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

Dr Martie van Deventer

Dr Heila Pienaar
Roadmap

Lessons learned

- Introduction
- Learning from international colleagues
- National: DIRISA
- RDM @ CSIR
- NeDICCoP
- 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

Australia
(Monash University)

US
(Purdue Library)

UK
(JISC, DCC, MIMAS, EDINA, EEVL, British Library, Oxford University)
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/](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
Research Data Life Cycle

- Creating data
- Processing data
- Analysing data
- Preserving data
- Giving access to data
- Re-using data

- 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

- enter data, digitise, transcribe, translate
- check, validate, clean data
- anonymise data where necessary
- describe data
- manage and store data

- distribute data
- share data
- control access
- establish copyright
- promote data

- migrate data to best format
- migrate data to suitable medium
- back-up and store data
- create metadata and documentation
- archive data

Based on UK Data Archive documentation:
http://www.hrc.ac.uk/Page-156.html
http://www.data-archive.ac.uk/About/2334/managingsharing.pdf
http://www.data-archive.ac.uk/
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 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 better 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
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!!!

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.