Research Data Management (RDM) in a Developing Country: a personal journey

Dr Martie van Deventer

Dr Heila Pienaar

CSIR
our future through science

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Roadmap

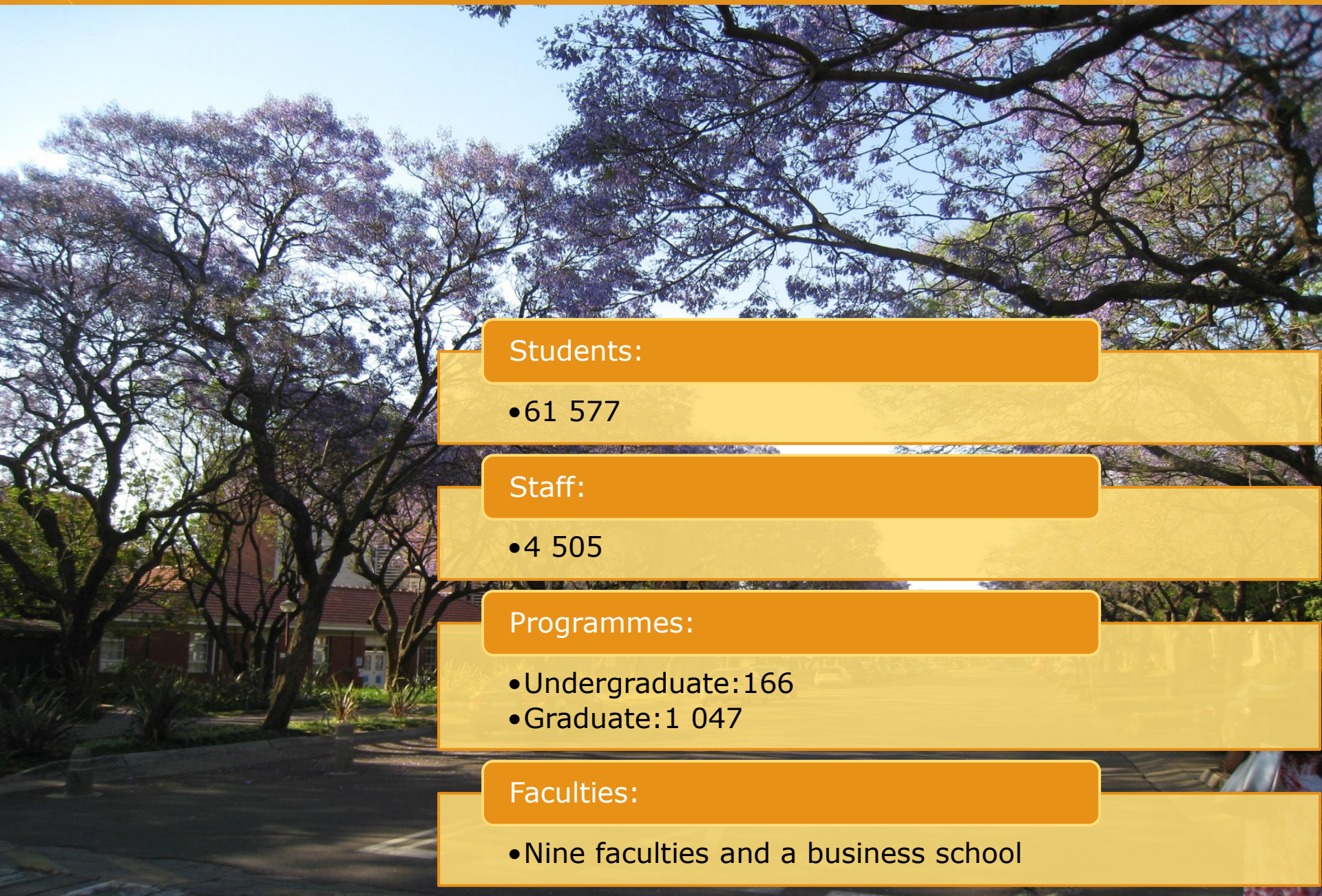
Introduction
National: DIRISA
Learning from international colleagues
Lessons learned
RDM @ CSIR
NeDICC CoP
RDM @ UP

The CSIR at a glance

- The CSIR is a science council, classified as a national government business enterprise – conducting research on behalf of government (but also for the private sector)
- The CSIR's Executive Authority is the Minister responsible for the Department of Science and Technology

69	years in 2014
2411	total staff
1 691	total in SET base
310	doctoral qualifications
~R2.15 bn	total operating income

University of Pretoria at a glance



Students:

- 61 577

Staff:

- 4 505

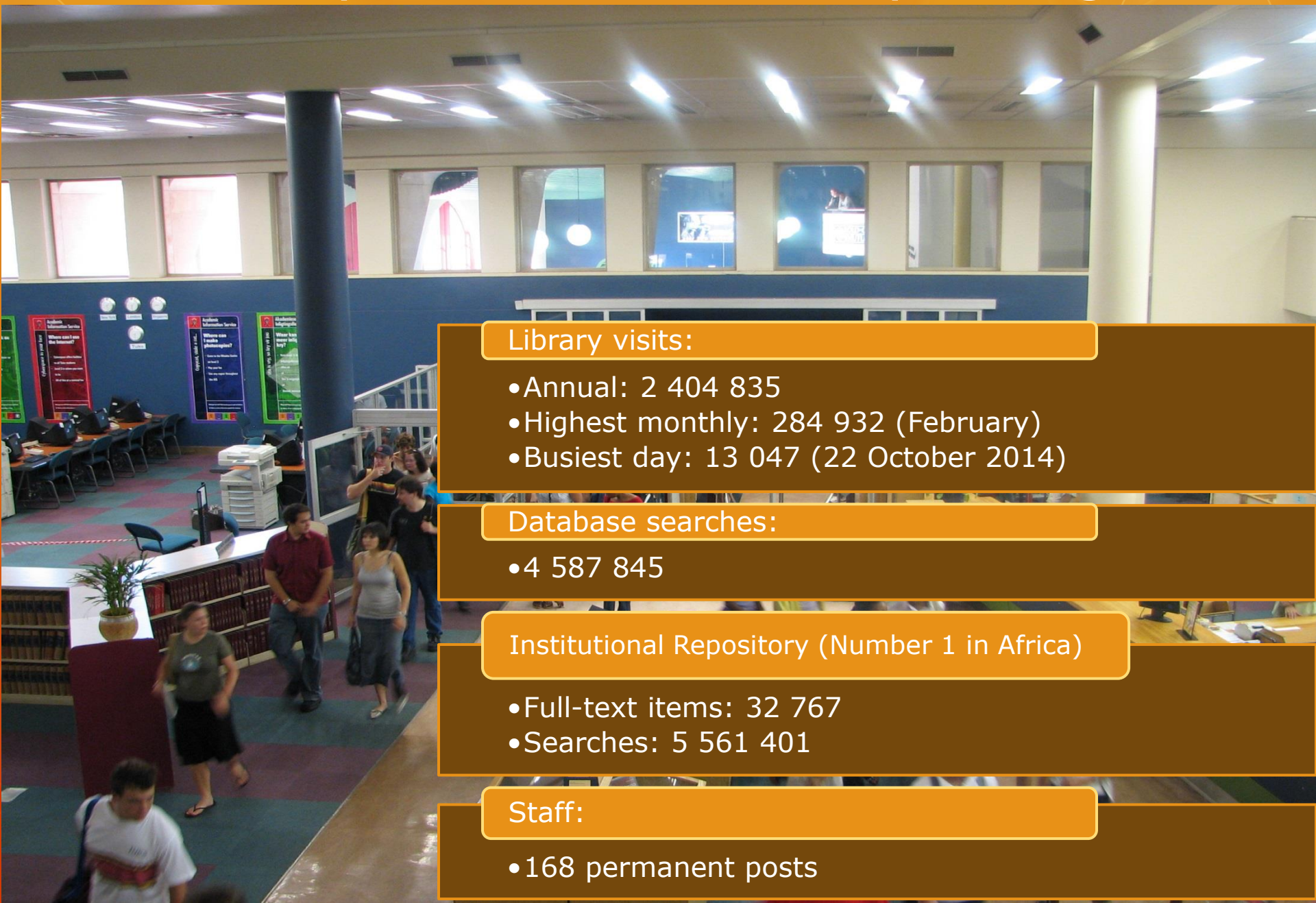
Programmes:

- Undergraduate:166
- Graduate:1 047

Faculties:

- Nine faculties and a business school

University of Pretoria Library at a glance



Library visits:

- Annual: 2 404 835
- Highest monthly: 284 932 (February)
- Busiest day: 13 047 (22 October 2014)

Database searches:

- 4 587 845

Institutional Repository (Number 1 in Africa)

- Full-text items: 32 767
- Searches: 5 561 401

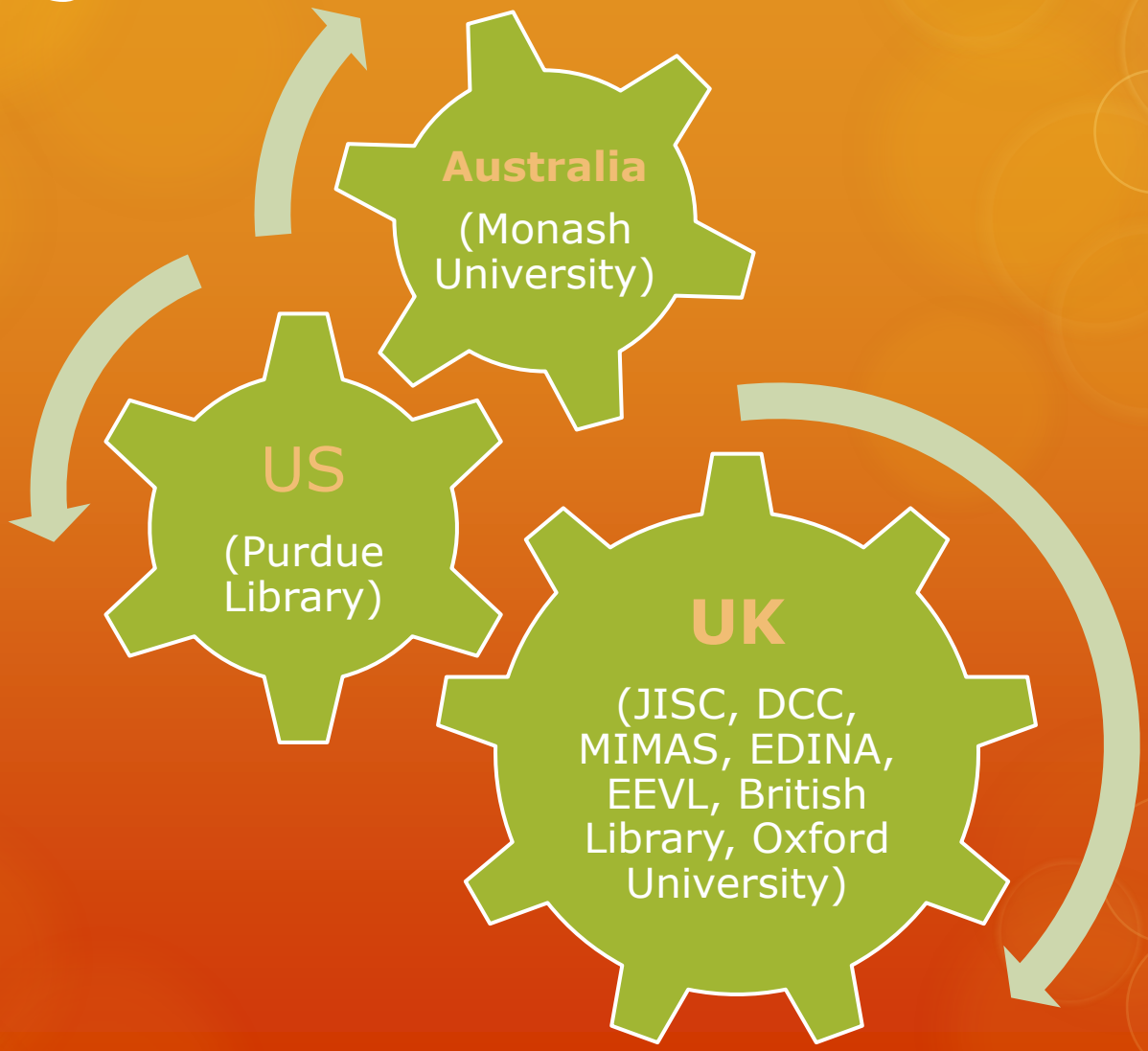
Staff:

- 168 permanent posts

Introduction

- 2004: South African Research Information Services **(SARIS)** project with Ford Foundation Funding
- We were **members** of the 6 person team
- **Initial focus**: high cost of electronic literature
- After we visited the **UK** the focus was expanded to include **e-Research**, including **RDM**
- We then initiated several **RDM initiatives in SA** e.g. conferences & workshops
- Created a **RDM CoP**: NeDICC (Network of Data and Information Curation Communities)
- Also initiated and managed RDM projects at the **CSIR and UP**
- Visited several **international** initiatives on RDM
- National initiatives are being implemented via **DIRISA** (Data Intensive Research Infrastructure for SA)

Learning from international colleagues



Learning from international colleagues

Purdue

Oxford

NeDICC CoP

- NeDICC aims to promote the development and use of research data and information curation standards and practices to ensure the long term preservation and accessibility of digital research outputs in support of e-Research http://nedicc.com/
- We are both members of the **steering committee**, with Martie as the **chair**
- **Members**: 3 large universities and 4 national research councils
- **Activities**: discussions; organising conferences / conference tracks; workshops; training e.g. Carnegie CPD
- Involvement in rest of **Africa** via the conferences and Carnegie CPD

Carnegie CPD module on RDM, 2014, Pretoria

2013 African Digital Scholarship and Curation conference, Durban, South Africa

RDM @ UP

- 2007: Policy for the preservation and retention of research
- 2010: **Survey** of RDM practices at UP
- 2013: Assignment of a **full-time staff** member to RDM
- **Surveys and documents**:
  - 2013: **Survey** of the Deputy Deans: Research to determine "Essential research data that the University should manage"
  - 2014: High level **report** compiled on RDM internationally, nationally and at UP
  - 2014: Proposed University **policy** sent through to University Executive for review
  - 2015: Task team (Library & IT) to investigate **infrastructure** needed for RDM across the University
- **RDM pilots (proof of concept)**: Alfresco
  - Institute for Cellular and Molecular Medicine (ICMM)
  - Word of mouth: Neuro Physio-therapy; Potato Pathology; Powdery Scab; Psychiatry Dissociation

Repository Browser

Select ▾ + Create... ▾ Upload Selected Items... ▾ Name ▾ Options ▾

Repository > Sites > stemcell > documentLibrary > Completed Research Projects > Madelein Meissner-Roloff PhD Project - Completed 2012

- ▾ Documents
  - I'm Editing
  - My Favorites
- ▾ Repository
  - Repository
    - Data Dictionary
    - Imap Attachments
    - Sites
      - stemcell
        - calendar
        - dataLists
        - discussions
        - documentLibr
          - Common S
          - Completed
            - Carla D
            - Dr Mari
            - Fatima
            - Juanita
            - Karlien
            - Madele
              - Artic
              - CFU
              - Flow
              - Lab
              - PhD
              - Raw
            - Tyren D
        - Instrument
        - Projects
        - Sync Softw
- ▾ Categories

Articles
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

CFU pictures
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

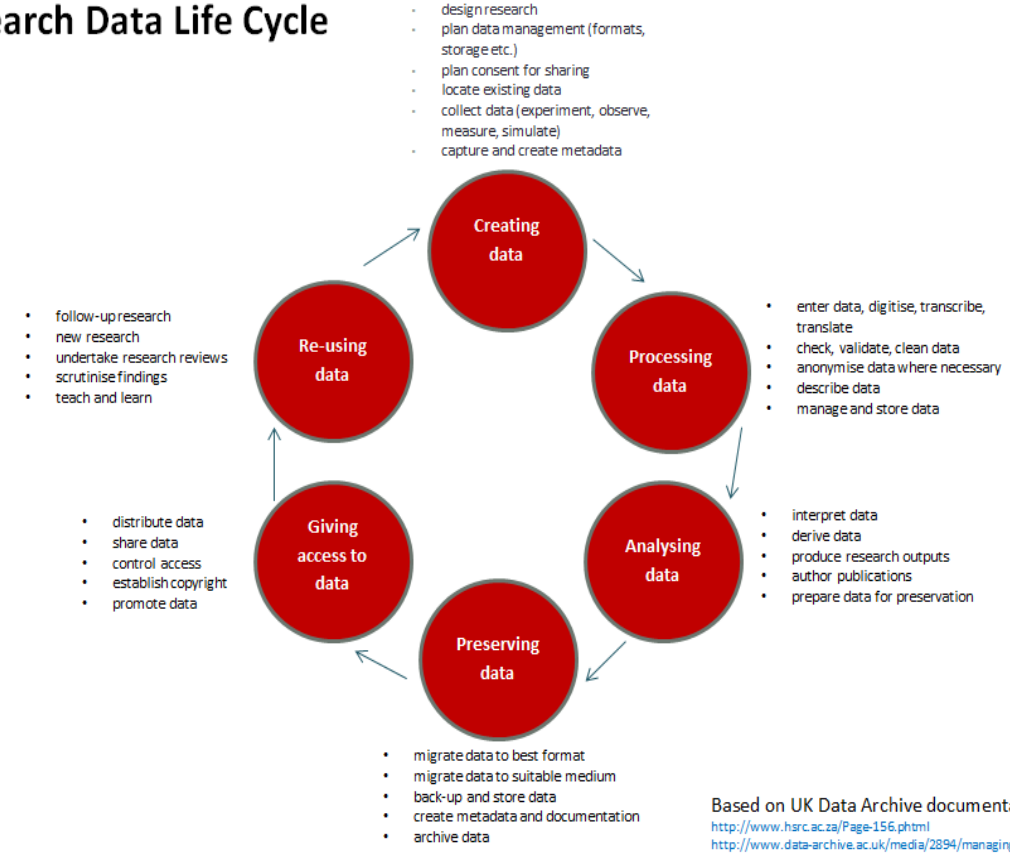
Flow cytometry data
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

Lab book
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

PhD Thesis Final 2013
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

Raw data
Modified about a month ago by Administrator
No Description
No Tags
Favorite | Like 0 | Comment

Research Data Life Cycle



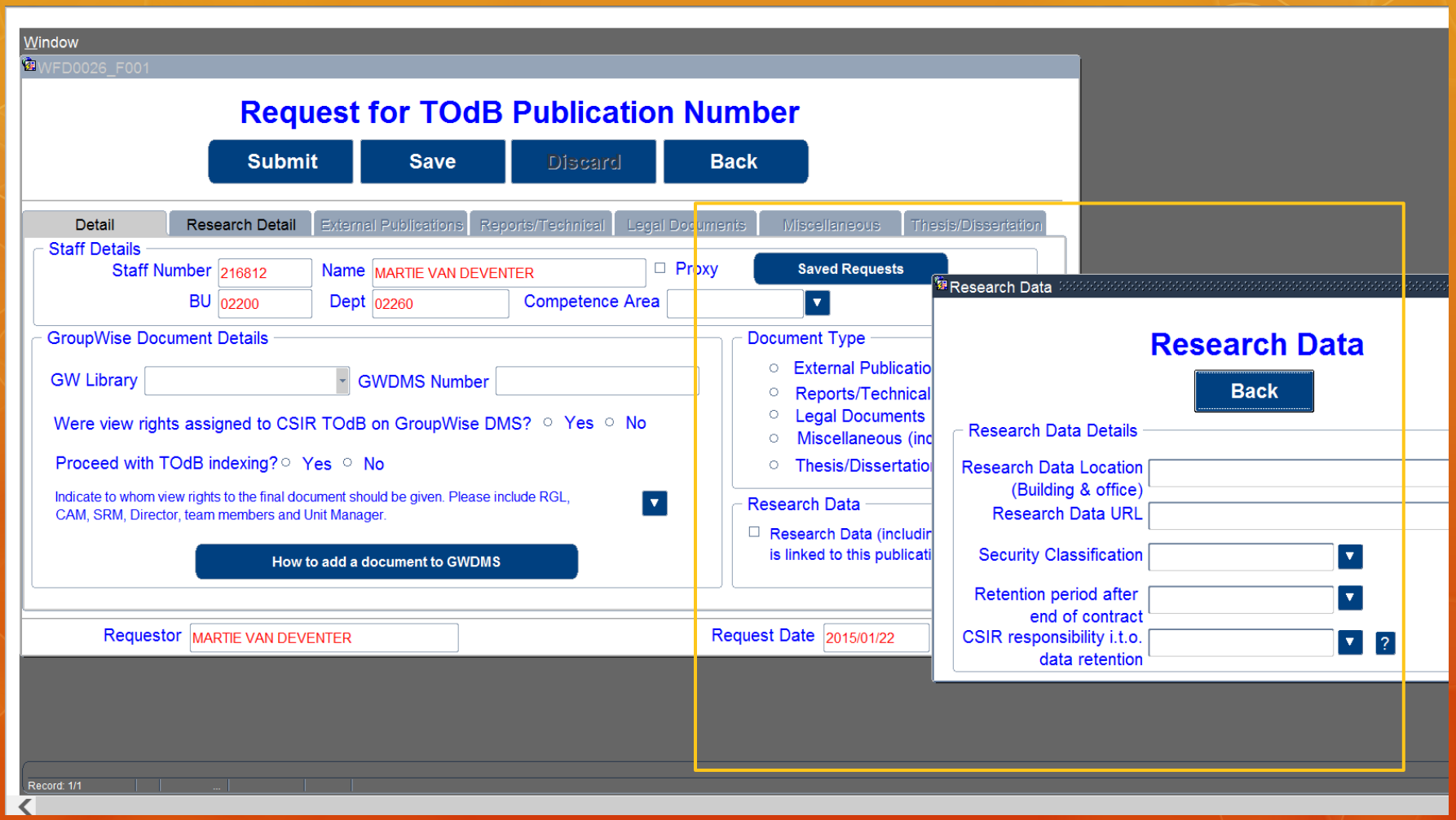
Based on UK Data Archive documentation:
http://www.hsrc.ac.za/Page-156.phtml
http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
http://www.data-archive.ac.uk/

100%

RDM @ CSIR

- **Activity started in 2010**
  - Assisted our Natural Resources researchers to gain control over their **geospatial datasets**.
  - Involved **destroying data** (discs/ CDs/ DVDs) for which there just was no provenance, reliable documentation or metadata.
  - A **geoportal** was the established with a central storage server.
  - Formal **workflow** – to surface geospatial data.
- **In 2014 a data librarian was appointed**
  - A CSIR-wide **survey** was completed by August 2014.
  - Now have a beter understanding of the spectrum and formats of data generated. Researcher behaviour also documented.
  - Completed **CARDIO-model** exercise with our ICT department to establish a benchmark – level 2.
  - Workflow expanded to include other data.

Data is acknowledged as a research record



National: NICIS & **DIRISA**

- Formation of the National Integrated Cyber-Infrastructure System (NICIS) in 2014
- NICIS framework includes **four initiatives**;
  - High Performance Computing (HPC)
  - SA National Research Network (SANReN)
  - SA Grid (SA-GRID)
  - **Data Intensive Research Infrastructure for SA (DIRISA)**
- **DIRISA** will primarily fulfil enabling, supporting and facilitation roles
- The programme manager is currently completing a business plan with a 3 year implementation horizon
- The plan will make ample provision for community and stakeholder engagement e.g. NeDICC

DIRISA members and the international review panel

Lessons learned

- International practices and guidelines must be **adapted** to suit your own set of researchers & implementation takes place at your own pace
- A developing country can also make valuable **contributions** to RDM theory and practice: Incorporating RDM in VRE development and vice versa
- Commitment from very **senior decision makers** is essential for national implementation
- Both **bottom up** and **top down** initiatives have a role to play
- Learning involves more than travelling, attending conferences and building a professional network. True learning comes from 'elbow grease'
- Librarians should define the **role** that they want to play w.r.t. RDM, as there are many role players involved
- Do not see RDM as a burden only – there is also much fun to be had during the journey!!!

Hot off the press!!!

National Research Foundation

Statement on Open Access to Research Publications from the Nat Foundation (NRF)-Funded Research

Background

The National Research Foundation (NRF) was established through t Foundation Act (Act No.23 of 1998). As an independent statutory a promotes and supports research in South Africa largely throug Education Institutions (HEIs), National Research Facilities and Sci to generating knowledge and promoting high-level research c System of Innovation (NSI).

Supporting scientific research through public funding is impor economy, promoting innovation and stimulating appropriate NRF-funded research outputs contributes to the knowl access to this knowledge base facilitates:

- Dissemination of knowledge;
- Utilisation/uptake/application of this knowledge
- Accelerated transformation and globalisation of science,
- Rapid transformation of this knowledge into innovative an applications for the benefit of society.

Statement

The NRF recognises the importance of Open Access to science and research while at the same time appreciating that Open Access will continue to evolve in response to societal needs, achieving overarching policy harmonisation and new innovative publishing business models.

From 01 March 2015, authors of research papers generated from research either fully or partially funded by NRF, when submitting and publishing in academic journals, should deposit their final peer-reviewed manuscripts that have been accepted by the journals, to the administering Institution Repository with an embargo period of no more than 12 months. Earlier Open Access may be provided should this be allowed by the publisher. If the paper is published in an Open Access journal or the publisher allows the deposit of the published version in PDF format, such version should be deposited into the administering Institutional Repository and Open Access should be provided as soon as possible.

1

In addition, the data supporting the publication should be deposited in an accredited Open Access repository, with the provision of a Digital Object Identifier for future citation and referencing.

The NRF encourages its stakeholder community, including NRF's Business Units and National Research Facilities, to:

- Formulate detailed policies on Open Access of publications and data from its funded research;
- Establish Open Access repositories; and
- Support public access to the repositories through web search to international standards and best practice.

The NRF requires its relevant Business Units and collaborate with relevant governmental research institutions to facilitate Open research. The NRF requires its international scientific from publicly funde

From 01 March 2015 authors of research papers generated from research funded by the NRF … deposit their final peer-reviewed manuscripts to … Institution Repository. The data supporting the publication should be deposited with a DOI in an Open repository