



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

<https://www.flickr.com/photos/infocux/8450>



RDM @ University of Pretoria

Presentation to NWU

Johann van Wyk

23 June 2016



Chronological Development of RDM



Mike Prins Wikipedia



By JMK on Wikimedia Commons

- **2007** - Policy for the preservation and retention of research data
- **2010** - Survey of RDM practices at UP (October 2009 – March 2010)
- **April 2013** – Meeting - Director of Institute for Cellular and Molecular Medicine (ICMM) on possibility of pilot project with students, which was subsequently implemented
- **August-November 2013** - Interviews with Deputy Deans Research Faculties to determine the “Essential Research Data that the University should manage”
- **December 2013** – 2nd Pilot project – Neuro-Physiotherapy

Chronological Development of RDM



Mike Prins Wikipedia



By JMK on Wikimedia Commons

- **April 2014** – Visit by Deputy Director Innovation and Technology and Library IT Specialist (UP Library Services) to Purdue University in USA: investigate Purdue’s Research Data Repository (PURR) and long-term preservation processes as possibility for replication at UP
- **June 2014** - Assistant Director RDM attended CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, in Beijing, China.

Chronological Development of RDM



Mike Prins Wikipedia



By JMK on Wikimedia Commons

- **July 2014** - High Level Report on RDM sent through to University Executive for review
- **August 2014** – Proposed new University policy on RDM sent through to University Executive for review
- **Jan 2015** – Task Team to investigate infrastructure and policy for RDM across the University
- **Nov 2015** – Revisions done on RDM policy, and final version sent through to UP Executive for review

Chronological Development of RDM



Mike Prins Wikipedia



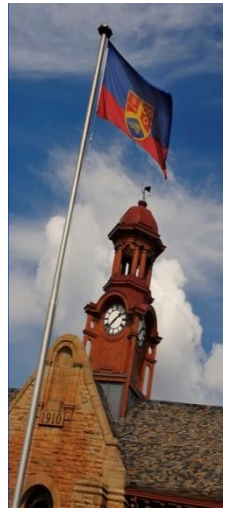
By JMK on Wikimedia Commons

- **2016** – Variety of workshops planned in different aspects of RDM for information specialists
- **Jan-May 2016** – Involved in testing of DMP Tool for DIRISA
- **Jan – Mar 2016** – Involved in testing the DOI system for NRF
- **2016** – Collaborate with DIRISA in developing a national Data Repository platform
- **2016** – Approval of RDM Policy by UP Executive?
- **2016/17** – Roll-out of RDM infrastructure across university?

Survey - October 2009 – March 2010: Findings

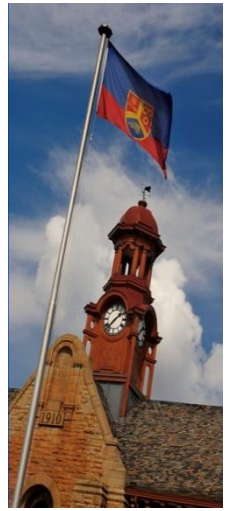
- Funding: In most cases no need for data management or data sharing plans, depending on funding agency requirements
- Data Collection: Wide variety of data collection methods used. Both primary and secondary data. Data sets are often small.
- Data Storage: Ad hoc storage of data, both on paper and electronically
- Publishing: In general raw data not published
- Support: Lack of support with regard to storage of data (physical and electronic).
- Concern: Impact on researchers' time

RDM does not exist in any formal manner (with the exception of one or two departments) at the University of Pretoria



Survey: Oct 2009 – Mar 2010: Recommendations

- Investigate Very Large Database Initiative (now DIRISA) from the Department of Science and Technology as a possibility to support UP's RDM needs
- Central UP server or repository
- Address the need for physical storage space
- Create a formal staff position of 'research data manager' to drive RDM at UP



Survey : Aug – Oct 2013

- Interviews with Deputy Deans Research of each of the Faculties
- Focus: 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



Survey : Aug – Oct 2013: Trends

- **Essential Data**

Interview data, Questionnaires, Spread sheet data, Lab books, Experiment / laboratory data, Images (e.g. graphs, models, Sketches, X-Rays, scans etc.), Literature reviews, Sequencing data, Computer-generated data.

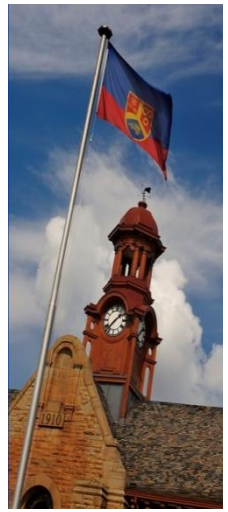
- **Level of Data**

Some see only raw data as essential, some see processed /analysed data as essential, while others see both raw and processed/analysed data as essential

- **Volume of data**

A small number work only with small data sets.

Majority work with small and big data sets, with exponential increase in big data sets



Survey : Aug – Oct 2013: Trends

- **Data Formats used**

Excel, Pdf, MS Word, Text Format, images in various formats, video, sound, various computer generated formats, SPSS, SAS, AMOS, Qualtrics data, SurveyMonkey data, simulation data formats, and even data from social media. Some in paper format.

- **RDM Plans**

None have Research Data Management Plans in place.

- **Uploading capacity**

No capacity to upload these data sets to a repository

- **Willingness to share data**

Majority willing to share their data under certain conditions.

Health Sciences not willing to share their data



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

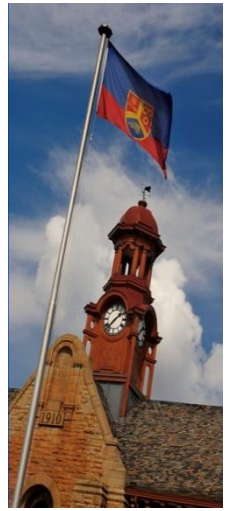
<https://www.flickr.com/photos/rosefirerising/6776182890/>



http://en.wikipedia.org/wiki/File:Open_Data_stickers.jpg

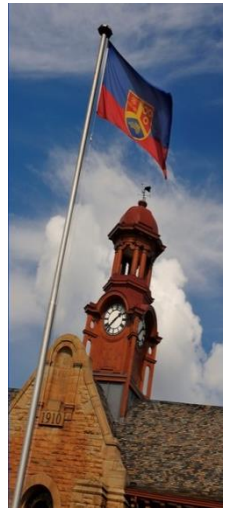
Survey: Aug – Oct 2013: Recommendations

- A new Research Data Management Policy for UP
- Establishment of a central research data management office
- Establishment of a RDM presence in each Faculty
- Consider impact of RDM on workload and time of researchers and students
- Establishment of data repository for UP
- Investigate necessary IT infrastructure for RDM – (handle small and big data sets, and HPC)
- Determine a time frame for the roll-out of a RDM system for UP



Pilot Projects at UP

- Five data management pilot projects in 2013-2016: Institute for Cellular and Molecular Medicine (ICMM) and the Neuro Physio Group, Potato Pathology Programme, Powdery Scab, and Psychiatry Dissociation
- An Open Source Document Management System Alfresco was customised for this purpose



Why Alfresco?

Open Source

Captured provenance of data

Has a versioning function

Good metadata function

Easy to integrate with other software

Workflow function gave supervisor overview of progress of students

Sync function with Dropbox and Google Drive

Drag and Drop function

File Sharing function

Mobile App

IT Infrastructure for RDM at the University



RESEARCH INFRASTRUCTURED DRIVERS – RDM

1. Implementation of UP's Research Data Management(RDM) Policy:

Governance of the management of *research data* at UP is essential whilst aiming to ensure that all research data generated at UP is managed and curated effectively and efficiently in order to support the University's positioning as an international research-intensive institution.

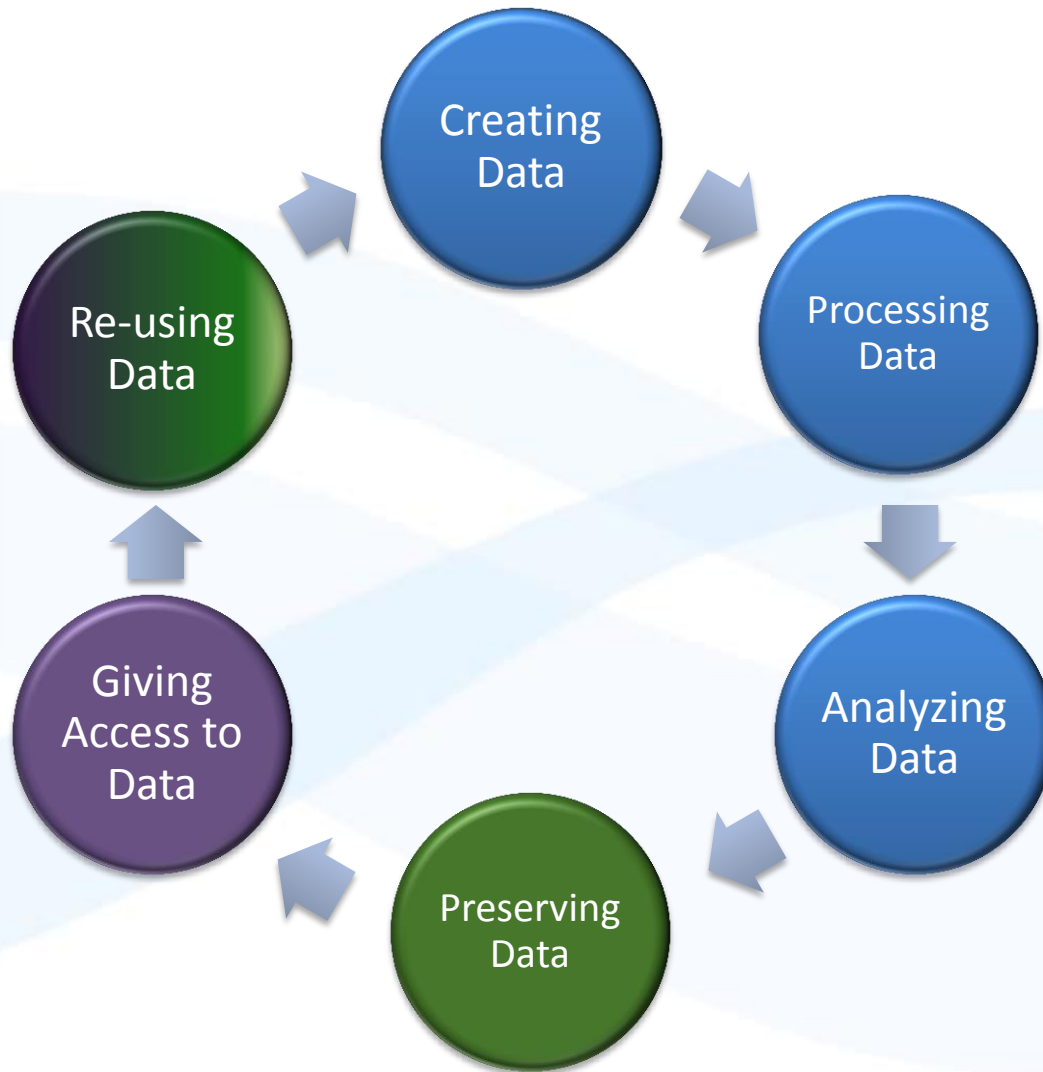
2. Funder's Requirements

- **National Research Foundation(NRF) requirement:**

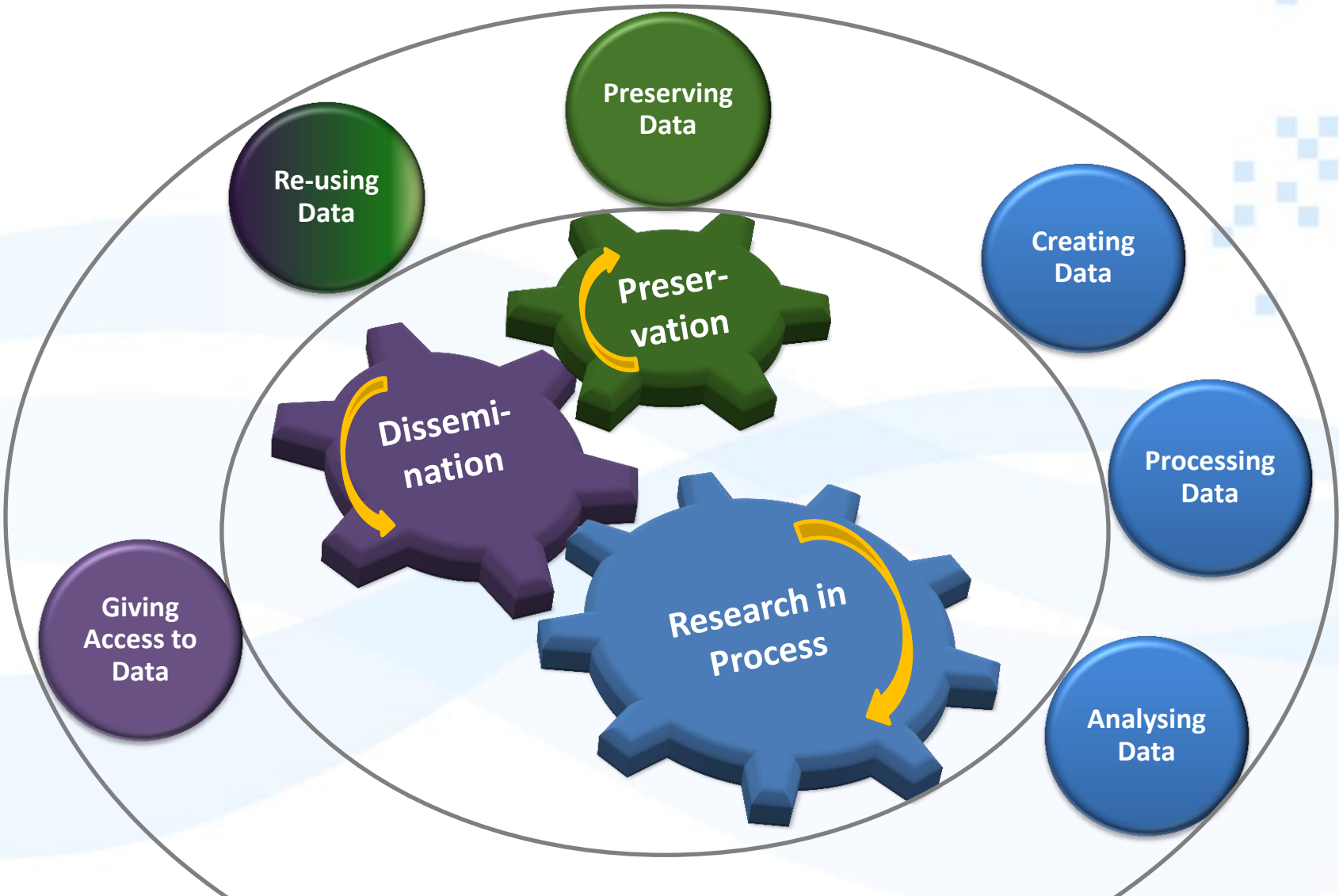
Need to deposit NRF-funded *research output/s* and the *data supporting the output/s* in an accredited Open Access (OA) data repository, with the provision of a Digital Object Identifier (DOI) for future citation and referencing.

- **Compliance with similar requirements of other funding bodies requirements**

RESEARCH DATA LIFE CYCLE

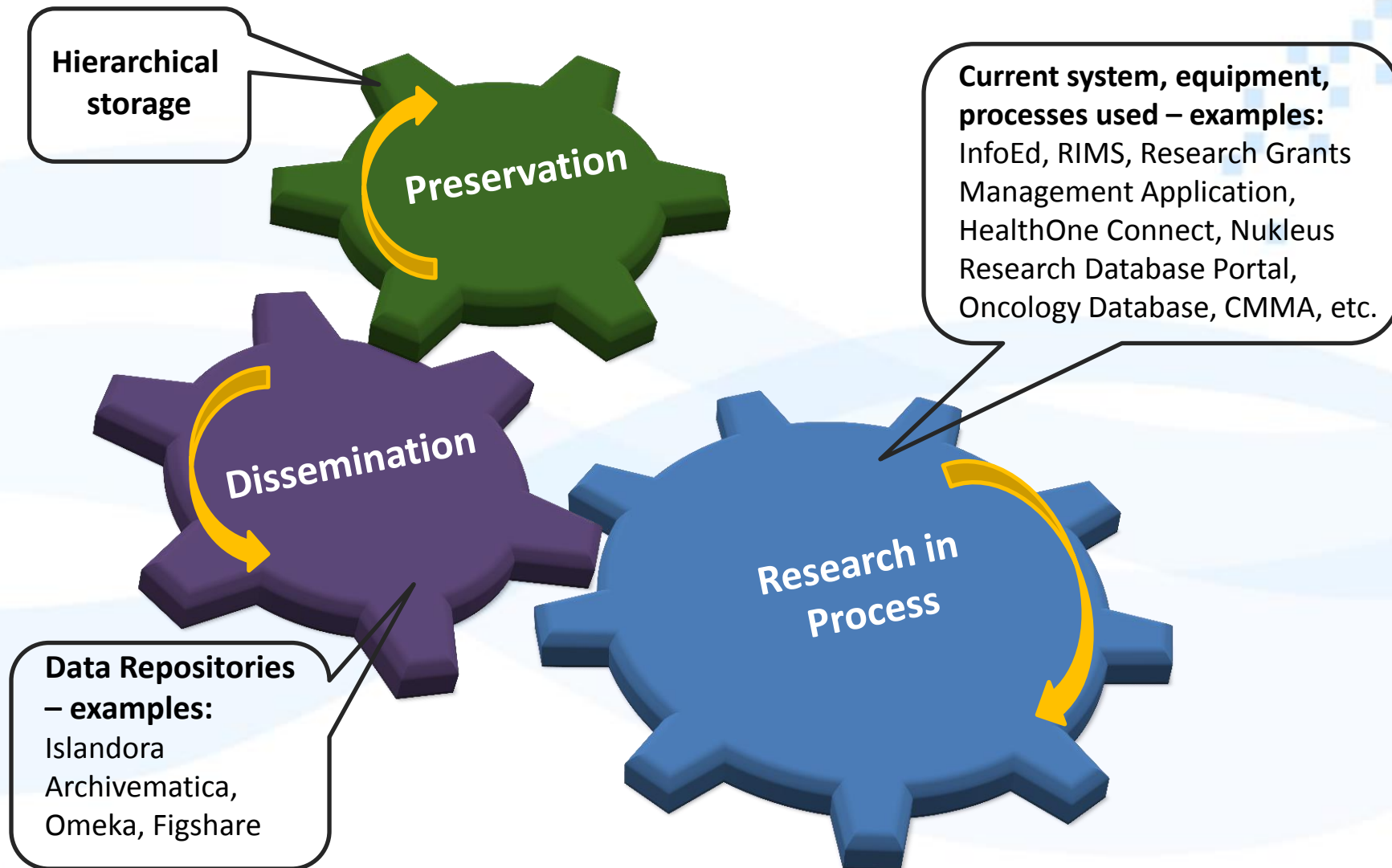


PROCESSES within the RESEARCH DATA LIFE CYCLE

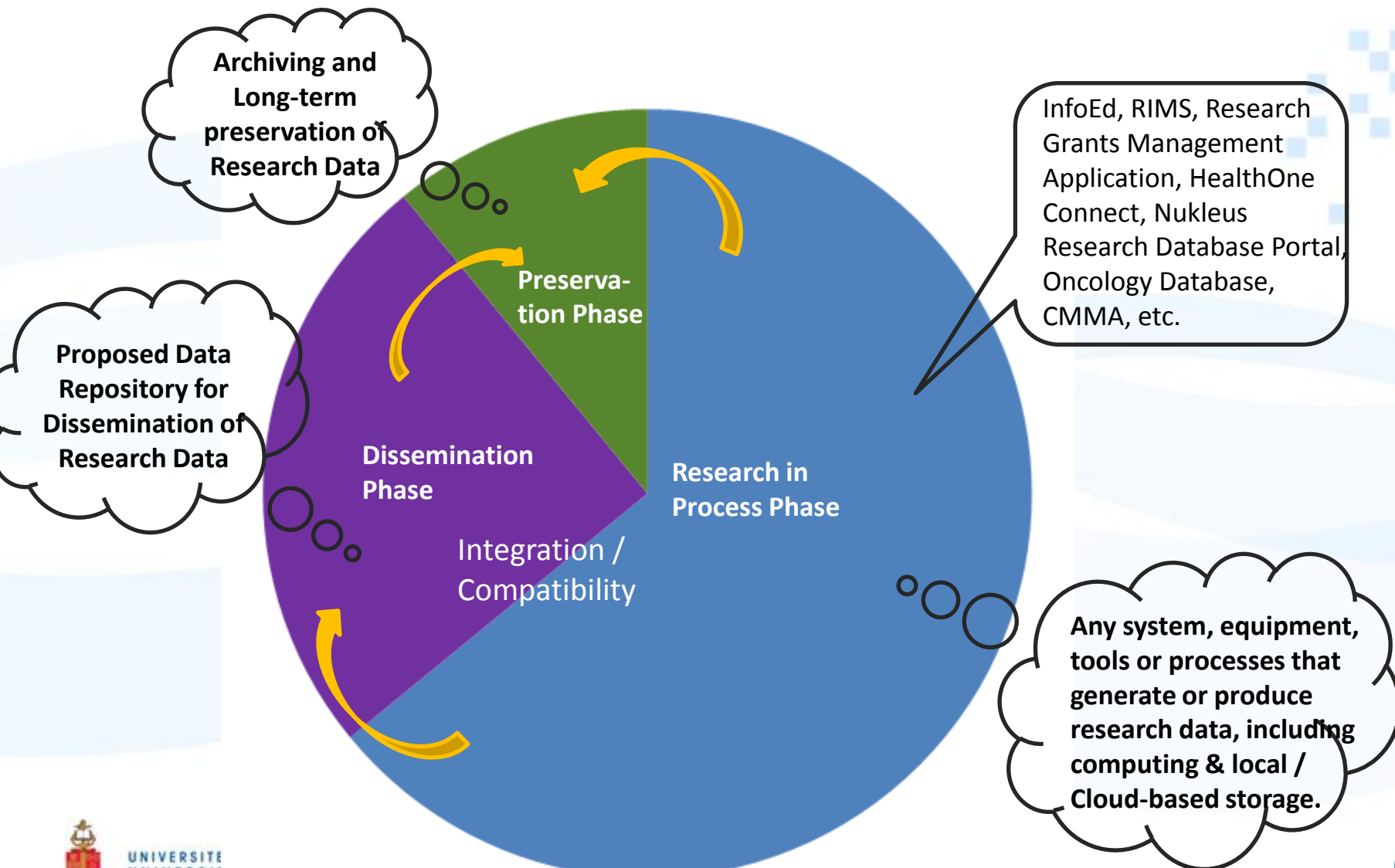


Research Data Life Cycle

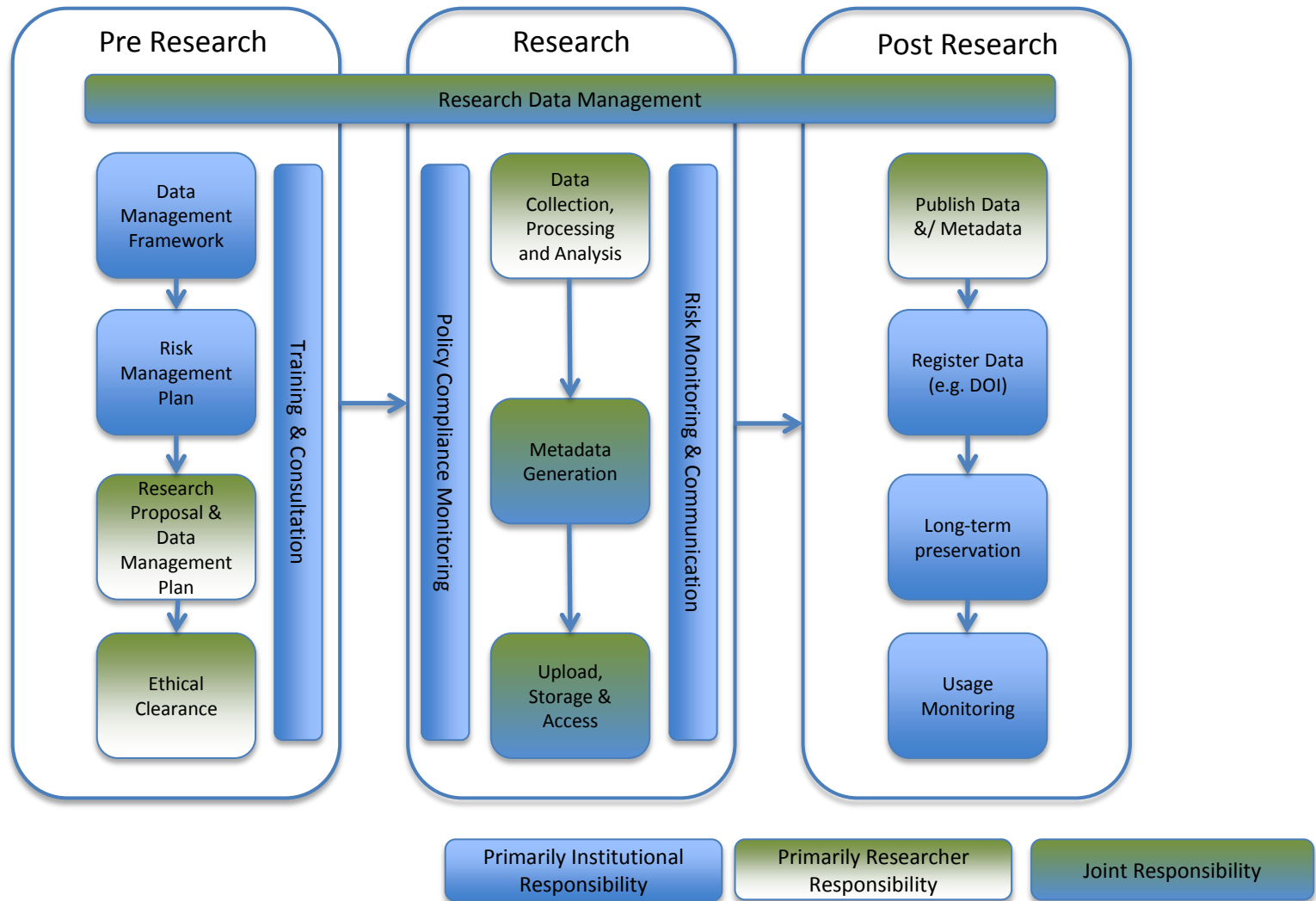
DATA FLOW within the RESEARCH DATA LIFE CYCLE



OVERVIEW OF INTEGRATED INFRASTRUCTURE REQUIREMENTS



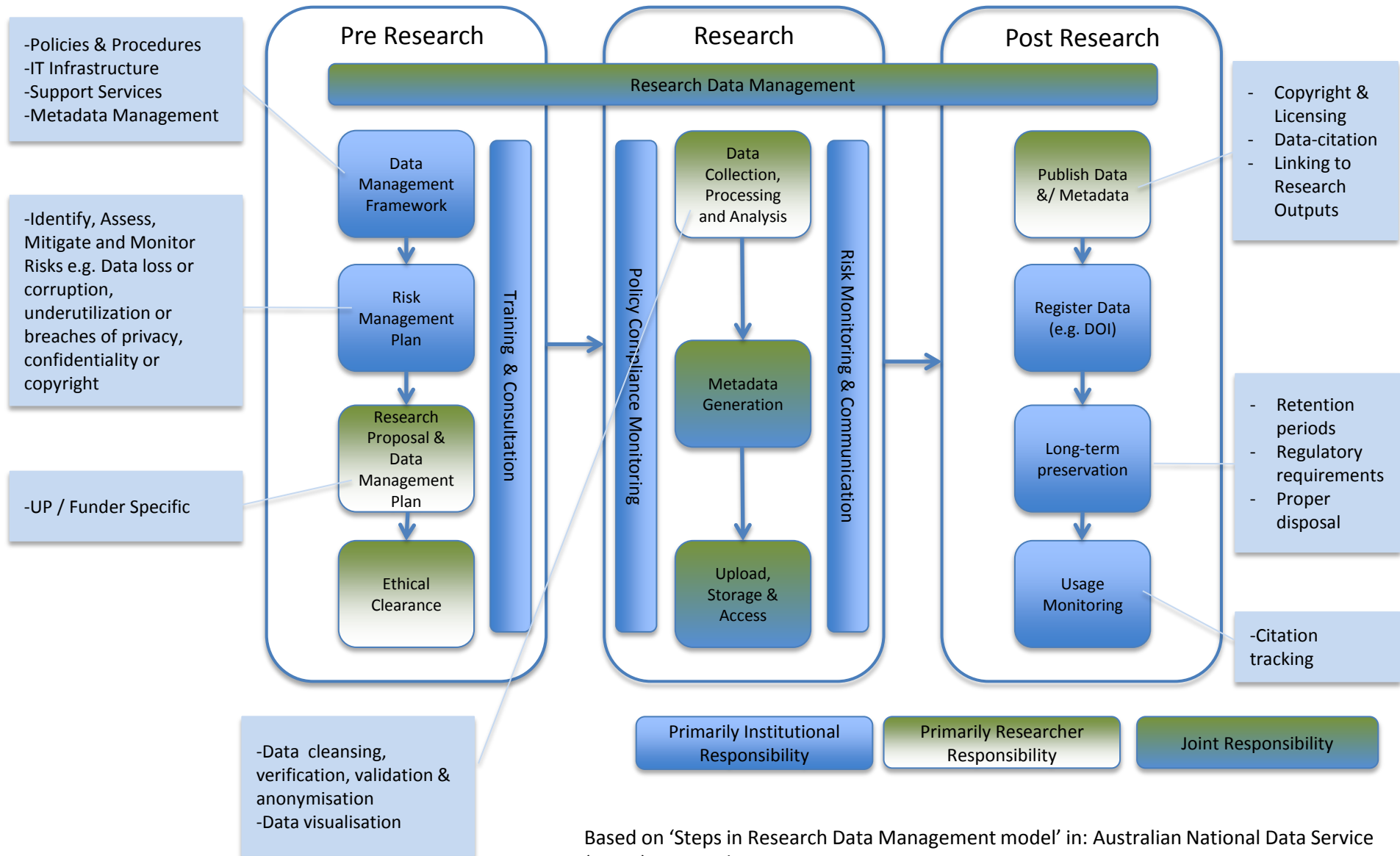
UP: Key Steps in Research Data Management



Based on 'Steps in Research Data Management model' in: Australian National Data Service (ANDS): Research Data Management in Practice. May 2013.

http://ands.org.au/_data/assets/pdf_file/0009/394056/research-data-management-in-practice.pdf

UP: Key Steps in Research Data Management (more detail)



Based on 'Steps in Research Data Management model' in: Australian National Data Service (ANDS): Research Data Management in Practice. May 2013.

http://ands.org.au/_data/assets/pdf_file/0009/394056/research-data-management-in-practice.pdf