



studio.cul.columbia.edu

DEVELOPING AN INSTITUTIONAL RESEARCH DATA MANAGEMENT PLAN: GUIDELINES

Unisa 2014 LIS Research Symposium
Research Data Management, July 24 2014

Dr Heila Pienaar

Deputy Director: Innovation & Technology
University of Pretoria Library

CONTENT

- Research data (management) cycle
- What is a data management plan?
- Benefits of a research data management (RDM) plan
- The two best-known international RDM plans
- Examples of university RDM plans
- Guidelines



Living in an Ivory Basement

Stochastic thoughts on science, testing, and programming.

misc personal python **science** teaching testing

My Data Management Plan - a satire

Dear NSF,

I am happy to respond to [your request](#) for a 2-page Data Management Plan.

First of all, let me say how enthusiastic I am that you have embraced this new field of "large scale data analysis". Ever since I started working with large Avida data sets in 1993, then with large meteorological data sets in 1995, and then again with large sequence data sets in 1999, I have seen the need for a systematic plan to manage the data. It is nice to see NSF stepping up to the plate in such a timely manner, and I am happy to comply.

Now, as to my actual data management plan, here is how I plan to deal with research data in the future.

I will store all data on at least one, and possibly up to 50, hard drives in my lab. The directory structure will be custom, not self-explanatory, and in no way documented or described. Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is.

Backups will rarely, if ever, be done.

When required to make the data available by my program manager, my collaborators, and ultimately by law, I will grudgingly do so by placing the raw data on an FTP site, named with UUIDs like 4e283d36-61c4-11df-9a26-edddf420622d. I will under no circumstances make any attempt to provide analysis source code, documentation for formats, or any metadata with the raw data. When requested (and ONLY when requested), I will provide an Excel spreadsheet linking the names to data sets with published results. This spreadsheet will likely be wrong -- but since no one will be able to analyze the data, that won't matter.

Did I mention the dick-through license? "You are provided this data for the sole purpose of reproducing our published results. Any attempt to publish your own analyses of this data will be rejected, if necessary during the anonymous review process, by pointing out all of the data cleanup steps you forgot to do correctly in your analysis. (We don't remember all of them ourselves, but there sure were a lot!) Give up now."

Mon 17 May 2010

By [C. Titus Brown](#)

In [science](#).

tags: [science](#)



‘I will store all data on at least one, and possibly up to 50, hard drives in my lab.

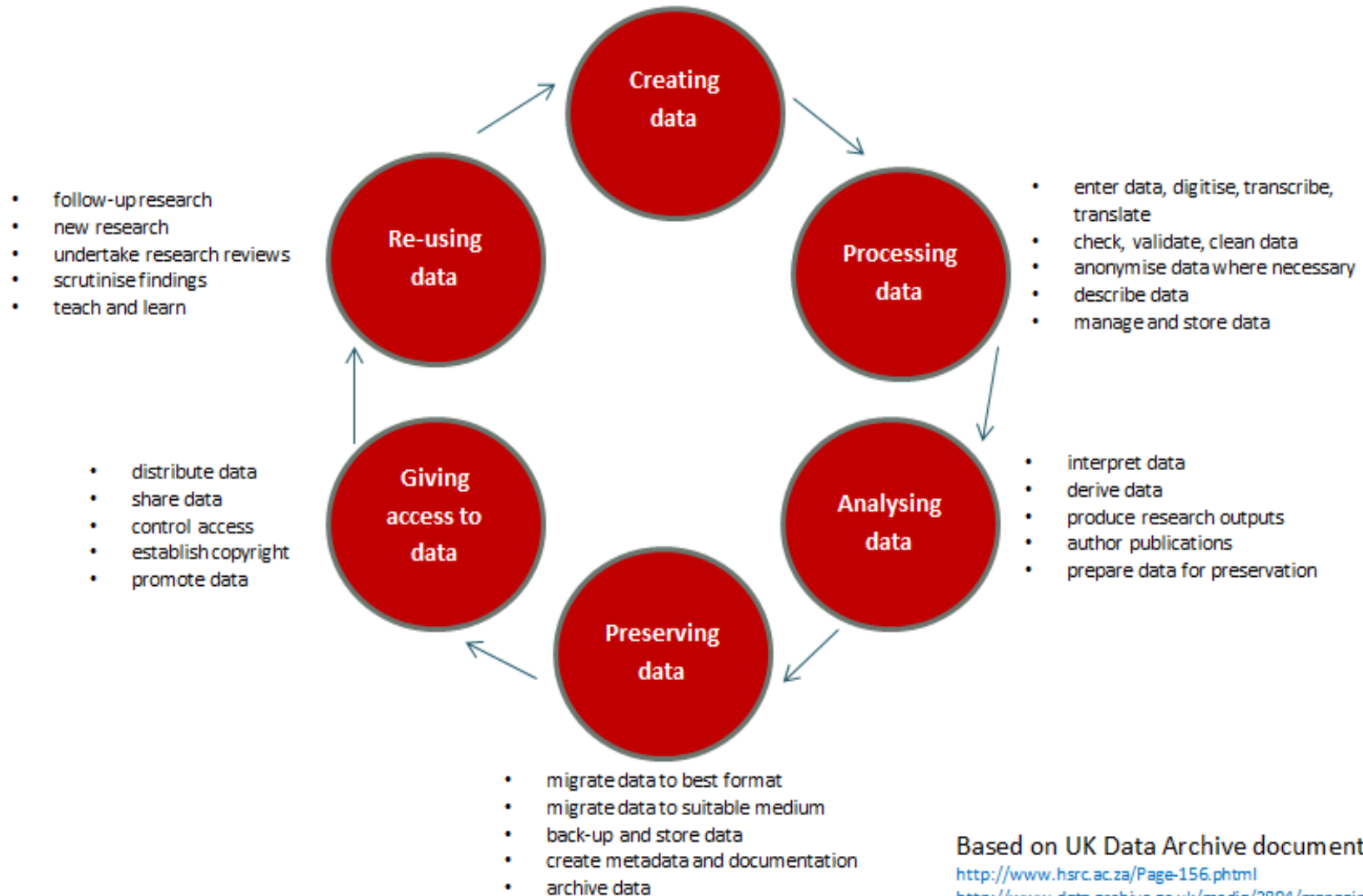
The directory structure will be custom, not self-explanatory, and in no way documented or described.

Students working with the data will be encouraged to make their own copies and modify them as they please, in order to ensure that no one can ever figure out what the actual real raw data is.

Backups will rarely, if ever, be done.’

Research Data Life Cycle

- design research
- plan data management (formats, storage etc.)
- plan consent for sharing
- locate existing data
- collect data (experiment, observe, measure, simulate)
- capture and create metadata



Based on UK Data Archive documentation:

<http://www.hsrc.ac.za/Page-156.html>

<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

<http://www.data-archive.ac.uk/>

WHAT IS A RDM PLAN?

- ⦿ A data management plan (DMP) is a **document** that describes how you will collect, organise, manage, store, secure, back- up, preserve, and where applicable, share your data
- ⦿ A data management plan should be created at the **start** of a research project so that good practices are established early

BENEFITS OF A RDM PLAN

- ⦿ Helping you to identify issues and strategies early in your research project
- ⦿ Ensuring that you have documented your compliance with institutional and funder **policies** and **ethics** approval requirements
- ⦿ Helping you think about **data sharing** and reuse opportunities of your research data
- ⦿ Ensuring that your research data remains useful and stored safely for **future use**

<http://www.sheffield.ac.uk/library/rdm/dmp>

THE TWO BEST-KNOWN INTERNATIONAL RDM PLANS



The screenshot shows the DMPTool website homepage. At the top left is the logo for DMPTool, which consists of three green squares of varying shades to the left of the text "DMPTool". Below the logo is the tagline "Guidance and Resources for your Data Management Plan". To the right of the logo are links for "Contact Us", "Get Started", and "Login". Below the logo is a navigation menu with links for "Home", "About DMP Tool", "DMP News", "My Plans", "Funder Requirements", and "Help". The main content area features a large image of a scientist in a white lab coat and safety glasses working in a laboratory. Overlaid on this image is a text box that reads "Create ready-to-use data management plans for specific funding agencies." Below the image is a caption: "The DMP Tool allows you to: 1 2 3 4". To the right of the image is a "Get Started!" button. Below the button is a section titled "Data Management Plan: Sample Plan Created at the DataONE Best Practices Workshop - Santa Fe NM 7/2011 Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013". Below this is a "Type of this product" section with a "See a plan created with the DMP Tool" link. At the bottom of the page is a "Recent DMP News" section with links for "Support for NIH now available", "DMPTool unavailable, Thursday February 23, 2012, 5:00pm (PT)", and "More on user survey results".

DMPTool is a service of the University of California Center of the California Digital Library
Copyright © 2010-2012 The Regents of the University of California
Privacy Policy | Terms of Use | Photo Credits



DCC

<https://dmponline.dcc.ac.uk/>

<https://www.flickr.com/photos/rosefirerising/6776182890/>

DMP CHECKLIST, DCC, UK

- ◉ What data will you collect or create?
- ◉ How will the data be collected or created?
- ◉ What documentation and metadata will accompany the data?
- ◉ How will you manage any ethical issues?
- ◉ How will you manage copyright and Intellectual Property Rights (IPR) issues?
- ◉ How will the data be stored and backed up during the research?
- ◉ How will you manage access and security?
- ◉ Which data should be retained, shared, and/or preserved?
- ◉ What is the long-term preservation plan for the dataset?
- ◉ How will you share the data?
- ◉ Are any restrictions on data sharing required?
- ◉ Who will be responsible for data management?
- ◉ What resources will you require to deliver your plan?

DMP ONLINE, DCC

The screenshot shows a Windows Internet Explorer browser window displaying the DMPonline website. The address bar shows the URL <https://dmponline.dcc.ac.uk/>. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The Favorites bar contains several links, including 'DMPonline', 'Suggested Sites', 'Web Slice Gallery', 'Customize Links', 'De Rebus', 'Free Hotmail', 'Law Society of South Africa', 'Practical Legal Training', 'ProStandard', and 'RealPlayer'. The website content features the 'DMP ONLINE' logo in the top left, a navigation menu with 'Home', 'About', 'News', and 'Help' in the top right, and a main content area with a 'Welcome' message, a 'Sign in' form, a 'Sign up' section, and a 'Screencast on how to use DMPonline' video player. The footer contains contact information, copyright notices, and logos for DCC and Jisc.

DMPonline - Windows Internet Explorer

<https://dmponline.dcc.ac.uk/>

Certificate Error Google

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery Customize Links De Rebus Free Hotmail Law Society of South Africa Practical Legal Training ProStandard RealPlayer

DMPonline

Home About News Help

Welcome.

DMPonline has been developed by the **Digital Curation Centre** to help you write data management plans.

Screencast on how to use DMPonline

Sign in

Email address *

Password *

[Forgot your password?](#)

Remember me

Sign in

[Or, sign in with your institutional credentials](#) (UK users only)

Sign up

New to DMPonline? Sign up today.

Contact us | Terms of use | DMPonline previous version

© 2004 - 2014 Digital Curation Centre (DCC)

DCC Jisc

Internet | Protected Mode: Off 100%

<https://dmponline.dcc.ac.uk/>

DMPTOOL, CALIFORNIA DIGITAL LIBRARY

The screenshot shows a web browser window with the URL https://dmptool.org/user_sessions/institution. The page features a green header with the DMPTool logo and a navigation menu with links for Home, DMP Requirements, Public DMPs, News, Help, Contact Us, About, and Log In. Below the header is a grey bar with a computer icon and the text "INSTITUTION LOG IN".

The main content area is titled "Log in through your institution" and includes the instruction: "Select your institution below and you will be directed to your institutional log in page." Below this is a dropdown menu labeled "Select your institution" and a green "Next >>" button.

A note below the button reads: "If you do not see your institution in the list, please select 'Not in List' and click Next."

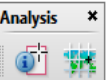
The footer contains the DMPTool logo, the text "DMPTOOL is a service of the University of California Curation Center of the California Digital Library Copyright © 2010-2014 The Regents of the University of California", and links for Privacy Policy, Accessibility Policy, Terms of Use, Contact Us, and About. There are also social media icons for Facebook, Twitter, and RSS.

https://dmptool.org/user_sessions/institution

DMPTOOL: REQUIREMENTS

The screenshot shows the DMPTool website interface. At the top, there is a navigation menu with links for Home, DMP Requirements, Public DMPs, News, Help, Contact Us, About, and Log In. Below the navigation is a search bar and a section for filtering templates by institution, with buttons for A-F, G-L, M-S, T-Z, and All. The main content is a table listing various DMP templates.

Template	Funder	Funder Links	Sample Plans (if available)
Gordon and Betty Moore Foundation	Gordon and Betty Moore Foundation	Guidelines	GBMF: Sample Plan #1 GBMF: Sample Plan #2 GBMF: Sample Plan #3
Gulf of Mexico Research Initiative	Gulf of Mexico Research Initiative		
Institute of Education Sciences (US Dept of Education)	Institute of Education Sciences (US Dept of Education)	IES Data Sharing Implementation Guide	
Institute of Museum and Library Services	Institute of Museum and Library Services	Guidance	
Joint Fire Science Program	Joint Fire Science Program		
National Institutes of Health	National Institutes of Health	Guidance	NIH: Sample Plans
National Oceanic and Atmospheric Administration	National Oceanic and Atmospheric Administration	Procedural Directive	NOAA: Sample Plans
NEH-ODH	University of Connecticut	NEH Digital Humanities Start-Up Grants	
NEH-ODH: Office of Digital Humanities	National Endowment for the Humanities	Guidelines	NEH-ODH Sample
NSF-AGS: Atmospheric and Geospace Sciences	National Science Foundation	AGS Advice	



This is an example Data Management Plan created by participants of a DataONE best Practices Workshop.

Data Management Plan

I. Products of Research

Every two days, we will subsample *E. affinis* populations growing at our treatment conditions. We will use a microscope to identify the stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook, then copy the data into an Excel spreadsheet. For quality control, values will be entered separately by two different people to ensure accuracy. The Excel spreadsheet will be saved as a comma-separated value (.csv) file daily and backed up to a server. After all data are collected, the Excel spreadsheet will be saved as a .csv file and imported into the program R for statistical analysis. Strasser will be responsible for all data management during and after data collection.

II. Data Storage and Preservation

Our short-term data storage plan, which will be used during the experiment, will be to save copies of 1) the .txt metadata file and 2) the Excel spreadsheet as .csv files to an external drive, and to take the external drive off site nightly. We will use the Subversion version control system to update our data and metadata files daily on the University of Alberta Mathematics Department server. We will also have the laboratory notebook as a hard copy backup.

PURDUE UNIVERSITY, USA

The screenshot shows a Windows Internet Explorer browser window displaying the Purdue University Research Repository (PURR) website. The browser's address bar shows the URL <https://purrr.purdue.edu/>. The website header features the Purdue University logo and the text "Purdue University Research Repository" with "PURR" in large letters. Navigation links include "Home", "Publications", "Projects", "Get Started", "Policies", and "Contact Us". A search bar is located in the top right. The main content area is titled "What is PURR?" and describes the repository as an online, collaborative working space and data-sharing platform. A "Learn More" button is provided. To the right is a video player showing a PURR logo and the URL <http://purrr.purdue.edu>. Below this, there are three columns: "Start Your Research Project" with three sub-sections (Create a Data Management Plan, Upload Research Data to Your Project, Publish your Dataset), "Featured Dataset" for "Maize grain yield record for the WQFS (1995-2012)" with a bar chart and description, and "Do you have a question?" with a "Chat with Us" form.

PURDUE UNIVERSITY | **Purdue University Research Repository** | PURR

Login Register Report a bug





Home Publications Projects Get Started Policies Contact Us Search

What is PURR?

The [Purdue University Research Repository \(PURR\)](#) provides an online, collaborative working space and data-sharing platform to support the data management needs of Purdue researchers and their collaborators.

[Learn More](#)

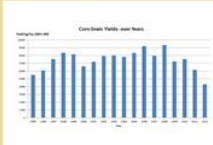
Start Your Research Project

-   **Create a Data Management Plan**
Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. [Get Started](#)
-  **Upload Research Data to Your Project**
Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. [Create a Project](#)
-  **Publish your Dataset**
Package, describe, and publish your dataset with a Datacite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long-term. [See Published Datasets](#)

Featured Dataset

Maize grain yield record for the WQFS (1995-2012)

By S.M. Brouder, N. De Armond, R.F. Turco, J.J. Volenec
Purdue University



The data included here are for the WQFS maize grain yields only; see companion publications for other data including yields of rotation crops, maize stover biomass, nitrate loss in drainage water and other environmental impact data.

Do you have a question?

Chat with Us

Send us your email question.

Enter your email address

Your Question/Message

[Send](#)

Done Internet | Protected Mode: Off 100%

You are here:

Data Management Plan (DMP)

Data management planning (DMP) is an essential element of research, and seen as a necessity by many funders. Because "sharing" or "publishing" data may not have always been an integral part of research, PURR offers several levels of assistance in data management planning. For those who have a data management plan and intend to use PURR, there is a guide to get you going. For those who are new to preparing a data management plan, here are a couple of options:

- Review the [Data Management Plan overview](#)
- Go directly to the [Data Management Planning Self-Assessment Tool](#)
- Use the [DMPTool](#) to create ready-to-use data management plans for specific funding agencies
- Contact a Libraries faculty member for consultation on discipline specific questions using the "Ask a Librarian" form below.

Watch the *Data Management Planning (DMP) Workshop!*



NOTE: If you need more information about PURR as a repository solution for data management planning, please see the links under Contact Us.

<p>Overview:</p> <p>This is where you'll find links to more details about writing your Data Management Plan.</p> <ul style="list-style-type: none">OverviewDMP ExamplesDMP Self-Assessment ToolDMP Workshop	<p>Guides and Tutorials:</p> <p>In this section, we will eventually provide tutorials to help you get more acquainted with PURR, and how to use groups, projects and much more for collaboration and publication.</p> <ul style="list-style-type: none">Boilerplate Text for DMPs planning to use PURRGuidance on developing a DMP for NSF proposals	<p>Contact Us:</p> <p>If you can't seem to find what you need, or you've noticed a bug on our site, please contact us and let us know.</p> <ul style="list-style-type: none">For questions on getting started with a grant proposal: Contact SPSFor information using PURR and data management, contact Courtney Matthews, 496-2770, courtneyearmatthews@purdue.edu, PURR Digital Data Repository SpecialistFor technical support or to report a bug: open a trouble ticket
---	--	--



Text for Your Proposal

> [GET STARTED](#)

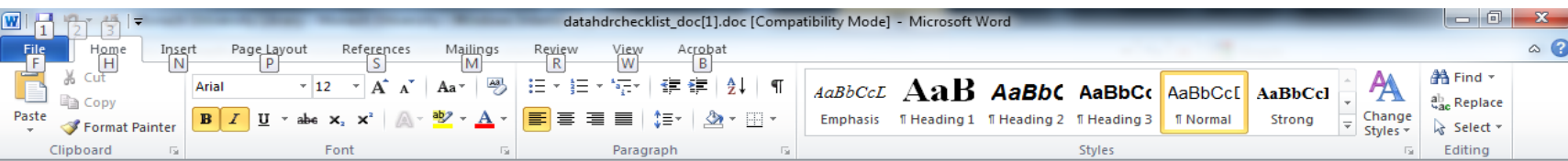
The following boilerplate text is a great way to describe PURR in your Data Management Plan:

Support for data management for this project throughout its lifecycle will occur using the Purdue University Research Repository (PURR), Purdue's institutional data repository. PURR utilizes HUBzero®, a web-mediated software platform designed for scientific collaboration and sharing of scientific data that was developed with support from the National Science Foundation and Purdue University. PURR provides workflows and tools for ingestion, identification and dissemination of data as well as services to ensure data security, fidelity, backup, and mirroring. Purdue Libraries will consult with investigators to facilitate selection and ingestion of data with the application of appropriate descriptive metadata and data standards as well as to provide appraisal of data for long-term preservation and stewardship. PURR is working towards the ISO 16363 process to become a certified Trusted Digital Repository. PURR comes with a set of default policies and functionality that addresses privacy and confidentiality, intellectual property and copyright, and access and sharing of data. Datasets published using PURR will be assigned Digital Objects Identifiers (DOIs) and will be exposed to the web using open standards to maximize discoverability and scholarly reuse of data. An allocation of resources from PURR has been reserved for this project and will be appropriated upon its award.

Suggested citation for proposal bibliography:

Purdue University. (2012). Purdue University Research Repository. Retrieved from <http://purrr.purdue.edu>

MONASH UNIVERSITY, AUSTRALIA



Research Data Planning Checklist (Higher Degree by Research Student)

Name

Faculty / division / unit

Contact details

Title / description of the Higher Degree by Research (HDR) project

A. OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP)

Copyright protection

- The data is protected by copyright.
This will apply to most research data.
- The data will be collected, created or compiled
 - in Australia - Australian copyright applies.
 - outside of Australia.

Ownership of copyright and IP

- The copyright and other IP in the data is owned by:
 - the Higher Degree Research Student
Research by Monash HDR student in the normal course of study, which does not fall into any of the other categories below.
 - Monash University
I have assigned IP to the University because it falls into one of the categories prescribed under the Statutes and Regulations.
 - Monash University (joint ownership)
Research conducted by Monash in collaboration with another HDR student

KEY DOCUMENTS ON THIS TOPIC

[Research data management guidelines: ownership, copyright and IP](#)
[Intellectual Property Framework](#)
[Statute 11.2 IP and Copyright](#) and [IP Regulations](#)
[Explanatory Memorandum for IP Statute and Regulations](#)
[Copyright at Monash website](#)
[Practical Data Management: A Legal and Policy Guide \[national guide\]](#)

Consult the Copyright Advisers or University Solicitors.

MIGR Handbook Chapter 6: Intellectual Property - [Assignment and Licensing](#)
Provide a copy of [MIGR IP and Assignment Forms](#) to help clarify ownership of the data.

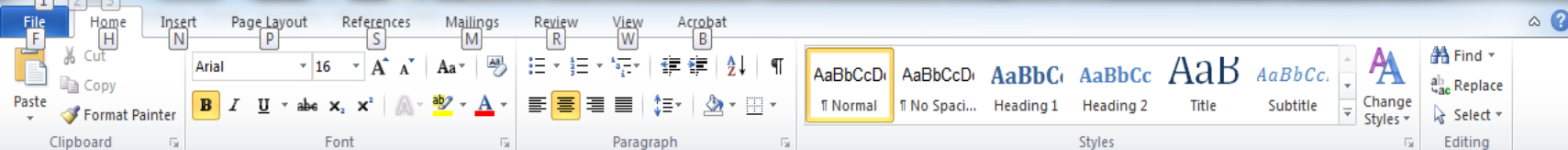


Data sets

The HSRC Research Data Service provides a digital repository facility for the HSRC's research data in support of evidence based human and social development in South Africa and the broader region. Access to data is dependent on ethical requirements for protecting research participants, as well as on legal agreements with the owners, funders or in the case of data owned by the HSRC, the requirements of the depositors of the data. We facilitate data use by preparing comprehensive metadata and disseminating data and related documents to appropriate target audiences. Data sharing is subject to an End User License agreement. Data sets from the following projects have been curated and are available for use:

- **A comparative analysis of self-identifying Gender Non-Conforming youth against the realities of Gender Non-Conforming youth in Cape Town, South Africa (NCG)**
The study, funded through the International Congress of Psychology, provides a comparative analysis of self-identifying gender non-conforming youth residing on the urban peripheries of Cape Town, against the realities of gender non-conforming youth in more affluent, resourced communities in Cape Town, South Africa.

The study explored the dynamics that make for different possibilities to express gender non-conformity among youth, as well as the opportunities available for these young people to recreate their identities in ways that make sense to themselves, and allow for a more fulfilling sense of self in contemporary South Africa.
- **A longitudinal view of unemployment in South Africa (UNEMPL)**
Global unemployment has risen in the past few years. Spatial data are required to address the problem effectively. South African unemployment literature focused mostly at a national level of spatial analysis. Some literature refer to spatial aspects that affect unemployment trends, but does not assign a location, e.g. a suburb or municipality. The research was conducted to obtain an understanding of geographical unemployment changes over time. The objectives were to provide an overview of South African unemployment spatial data sets, assess the usefulness of unemployment data for the geographical analysis of unemployment and recommend data sets that should be used for the spatial analysis of unemployment.
- **Agrarian Reform and rural poverty reduction (AR)**
Phase one of the project 'Agrarian Reform and Poverty Reduction' produced a comprehensive status report on this topic, including a review of the conceptual approaches and methodological best practices to guide further empirical research. This desktop synthesis of the literature identified prominent mechanisms or pathways through which reforms are likely to interact with poverty and to define meaningful indicators of human well-being as measures of impacts.



Human Sciences Research Council
 Lekgotla la Dinyakisišo tša Se mahlale tša Setho
 Raad vir Geesteswetenskaplike Navorsing
 Umkhandlu Wezokucwaninga Ngesayensi Yesintu
 Ibhunga Lophando Ngenzulu-Lwazi Kantu

Research Data Management Centre

Research Data Management Plan

Please supply the following information about how the sharing of the research data¹ related to the project will be done. Provide the completed form as part of the submission of the ethic application to the HSRC Research Ethics Committee.

Note: For clarity about the required information, please refer to the footnotes provided.

1	Study / collection title	
2	Person who will be responsible for the data	
3	Research data files	
3.1	What will be deposited for preservation?	
	Select one or more of the data types indicated below.	Give a short description of the contents of the data files.
		Indicate the electronic file format, e.g. <i>.sav</i>
	Quantitative tabular data	
	GIS and CAD data	
	Qualitative data	
	Digital image data	
	Digital audio data	
	Digital video data	
	Other If other, please describe.	

GUIDELINES

- ◉ Do not try to create a data management plan from scratch: many good examples are available
- ◉ Use current examples to co-create with role players a data management plan framework with minimum requirements for your institution
- ◉ Each research group can adapt the plan as long as the minimum requirements are included
- ◉ Plan should be short, easy and available online
- ◉ Must link with or include ethical requirements
- ◉ Must form an integral part of research planning & process