

Data Management Plan for the "Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights" project.

Project Name IKS & IPR Climate Change Project

Project Identifier 107650 (IDRC-OCSD Sub-Grant 2)

Principal Investigator / Researcher Cath Traynor

Description The 'Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights' project examines processes of open and collaborative science related to indigenous peoples knowledge, climate change and intellectual property rights. It assumes and challenges practices of open science as a process, one that should involve modes of being both open and closed. The project takes history into account when considering how indigenous peoples' are producing knowledge related to climate change and how such knowledge maybe characterized as indigenous peoples' intellectual property and/or impacted by dominant intellectual property regimes. The central questions the research is addressing are: (i) How is climate change impacting indigenous Nama and Griqua communities? (ii) How are these communities producing indigenous knowledge related to addressing climate change and offering alternative strategies? (iii)How do indigenous Nama and Griqua characterize their knowledge as indigenous intellectual property (or not) and decide to openly share their knowledge (or not) internally or with the outside public? (iv) How and what types of laws and policies (including intellectual property rights) promote and/or hinder these indigenous strategies and open collaboration with the public? The data are being collected and created to answer these main questions. Furthermore, the researchers are critically tracking the research process itself and this data will be scrutinized to provide information on the open and collaborative science process and dilemmas and tensions around openness issues.

Institution Portage

Data Collection

What types of data will you collect, create, link to, acquire and/or record?

Activity	Data type	Notes
Interviews with human subjects		
Interview questions and prompts	Text	
Signed Consent Form	Text	
Recording of Interview (if consent granted)	Audio	
Researcher notes	Text	
Images of interviewee (if consent granted) and environment/surroundings;	Images	

Analysis of interviews		
Transcription of audio	Text	
Coding of interview text software type (i) Excel	Tabular data	Excel spreadsheet
Coding of interview text software type (ii) Nvivo	Tabular data	NVivo Software
Research-related correspondence		
Researcher – University correspondence	Text	Email correspondence
Researcher – University correspondence	Tabular	IU Ethics software
Researcher – University correspondence	Text	Researcher notes
Researcher – Researcher correspondence	Text	Email correspondence
Researcher – Researcher correspondence	Text	Researcher notes
Researcher-Researcher Team meetings	Tabular	Record of team meetings
Researcher-Researcher Feedback Loop Sessions & M&E	Audio	Recording of discussion
Researcher-Researcher Feedback Loop Sessions & M&E	Text	Researcher notes
Researcher-Researcher Feedback Loop Sessions & M&E	Images	
Researcher – Indigenous peoples correspondence	Text	Email correspondence
Analysis of researcher-related correspondence		
Coding of interview text and correspondence (i) Excel Software	Tabular data	Excel spreadsheet

Coding of interview text and correspondence (ii) Nvivo Software	Tabular data	NVivo Software
Community meeting and peer-to-peer learning		
Meeting/activity resources	Text & tabular data	Programme, participants list,
Meeting documentation	Text	Researcher notes
Recordings of discussion (if consent granted)	Audio	
Photos of activities (if consent granted)	Images	
Analysis of meetings/activities		
Transcription of audio	Text	
Coding of interview text software type (i) Excel	Tabular data	Excel spreadsheet

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Type of data	File format	Allow for:		
		Re-use	Sharing	Long-term Access
Text (Qualitative)	Rich text format (.rtf)	Y	Y	Medium-term - widely used propriety format e.g. MS Word (.docx)
Documentation and scripts	Rich Text Format (.rtf), PDF (.pdf) & HTML (.html)	Y	Y	Medium-term - widely used propriety format (MS).
Tabular (Qualitative)	Excel (.xlsx)	Y	Y	Medium-term - widely used propriety format e.g. MS Excel (.xlsx)
Tabular (Qualitative)	NVivo (?)	Y	Y	Y - as long as access to proprietary software
Audio (Digital)	MP3 Format Sound (.mp3)	Y	Y	Yes: Created as MPEG-1 Audio Layer 3 (.mp3)
Images (Digital)	JPEG Image (.jpeg)	Y	Y	Yes: ascreated in JPEG format

What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

- Files will be uniquely identified using a systematic naming convention; e.g. OCSD_Team_Meetings
- The date will be recorded in the file name; e.g. OCSD_Team_Meetings_2016_05_11
- Version description will be added to the file name (where relevant); OCSD_Team_Meetings_2016_05_11_Final
- Only current versions of files will be used between researchers – older versions will be stored in ‘archive’ folders.
- Human subject interview audio, transcription, and coding file names will be coded using a systematic naming convention: e.g. CC_001, CC_002;

- A Master Sheet for interviews will record relationships between, audio, transcription and coding files;

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future?

The following documentation will be needed for the data to be read and interpreted correctly in the future:

- Reserach proposal;
- Reserach methodology;
- Format and file type of data;
- Data capture and collection methods;
- Data coding and analysis information;
- Record of reserachers and for each task performed - identity of reseracher(s);
- To be 'interpreted correctly' a political-ecological approach is required to the data, so that the relationship between climate change, indigenous peoples and IPR can be undertdood, furthermore, when interpreting the data the histories of the indigenous Nama and Griqua peoples needs to be taken into account.

How will you make sure that documentation is created or captured consistently throughout your project?

Documentation will be at (i) study level, and (ii) data-level.

- Study-level:
 - Context of data collection: Information of OCSDNet & IDRC Data Sharing programme, project history and project proposal;
 - Summary of data collection methods;
 - Data files: structure, number of records, relationships between files;
 - Data sources: provenance of interview audio transcribed files;
 - Data validation procedures;
 - Modification history (if any);
 - Interview data: information on regards privacy, confidentiality, on/off-record, etc. and institutional ethics requirements, and researcher protocols regards sharing data;
- Data-level:
 - Data documentation will be embedded in the data;
 - A Master Sheet will be developed to summarise data documentation (e.g. names, labels, descriptions; For analysis – codes; For interviews – data list describing individuals interviewed – conforming to ethics requirements);

If you are using a metadata standard and/or tools to document and describe your data, please list here.

If through the IDRC Data Sharing study, research data is identified as suitable for sharing, then the possibility of attaching metadata for online data catalogues and discovery purposes will be considered.

Storage and Backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

The anticipated storage requirements are:

- 10 GB for audio data;
- 40 GB for images;
- 10 GB for all other data (text, tabular, PDFs, etc);
- Plus back-up (60 GB).

How and where will your data be stored and backed up during your research project?

The 3-2-1 back-up rule will be used;

Three copies of the data:

1. Stored on Project Manager's laptop compute (fixd);
2. Stored on Project Manager's institutional Google Drive account (cloud-based);
3. Stored on Academics principle investigators institutional 'Box' account (cloud-based).

Two different storage media:

1. Fixed: Project Manager's laptop and external hard drive;
2. Cloud-based: Seperate institutional Google Drive and 'Box' Account;

One back-up off-site:

1. Back-up off site from institutional offices: Academic 'Box' account.

Note: Storage and back-up need to adhere to University Ethics agreements and for human subject research data, (e.g. interview audio, text, and coding), to agreements in individual interviewee consent forms. In our case, physical copies stored in a locked filing cabinet that only designated persons have access to. Password-protected computers and cloud accounts.

How will the research team and other collaborators access, modify, and contribute data throughout the project?

The 3 principal researchers will modify and contribute data throughout the project through different means regards whether data is sensitive (human subject data and related analysis) or not;

1. Sensitive data (e.g. interview audio's, transcripts, coding), shared through academic 'Box' account and institutions Google Drive;
2. Other data: shared through institutional 'Box', and Google Drive accounts and limited sharing through emails.

Preservation

Where will you deposit your data for long-term preservation and access at the end of your research project?

In our submission for ethics clearance to the University of Cape Town Research Ethics Committee (REC) - which has been approved for the full duration of the project - we indicated that we planned to preserve the data after the project completion date (31 December 2016).

In the short-term, the data will be preserved as-per storage guidelines.

In the long-term, as the interview data with indigenous peoples' may contain IK which could be of lasting value, possible data repository or archives will be investigated. Nationally in South Africa, the National Recordal System (NRS), which has been specifically developed to record, document, preserve and protect indigenous knowledge in South Africa. The NRS includes a cyber infrastructure that aims to provide authentic, legal and secured IKS information. The issues regards depositing the data in the long-term on such a system will need to be evaluated, and considerations will need to include consent from interviewees, and impact of the 'Community-Reseracher' Agreement and consent from the community representatives, in this case the Traditional Leader and associated decision-making structures. Furthermore, the Principle Researchers would also need to agree to the form of data made available, e.g. raw data, audios, transcripts, or selected 'IK' elements, etc.

Preservation opportunities of especially the "IK" element of the data could also be investigated within existing community governance structures, and barriers and challenges (if any) discussed.

An important ethical issue regards preservation of the "IK" component relates to the Griqua and Nama community control over, and access to the data - wherever it may (or may not) be preserved.

Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

If it is decided that data are preserved for the long-term, a process will need to be developed for interview data, to ensure that is is, anonymous (to ensure confidentiality as per the interviewee consent form); furthermore, to ensure privacy as per the interviewee consent form, as privacy includes sharing oneself intellectually and is in the eye of the interviewee (not the researchers or the IRB/REC), this implies that interviewees would need to be consulted regards whether the interview data, be that unedited/edited audio, notes, transcripts, and coding analysis meet their privacy requirements. To ensure confidentiality, if identifiable data is to be preserved, then there would need to be clarity on who has access to this data and also measures to maintain confidentiality developed. Additional ethical and legal requirements will need to be considered, for example SA laws and policies regards IKS.

Where it is agreed that data be preserved, the following conversions will be used to ensure that the file formats are preservation friendly:

Type of data	File format	Conversion to preservation friendly file format
Text (Qualitative)	Rich text format (.rtf)	PDF-A
Documentation and scripts	Rich Text Format (.rtf), PDF (.pdf) & HTML (.html)	PDF-A
Tabular (Qualitative)	Excel (.xlsx)	PDF-A
Tabular (Qualitative)	NVivo (?)	PDF-A (?)
Audio (Digital)	MP3 Format Sound (.mp3)	Original digital audio data were created as MPEG-1 Audio Layer 3 (.mp3) files, and these are suitable for preservation, thus, no conversion required.
Images (Digital)	JPEG Image (.jpeg)	TIFF/PNG

Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

The default for the project is sharing of data, the following key issues with regards to data provided by indigenous peoples, will need to be investigated further and clarified and where relevant consent provided, prior to any sharing taking place:

Form that data could be shared in	Individual Indigenous Interviewee
All interview data	
(i) On-the-Record	
(ii) Off-the -Record	Level of anonymity required to meet informed consent process needs to be clarified. Then data to be shared needs to be processed to remove personal identifiers.
Raw data	

Audio interview	<p>Audio data may contain ‘direct identifiers’ such as family name, home location information or similar. Audio may also contain ‘indirect identifiers’ which could be used to identify individuals. This is particularly so at the community level, e.g. interviewees provided summaries of family histories which they could be linked to. Furthermore, individuals could be identified through voice recognition by people that know them. Thus, individual interviewee consent would be required prior to sharing.</p>
Researcher notes	
Identifiable image of interviewee	<p>Prior to photographs, interviewees were asked for consent to photograph and given assurances that the photo would only be used in connection with the project. Thus verbal consent has been provided for the taking and use of photo’s <i>in relationship to the project only</i>.</p> <p>Individual consent would thus be required to share the photos further. Anonymising visual data can ‘reduce the usefulness’ of the data and is also time-consuming, thus, e.g. Data Archive UK recommend gaining the participants consent to use and share the data unaltered is recommended where confidentiality is an issue.</p> <p>However, the link that can be made between facial photos and the interview audio/transcripts/processed and analysed data needs to be considered.</p> <p>Pictures are ‘direct identifiers’ of a person’s identity – if photos are shared then it is possible that the interview data will not be anonymised fully – especially as the projects sample size is relatively small <30 Indigenous people interviewed. Furthermore, facial recognition software is available that can identify individuals by running facial photo’s through e.g. social media websites and their profiles can be located with relative ease e.g. see FindFace in Russia and report in the Guardian Newspaper.</p>
Images (non-personal)	<p>Depiction of indigenous peoples needs to be considered from an ethical perspective, e.g. dignity, equality, respect cultural sensitivities, and respect local norms and customs regards taking and using photographs.</p>

Processed data	Audio transcription files. Concerns largely as per audio files. If, transcriptions were to be shared, information would need to be anonymised fully. Following best practices, e.g. UK Data Archive , "identifiers should not be crudely removed or aggregated, as this can distort the data or even make them unusable. Instead pseudonyms, replacement terms or vaguer descriptors should be used. The objective should be to achieve a reasonable level of anonymisation, avoiding unrealistic or overtly harsh editing, whilst maintaining maximum content'. Best practices would need to be followed and an 'anonymisation log' developed.
Analysed data	Here, the results of 'coding in grounded theory'. Key issues relate to consent provided by interviewees and privacy and confidentiality issues.
Final data	

An important issue is that interviews with indigenous peoples may contain Indigenous knowledge (IK), IK is held by the owners of that knowledge, and IK may be individually and/or collectively held.^[1] The Griqua community in Vredendal and the Nama community in Kuboes are both part of national umbrella organisation in South Africa, they regard their IK as being held collectively (L. Jansen pers comm. 01.06.2016). Therefore, in such cases, community consent would be required to share all forms of the data from the research project. The process to gain community consent would be through the communities customary procedures and furthermore, it would need to be clarified whether consent would be required from the larger umbrella organisation.

The research project takes a political-ecological approach, with an emphasis upon contextualising the research approach, methodology and data in the socio-political and historical contexts of the communities we partner. Given, this approach attempting to extract the 'IK elements' from the 'non-IK elements' interviews would be inappropriate.

^[1] For example see the [Ottawa Traditional Knowledge Principles](#)

Have you considered what type of end-user license to include with your data?

The Sub-Grantee organisation has a Grant Agreement with iHUB LIMITED. Clause 8 Intellectual Property Rights states under section 8.1.4. Data and Report Sharing. The Sub-Grantee shall make the data collected in the course of the Project available through the OCSDNet online repository, on terms and conditions to be agreed upon between parties, acting reasonably.

However, regards human subject data derived from interviews, the Informed Consent Document the researchers have signed with those consenting to be interviewed, under confidentiality, states that the interviewees have "...the right to review the recordings that I make of your interview. You can ask me to change or erase some of the things you said or showed me." Furthermore, the consent form also states that the researchers will "...make

sure that all written documents regarding our conversation are kept confidential." and that only co-investigators will have access to them.

In order to share the human subject interview data, the researchers would need to return to each individual interviewee and negotiate what (if any) aspects of the data (audio, images, transcript, coding) could be shared publicly, and if so, a new agreement signed. Then, as per section above, the community governance structures would need to be approached for their consent to share the communally held IK. Thus, both individual and collective elements of IK need to be addressed. Furthermore, it would also need to be identified if any secret, or sensitive information had been shared, and if so the individuals and the community decide whether this should not be shared, or whether they want to share it (if the latter, the risks and implications of sharing will need to be thoroughly assessed and understood by all community members.)

Then as per the Grant Agreement, materials agreed to be shared would be upload onto the OCSDNet online repository. The Grant Agreement specifies that CC licence be used for 'project outputs and research findings' - thus, this would be the first end-user licence to be considered. The individual interviewee and the community would need to agree to an end-user licence.

The researchers are developing a Community-Researcher Contract, currently who will hold IPR is being negotiated. IPR holders are not clear, as the individual/community should hold the IPR to IK exclusively. The research outputs (e.g. academic papers, book chapters) will contain new knowledge created by the researchers, (specific publications may contain specific elements of IK), however, all of these outputs will have been developed with recourse to IK and interactions with the community and conversations around these issues.

What steps will be taken to help the research community know that your data exists?

This section will be addressed if it is agreed that IK can be shared.

Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

During the project, the project Manager will be responsible for managing the data.

The Principle Investigator at one of the partnering academic institutions will be responsible for managing the Box account (copy of crucial data).

Post-project, the project's ethics approval submission states that:

"...we plan to not discard of the research data once the project is completed as we are not dealing in sensitive health data or biological samples. The nature of our study is socio-behavioral and thus low risk in terms of the data collected. Storing the collected research data can be useful, among other things, for future research activities and/or funding applications. We shall safeguard collected data in the following manner:

Notes, recordings, transcripts, documents, and other data will be shared between principal investigator, co-investigators, and collaborators from the indigenous community via email, flash drive, and the institutional Box (located through one of the participating

universities). Hard copies of transcribed interviews and notes will be stored in locked file cabinets in locked offices at the three participating institutions. The principal investigator and co-investigators are the only persons who will have access to the data and codes. Only the principal investigator, co-investigators, a transcriber, and a translator will primarily see the data. Participant members from Indigenous communities as well as graduate and undergraduate researchers may also have access to the data, but only for the purposes of analysing the data.

The consent form will make it clear that the findings and data from the project will be publicly available and only in a form that will protect the identity of the subject, if they have requested "off-the record" sharing of their information.

The researchers will discuss with community representatives, the long-term retention and storage of the data within the participating communities, and capacity and resources to achieve this.

How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

The project is more than half-way through and no substantive changes in personnel are foreseen. However, if the current Project Manager departed the post, the organisation with whom the Grant Agreement is signed, in this case a Non-Profit Organisation (NPO) would be responsible for appointing a Project Manager to manage the data.

What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

The key resources are:

- Computers, 1 external hard drive (USD 110.00), internet access and cloud access & data bundles;
- Filing cabinet with lock to maintain hard copies of data (USD 150.00);
- Stationary: folders, pockets, etc;
- Printer, photocopy and scanner;

Ethics and Legal Compliance

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

The research includes sensitive data:

- Human subject research data derived from interviews;
- Indigenous knowledge;

During the project the following data management strategies will be applied:

To ensure confidentiality: Our research ethics submission states that: Hard copies of

transcribed interviews and notes will be stored in a locked file cabinet. The file cabinets are located in locked offices at the three participating institutions.

The principal investigator and co-investigators are the only persons who will have access to the data and codes. Only the principal investigator, co-investigators, a transcriber, and a translator will primarily see the data. Participant members from Indigenous communities as well as graduate and undergraduate researchers may also have access to the data, but only for the purposes of analyzing the data. The addition of graduate and undergraduate researchers will be added as study personnel at a later date via an amendment.

In order to protect confidentiality and subject's reputation, participants will have the option of having their interview data be "on the record", meaning that they can be cited by name, or "off the record," meaning that they will not be cited by name or in any other way identified. All prospective study participants may refuse to be interviewed, refuse to answer any particular questions, and may ask to have any comments be handled "off the record".

Consent forms will be used to communicate the risks and benefits of the study and procedures for safeguarding privacy and confidentiality of data."

Registered professionals (members of the Professional Editors' Guild - PEG) will be used for Transcription, and they will sign a 'Non-Disclosure Agreement' with the Project Management organisation.

Students who will participate in PAR and/or handle interview data will also sign a 'Non-Disclosure Agreement' with the NPO, and to conform with the institutional review board (IRB) and/or research ethics committee (REC) requirements of the partner universities, they will sign a 'Non-Affiliated Investigator Agreement' and also pass the CITI Program Human Research, Social/Behavioural researchers" course (an ethics requirement of the partner university based in USA).

Problems anticipated in sharing data:

1. Privacy: Interview Informed Consent Document states that "Efforts will be made to keep your information private and confidential." As privacy is from the interviewee's perspective, then this would need to be clarified on an individual basis.
2. Confidentiality: The Informed Consent Document states that "I will keep personal information about you secret and confidential." If interviews are "off the record" then all personal identifiers would need to be removed prior to sharing.
3. IK issues. currently consent at the individual level is clear, but at the community level it has not yet been agreed - this needs to be clarified prior to sharing.
4. Intellectual property right (IPR) issues. The Informed Consent Form does not cover IPR issues. The community-researcher contract refers to IPR - it is currently under negotiation.
5. The Sub-Grantee Agreement also refers to IPR.
6. The NPO managing the project has signed Memorandum of Agreement with individual researchers in the partner universities. These agreements state the partner shall respect IPR and that any intellectual property of the communities with which the 'Media and Reserach Contract' have been signed as part of this project, will be subject to the United Nations Declaration on the Rights of Indigenous

Peoples (UNDRIP) and South African intellectual property rights legislative and policy frameworks.

7. IPR issues between community-researchers, NPO-universities research partners, and NPO-Funder, need to be compatible.

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Question not answered.

How will you manage legal, ethical, and intellectual property issues?

The community-researcher agreement is the tool we hope will clarify legal and IPR issues between the community level and the research institutions (NPO and two partner universities).

Regards the IK and IPR issues the NPO will seek additional professional legal input to ensure that the communities rights, under international and national laws are adhered to.

The process of negotiating the agreement is being tracked as part of the research process.

The researchers within the universities are not commenting on the contract as this would be a conflict of interest with their respective institutions.

This document was generated by DMP Assistant (<https://assistant.portagenetwork.ca>)