

Shepherding data on the way to long term stewardship:

Research Data Management at the University of Pretoria: a case study

By Mr Johann van Wyk (UPLS, UP); Prof Michael S. Pepper (Institute for Cellular and Molecular Medicine, UP), Mr Isak van der Walt (UPLS, UP); Ms Candice Honing (Institute for Cellular and Molecular Medicine, UP)

**Presented (+ edited) by Dr Heila Pienaar and Isak van der Walt,
University of Pretoria Library Services, 5 June 2014, 12thSAOIM**



Wordle of the data management policy of the Office of Digital Humanities, National Endowment for the Humanities



Definitions

- **Research data:** Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results <http://www.ed.ac.uk/is/data-management>
- **Research data management:** “is the process of controlling the information generated during a research project. Any research will require some level of data management, and funding agencies are increasingly requiring scholars to plan and execute good data management practices”. Managing data is an integral part of the research process. How data is managed depends on the types of data involved, how data is collected and stored, and how it is used - throughout the research lifecycle.
http://www.libraries.psu.edu/psul/pubcur/what_is_dm.html

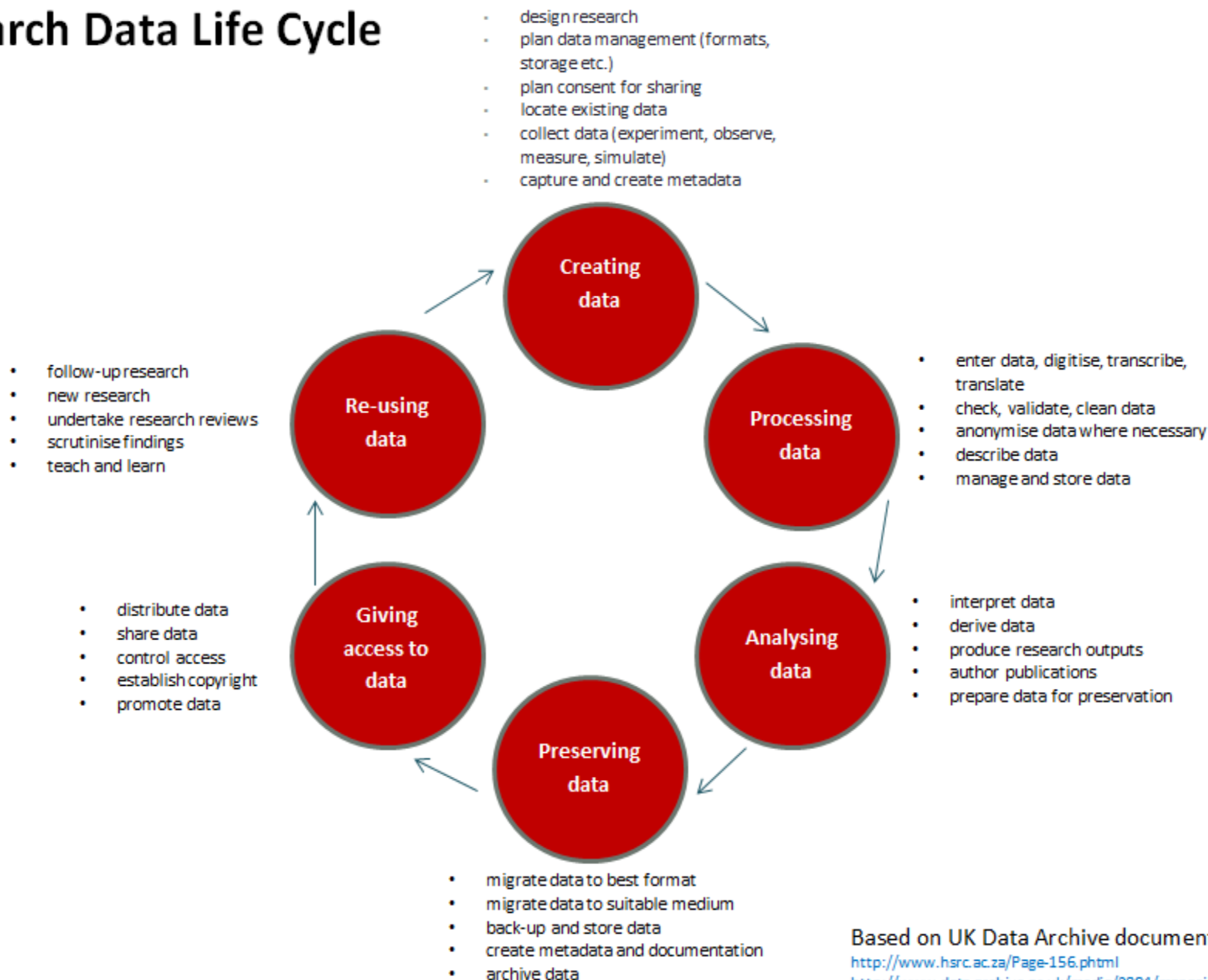
Why manage research data?

By managing research data you will:

- Meet funding body grant requirements;
- Ensure research integrity and replication;
- Ensure research data and records are accurate, complete, authentic and reliable;
- Increase your research efficiency;
- Save time and resources in the long run;
- Enhance data security and minimise the risk of data loss;
- Prevent duplication of effort by enabling others to use your data;
- Comply with practices conducted in industry and commerce; and
- Protect the University from reputational, financial and legal risk.



Research Data Life Cycle



Based on UK Data Archive documentation:

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

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

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

Team



Dr Heila Pienaar, Deputy Director: Innovation & Technology



Johann van Wyk, Assistant Director: UP Research Data Management



Isak van der Walt: IT Specialist in the Library

Chronological Development of RDM initiatives at UP

- **2007** - Policy for the preservation and retention of research data (Rt 306/07)
- **2010** - Survey of RDM practices at UP (October 2009 – March 2010)
- **11 April 2013** – Meeting with Prof Michael Pepper – Director of Institute for Cellular and Molecular Medicine (ICMM) to discuss the possibility of a pilot project with his students
- **August-November 2013** - Interviews with Deputy Deans Research of all the Faculties to determine “Essential Research Data that the University should manage” as requested by UP Vice Principal: Research – Johann van Wyk
- **December 2013** – 2nd Pilot project – Neuro-Physiotherapy
- **17 March 2014** – High Level Report compiled on RDM internationally, nationally and at UP for UP Vice Principal Research – Johann van Wyk

UP Survey of RDM: what is seen as essential data that should be managed?

- Interviews with Deputy Deans Research of each of the Faculties
- Focus: to determine what is seen as the **essential data** of the faculty that must be managed
- Conducted **eleven interviews**, August – October 2013
- **Trends** were then identified



Wordle of the data management policy of the Office of Digital Humanities, National Endowment for the Humanities

UP Survey of RDM: recommendations

- A Research Data Management Policy for UP
- Establishment of a central research data management office
- Establishment of a research data management presence in each Faculty
- Take into account the impact of RDM on workload and time of researchers and students
- Establishment of data repository for UP
- Investigate necessary IT infrastructure that can accommodate small and big data sets, as well as high performance computing
- Determine a time frame for the roll-out of a research data management system for UP



Pilot Studies

- **Current Projects:**
 - Institute for Cellular & Molecular Medicine project
 - Neuro-Physiotherapy project
- We plan to expand this so that we can have a pilot project in each Faculty as a learning experience
- Open invitation for information specialists to be involved in this



Example of a doctoral student's data

The screenshot shows a 'Repository Browser' interface with a left sidebar and a main content area. The sidebar lists 'Documents', 'Repository', and 'Categories'. The main content area shows a breadcrumb trail: 'Repository > Sites > stemcell > documentLibrary > Completed Research Projects > Madelein Meissner-Roloff PhD Project - Completed 2012'. Below this, a list of folders is displayed, each with a blue folder icon, a title, a modification date, and a list of actions (Favorite, Like, Comment). The folders are: 'Articles', 'CFU pictures', 'Flow cytometry data', 'Lab book', 'PhD Thesis Final 2013', and 'Raw data'. The titles 'Articles', 'Lab book', 'PhD Thesis Final 2013', and 'Raw data' are circled in red. The 'Flow cytometry data' folder is highlighted in light green.

On the right side of the screenshot, a circular diagram illustrates the data lifecycle. The diagram consists of six maroon circles arranged in a circle, connected by a pink arrow pointing clockwise. The circles are labeled: 'CREATING DATA' (top), 'PROCESSING DATA' (top-right), 'ANALYSING DATA' (bottom-right), 'PRESERVING DATA' (bottom), 'GIVING ACCESS TO DATA' (bottom-left), and 'RE-USING DATA' (top-left). In the center of the diagram, the text 'Analysing data' is displayed above a list of bullet points: 'interpret data', 'derive data', 'produce research outputs', 'author publications', and 'prepare data for preservation'.

Pilot Projects: ICMM

http://www.icmm.up.ac.za/share/page/document-details?nodeRef=workspace://

Data Management - Scholarly ... Building Research Data Manag... Alfresco » Document Details

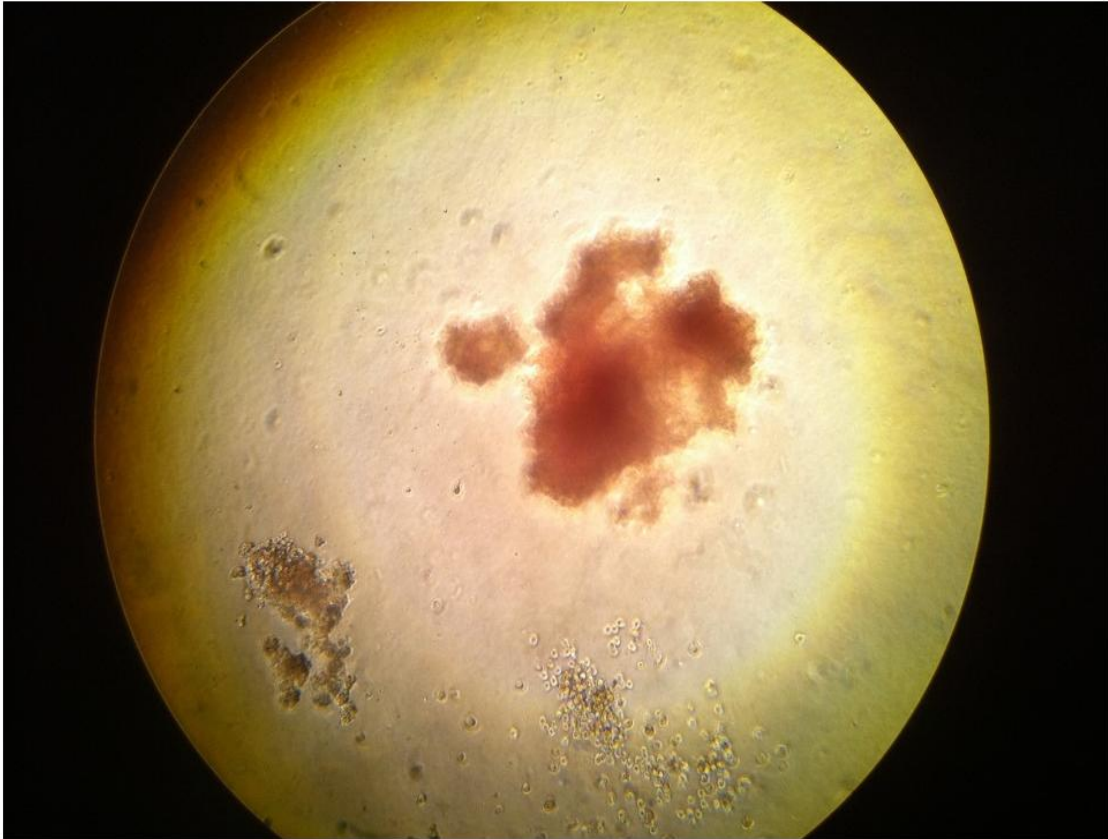
File Edit View Favorites Tools Help

Suggested Sites Get more Add-ons alianet LIS jobs Apple Disney Yahoo!

Page Safety Tools

IMG_1609.JPG 1.0 Download

Modified by Administrator on Thu 28 Nov 2013 10:02:56 Favorite Like 0 Comment Share



▼ Document Actions

- Download
- View In Browser
- Edit Properties
- Upload New Version
- Edit Offline
- Copy to...
- Move to...
- Delete Document
- Start Workflow
- Manage Permissions
- Manage Aspects
- Change Type
- Publish

▼ Share

Copy this link to share the current page

<http://www.icmm.up.ac.za/share/page/document-details?nodeRef=workspace://>

copy link to share this page

▼ Properties

Name: IMG_1609.JPG

100%

02:26 PM 2014/01/13

Next phase

- A campus-wide database / repository for open access data sets with DOI's

PURDUE UNIVERSITY | **Purdue University Research Repository**

Login Register Report a bug

Home Publications Projects Get Started Contact Us

Search

Publish Datasets with DOIs

Use PURR to publish datasets with Digital Object Identifiers (DOI) that make it easier for people to cite your data and give you credit. Purdue is a founding member of DataCite, the international agency that registers DOIs for data.

[Learn More](#)

```
<header>
<identifier> http://dx.doi.org/10.4231/D39P2W550
<datestamp>2012-10-24T12:28:53-04:00
<testSpec><Dataset></testSpec>
</header>
<metadata>
<dc:doi>
</dc:doi>
</metadata>
</pre>

DataCite



### 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



Evaluation of Function Predictions for Moonlighting Proteins



By Ishita K. Khan, Meghana Chitale, Catherine Rayon, Daisuke Kihara  
Purdue University, Université de Picardie Jules Verne



Supplemental datasets used for evaluation of function predictions for moonlighting proteins.



### Do you have a question?



Ask a Librarian



Send us your email question.



Enter your email address



Your Question/Message



Send


```

Long-term preservation

Archival information package (AIP)

- [Bagit](#) “bag” contains: (bag it and tag it)
 - bag declaration file, manifest file, data files
 - Metadata file (XML):
 - METS wrapper
 - Dublin Core and MODS (descriptive metadata)
 - PREMIS (preservation metadata)
- [MetaArchive](#): LOCKSS replication network (7 copies)-
USA example