
<tr>
<td>What are the reasonable research uses for the data?</td>
<td>- Will changes be applied by data repositories?</td>
</tr>
<tr>
<td>How long will the data be retained and preserved?</td>
<td>Guidance:</td>
</tr>
<tr>
<td>- Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.</td>
<td>Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions to consider:</th>
<th>Are any restrictions on data sharing required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>How will you share the data?</td>
<td>- What action will you take to overcome or minimise restrictions?</td>
</tr>
<tr>
<td>- Will you share data via a repository, handle requests directly or use another mechanism?</td>
<td>- For how long do you need exclusive use of the data and why?</td>
</tr>
<tr>
<td>- When will you make the data available?</td>
<td>- Will a data sharing agreement (or equivalent) be required?</td>
</tr>
<tr>
<td>- Will you pursue a persistent identifier for your data?</td>
<td>Guidance:</td>
</tr>
<tr>
<td>Guidance:</td>
<td>- Outline any expected difficulties in sharing data with acknowledged long-term value.</td>
</tr>
<tr>
<td>Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the topic, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the use of your data.</td>
<td></td>
</tr>
</tbody>
</table>
Key Issues

Data Collection

What data will be collected?

- Primary and secondary qualitative and quantitative data
- Data collected primarily concerns organizations, relations, policies and events
- Specified where individual personal data would be collected
- How personal data might be used and conditions of use
Key Issues

Data Collection

How will data be collected?

Data collected through interviews, archival and library research and field observation

Some data collected through participatory methods, e.g. workshops, interactive walks, participatory GIS

Participation by informed consent

Specific consent obtained for photography involving individuals or on private property
Key Issues

Data Collection

How will data be recorded?

- Interviews and workshops digitally recorded, with participant’s consent, and professionally transcribed
- If consent for recording not given, manual notes will be taken
- Data recording in archives and libraries will follow institutional regulations and may involve photographing or scanning of documents and/or manual note-taking
Key Issues

Data Transfer

How will data be transferred from the field site?

Interview recordings and notes containing personal or sensitive information will be stored during fieldwork in password-protected or encrypted files on a laptop or tablet computer and backed-up on a password-protected pen-drive or external hard drive.

Transfer of interview recordings for transcription?
Key Issues

Data Processing

How will confidentiality and privacy be protected?

Personal data will be anonymized or disguised as soon as possible after collection.

Use of pseudonyms

*Recording of appropriate metadata*

*Good practice in naming files*

*Delinking anonymized and non-anonymized data*
Key Issues

Data Storage

Where will data be stored?

Data, including interview transcripts and all associated documentation that could identify participants, will be stored on a secure drive on the departmental server.

Negotiation for appropriate secure driver space with university.

University policy required data to be stored on a server physically located in the EU.
## Key Issues

### Data Storage

| Who will have access? | Members of the GLOBAL-RURAL team
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Subsequently modified with permissions at folder level</em></td>
</tr>
</tbody>
</table>

| How will data be backed-up? | Automatic back-up on mirror server by university
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Additional back-up of some data on password protected or encrypted pcs and external hard disk drives</td>
</tr>
</tbody>
</table>
Key Issues

Data Retention & Sharing

What data will be retained?

Anonymized processed data, for example in interview transcripts and databases.

Transcripts and other data where effective anonymization is not possible will be with-held from archiving

Encouraged to think more expansively by Independent Ethics Advisor
Key Issues

Data Retention & Sharing

Where will data be archived?
UK Data Archive

But – evolving environment. Other possibilities?

Who will have access to archived data?
Anonymized transcripts and other data will be made available for use by other researchers, except in cases where consent has been withheld by the participant
Key Issues

Data Retention & Sharing

What data will be destroyed and when?

Raw data from which research participants may be identifiable, including recordings of interviews, field notes etc, will be retained for the duration of the project and no more than 12 months after the end.

These materials will be destroyed no later than 12 months after the end of the project.

But – compliance with university policy?
Specific Issues

Data from ‘non-conventional’ methods
Includes participatory methods, participant generated data (e.g. uploads to websites, social media data etc
Issues of consent, recording and anonymization, intellectual property and data ownership

Data from work with schools
Compliance with two sets of policy
Procedures for informed consent
Intellectual property and data sharing
Specific Issues

GIS analysis and mapping of data
Appropriate georeferencing
Safeguards that locational data does not compromise anonymity

Data from research in ICPC countries (Brazil, China, Liberia, South Africa and Zambia)
Attention to local legal regulations and cultural sensitivities
Advice and support from local partners
Role of field assistants and interpreters
Balancing security of data with local legal compliance
Exported data not to include any personal data or data with sensitive commercial or political applications

What data needs to be collected and what needs to be recorded?
Reflections

Make use of tools and resources to support preparation of data management plans

Take advice widely and particularly on technical issues

Avoid compromising the potential of research for easy options

Look for creative solutions

A data management plan is an evolving framework

A good data management plan can improve the research process and outcomes