https://www.e-education.psu.edu/dmpt/node/508

Research Data Management @ UP and the Role of NeDICC

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

By Johann van Wyk

Presentation at the Panel Discussion on Managing Research Data at the University of the Free State, 11 October 2017

Make today matter

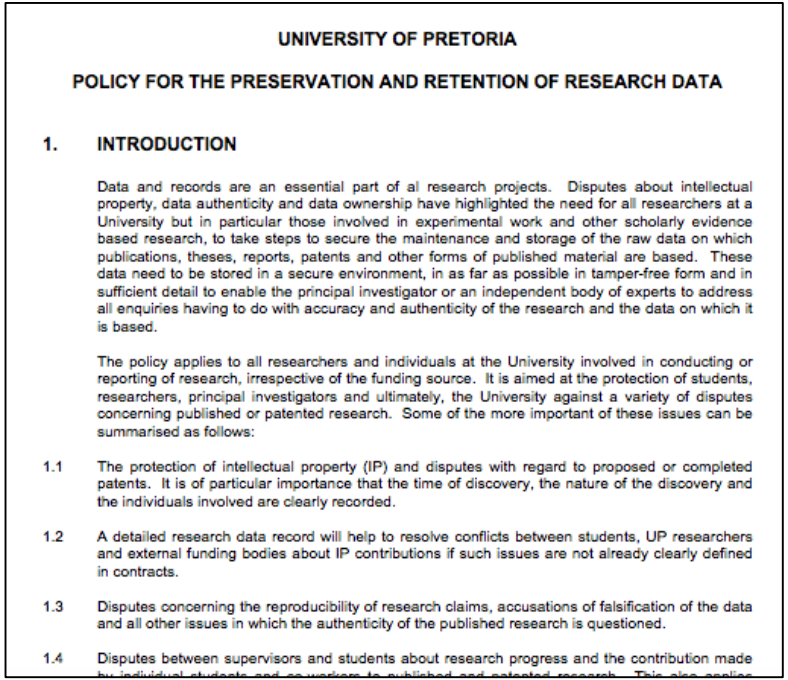
Agenda

- RDM at University of Pretoria
  - First RDM initiative
  - 1St Survey of RDM practices at UP
  - 2nd Survey of RDM practices at UP
  - Pilot projects
  - National and International engagement
  - New UP RDM Policy
  - Training
  - Other developments at UP
  - Research Data Lifecycle
  - Investigation of Research Data Repository Software Platforms
- NeDICC
  - Membership
  - Steering Committee
  - Aims of NeDICC
  - NeDICC activities and programme
  - Collaborative projects

First RDM initiative

2007

Policy for the preservation and retention of research data



UNIVERSITY OF PRETORIA

POLICY FOR THE PRESERVATION AND RETENTION OF RESEARCH DATA

1. INTRODUCTION

Data and records are an essential part of al research projects. Disputes about intellectual property, data authenticity and data ownership have highlighted the need for all researchers at a University but in particular those involved in experimental work and other scholarly evidence based research, to take steps to secure the maintenance and storage of the raw data on which publications, theses, reports, patents and other forms of published material are based. These data need to be stored in a secure environment, in as far as possible in tamper-free form and in sufficient detail to enable the principal investigator or an independent body of experts to address all enquiries having to do with accuracy and authenticity of the research and the data on which it is based.

The policy applies to all researchers and individuals at the University involved in conducting or reporting of research, irrespective of the funding source. It is aimed at the protection of students, researchers, principal investigators and ultimately, the University against a variety of disputes concerning published or patented research. Some of the more important of these issues can be summarised as follows:

1.1 The protection of intellectual property (IP) and disputes with regard to proposed or completed patents. It is of particular importance that the time of discovery, the nature of the discovery and the individuals involved are clearly recorded.

1.2 A detailed research data record will help to resolve conflicts between students, UP researchers and external funding bodies about IP contributions if such issues are not already clearly defined in contracts.

1.3 Disputes concerning the reproducibility of research claims, accusations of falsification of the data and all other issues in which the authenticity of the published research is questioned.

1.4 Disputes between supervisors and students about research progress and the contribution made

2009

1st Survey of RDM practices at UP

(October 2009 – March 2010)

Recommendations:
Survey - Oct 2009 – Mar 2010

- Investigate Very Large Database initiative (now DIRISA) from the Department of Science and Technology as 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 (research data manager was appointed in 2013)

2013

2nd Survey: Interviews with Deputy Deans Research from Faculties to determine the ‘Essential Research Data that the University should manage’

(August - November)

Recommendations

2nd Survey of RDM at UP: Aug – Oct 2013

- 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 – 2013 - 2014

Implemented 5 pilot projects using Alfresco as testing platform, to gain an understanding of researchers' RDM needs

- Institute for Cellular and Molecular Medicine (ICMM)
- Neuro-Physiotherapy
- Potato Pathology Programme
- Powdery scab
- Psychiatry Dissociation

National & International engagement

- **Since 2008** - Involvement in NeDICC – UP one of the founding members – Currently one of UP staff members are on the Steering Committee of NeDICC
- **April 2014** –Two library staff members visit Purdue University (USA): to investigate Purdue's Research Data Repository (PURR) and long-term preservation processes as possibility for replication at UP
- **June 2014** – One library staff member attends CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, in Beijing, China.

National & International engagement

- **July 2014** - **High Level Report** compiled on national and international developments on RDM (literature study), with some recommendations - sent through to University Executive.
- **June 2015** – Visited UCT's eResearch Centre to gain an understanding of how they manage their research data in the context of eResearch.
- **Aug 2015** – Presentation by Anthony Beitz (Interim Director UCT eResearch) - Overview of eResearch & Research Data Management
- **Oct 2016** – Members from the Library and IT Services attended the eResearch Australia Conference and visited a number of Higher Education institutions to **investigate potential RDM infrastructure**

New UP RDM Policy

- **August 2014** – 1st draft of proposed new University policy on RDM sent through to Vice-Principal responsible for Research and Postgraduate Education
- **Jan 2015** – UP RDM Task Team formed to investigate policy, infrastructure and processes needed for RDM across the University (ITS, Library Services, Research Office, 2 researchers, Vice Principal Research and Postgraduate Education)
- The proposed policy went through many iterations, and the **final version** went to UP Executive in July 2017.
- **September 2017** - RDM policy approved by UP Senate

Aspects covered in the RDM policy

- Introduction - background
- Organisational scope
- Who this policy is applicable to
- The purpose of the policy
- Policy statement
- Statement on ownership – intellectual property
- Statement on access to / sharing of research data and metadata
- Exceptions (e.g. research administration data, institutional governance information)
- Definitions and abbreviations of concepts
- Associated documents – other University policies and legislation
- Roles and responsibilities
- Policy lifecycle – reviewed every five years
- Addendum – Processes and procedures to be followed

Aspects covered in the RDM policy

Addendum – Processes and procedures to be followed

- Generation and collection of data
- Maintenance of data
- Preservation of data
- Destruction of data
- Sharing of data
- Support and training

Training

- **Nov 2015** – First version of the Data Management Planning (DMP) Tool developed in collaboration with DIRISA
- A number of in-house workshops were presented to library professionals since 2015, e.g. The role of library staff in RDM, Data Citation principles, Data Management Plans.
- The UP Library were also involved in presenting a number of NeDICC workshops

Other developments at UP

- 2016 – developed a LibGuide on RDM – https://up-za.libguides.com/c.php?g=356288
- 2016 – Assisted DIRISA in the testing and further development of the DMP (Data Management Planning) Tool that they downloaded from the DCC site onto DIRISA's site
- Mar 2016 – Were one of the institutions that tested the DOI system that the NRF is developing
- June 2016 – Jan 2017 - Investigation of Research Data Repository Software Solutions – a collaborative project between the UP Library Services and UP IT Services departments

Research Data Lifecycle

Creating Data
- Designing Data Management Plans
- Data Capture
- Data Storage
- Metadata creation

Processing Data
- Data Storage
- Metadata Creation
- Data Cleansing
- Data Verification
- Data Validation
- Data Anonymisation

Analysing Data
- Data Storage
- Metadata Creation
- Data Cleansing
- Data Verification
- Data Validation
- Data Anonymisation
- Data Interpretation & Analyis
- Data Visualisation

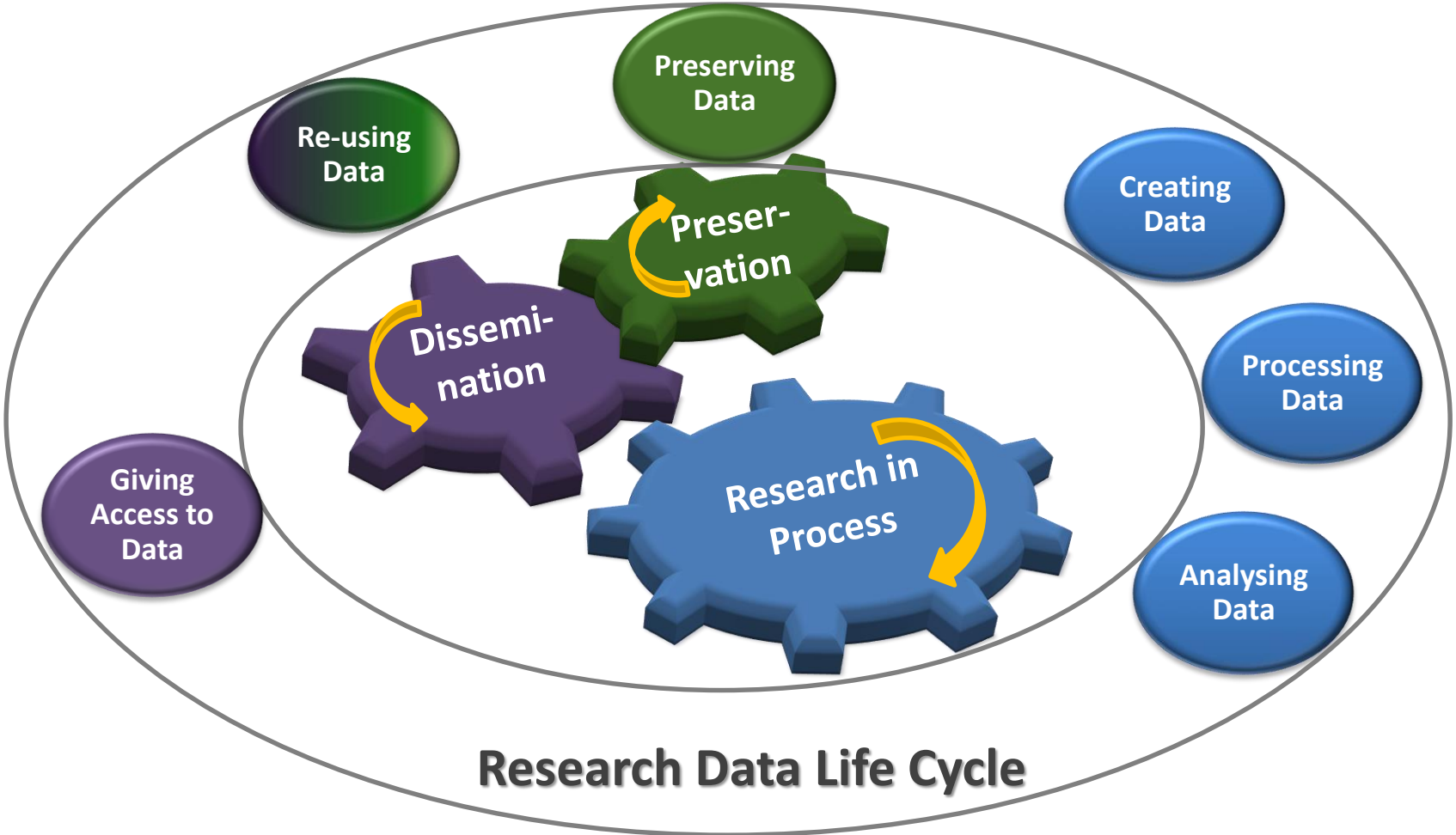
Preservin g Data
- Data Archiving
- Data Preservation
- Metadata Creation
- Link data to outputs

Giving Access to Data
- Data Sharing
- Data Publishing
- Link Data to Outputs

Re-using Data
- Data Repurposing/Re-use
- Data Citation

Based on UK Data Archive Lifecycle

Processes within the Research Data Lifecycle



Data flow with the Research Data Lifecycle

Investigation of Research Data Repository software solutions

Collaborative project between UP Library Services and IT Services

Finalisation of product evaluation criteria

- Consulted with various UP stakeholders to obtain their input. Consulted with external stakeholders at the NEDICC workshop held at the CSIR;
- Consulted with peer Universities; and
- Utilised **various selection criteria** from other institutions e.g. Leeds University, Texas Digital Library and the RDA RPRD IG Matrix (http://tinyurl.com/RPRD-matrix) selection criteria as a basis and adapted it according to UP specific requirements.

Product Short Listing

- Product scan of products being used internationally; and
- Most commonly used products at universities similar to UP (size and research activity).

Investigation of Research Data Repository software solutions

Product Evaluation

- UP’s formal Request For Information (RFI) process was followed.
- Product evaluation criteria list was compiled and send to short listed vendors together with standard RFI documentation.
- The requested information was received from the vendors and prepared for scoring.
- Products were scored and evaluated.

Investigation of Research Data Repository software solutions

Evaluation Criteria

- Functional / Business criteria: Deposit and Upload; Re-Usability; Identity and Access Management; Reporting; Discovery; Preservation
- Non Functional: Repository Architecture; Data Management; Data Governance
- Technical aspects: Back-end Management; Integration; Infrastructure
- Vendor specific: Support, Training, Usage of Product
- Performance requirements
- Integration requirements

Shortlisted products

Product	Vendor / Implementation Partner	RFI Feedback
DSpace	Atmire	Received information on criteria list, proposed implementation options and its associated cost.
Figshare	Digital Science	Received information on criteria list, proposed implementation options and its associated cost.
Islandora	Discoverygarden	Received information on criteria list, proposed implementation options and its associated cost.
Dataverse	Harvard University	Received insufficient information on criteria list, implementation options and cost.
PURR	Purdue University	Failed to respond to RFI.
Redbox	Queensland Cyber Infrastructure Foundation (QCIF)	Received information on criteria list, but Redbox is **only a meta data repository** and **not a data repository**.

Product Evaluation Results

Criteria	Figshare	Islandora	DSpace
BEEEE	All products and associated vendors/implementation partners are internationally based , therefore no weight was assigned in the scoring exercise.		
Requirements Criteria (incl functional, non-functional, vendor)	85% fit	96% fit	65% fit
Pricing	CONFIDENTIAL		
Preferential criteria: Hybrid Option (option 2)	100% Fit	10% fit – only available through huge custom development which poses huge risks to UP.	0% Fit
Preferential criteria: Consortial pricing	100% Fit	0% fit	0% fit

Next steps

- Investigate tools that can support the Research-in- Process phase, e.g. myTardis, Hubzero
- Finalise storage solution (e.g. African Research Cloud)
- Business Case to secure resources (financial and human)
- Implementation of repository solution
- Training of researchers / Library staff

NeDICC https://nedicc.com/

Network of Data and Information Curation Communities (NeDICC) aims to promote the development and use of standards and best practices to ensure the usability of digital material (e.g. research data, digital objects) in support of e-Research in South Africa

Membership

- Founding Members: University of Pretoria, NRF, HSRC, CSIR, and ASSAf
- Open to all Higher Education Institutions and Research Councils

Steering Committee - 2018

NeDICC
Network of Data & Information Curation Communities

- Mr Niklas Zimmer (UCT) (Chair)
- Ms Fatima Darries (UNISA)
- Mr Johann van Wyk (UP)
- Co-opted members:
    - Dr Martie van Deventer (lecturer, Dept of Information Science – formerly from CSIR)
    - Dr Heila Pienaar (UP)
    - Ms Louise Patterton (CSIR)

Aims of NeDICC

NeDICC
Network of Data & Information Curation Communities

- A forum to exchange experience and opinions
- Promote communication and co-operation between members through meetings, seminars, workshops and conferences and collaborative projects to:
    - Address issues of interest / concern
    - Promote awareness and best practices
    - Showcase work and initiatives
    - Expose the community to new developments and trends
    - Provide opportunities to engage with a wider audience
    - Active participation in forums (DIRISA, CODATA, RDA, etc.)
    - Develop the knowledge and skills of members
    - Develop shared tools

NeDICC Activities

NeDICC
Network of Data & Information Curation Communities

Carnegie CPD Programme

Enhancing librarians' ICT skills for research enablement in African universities: a Carnegie CPD programme

The Department of Information Science, University of Pretoria, was awarded a grant by the Carnegie Corporation of New York to offer eight four-week residential continuing professional development (CPD) programmes for qualified librarians from Sub-Saharan Africa, for the period 2014 – 2016.

The aims of the programme are to:

- Empower the next generation of library and information professionals within Sub-Saharan African countries with hands-on skills to apply current and emerging information and communication technologies (ICTs);
- Build capacity amongst library and information services (LIS) professionals to support and enable researchers.

NeDICC/DCC RDM Workshop participants,
12 August 2015, Knowledge Commons, CSIR (Pretoria)

NeDICC in collaboration with the North-West University and Talarify
invites you to participate in
the first **Library Carpentry Workshop** in Africa
lead by Prof Cameron Macdonell from MacEwan University, Canada.

The workshop is aimed at Librarians interested to learn skills relevant to the Library of the Future:

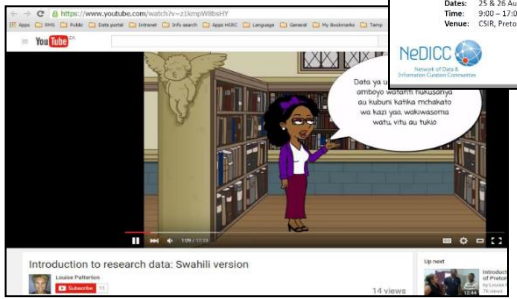
- Understanding the international Library Carpentry landscape
- Tools for cleaning data – OpenRefine
- Tools for collaboration – Git and GitHub

The workshop will also include a lesson sprint to familiarise participants with lesson design and contribution to allow further development of materials relevant to the South African Library Community

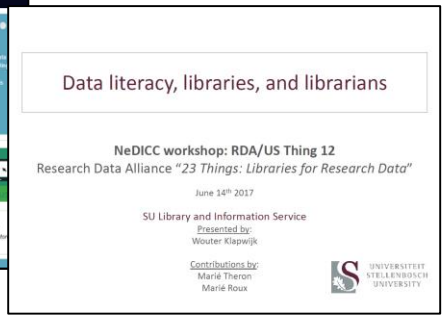
No prior programming experience is required. Everyone is welcome. Limited space available.

Dates: 25 & 26 August
Time: 9:00 – 17:00 daily
Venue: CSIR, Pretoria Campus
Cost: R1000 p/p (VAT excluded)
Registration: http://goo.gl/forms/BQTLLtn8x1tekTIl2
Enquiries: mvandeve AT csir DOT co DOT za

NeDICC | talarify



NeDICC Workshop for rookie data managers



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

NeDICC Programme 2017

NeDICC
Network of Data & Information Curation Communities

NeDICC Programme 2017

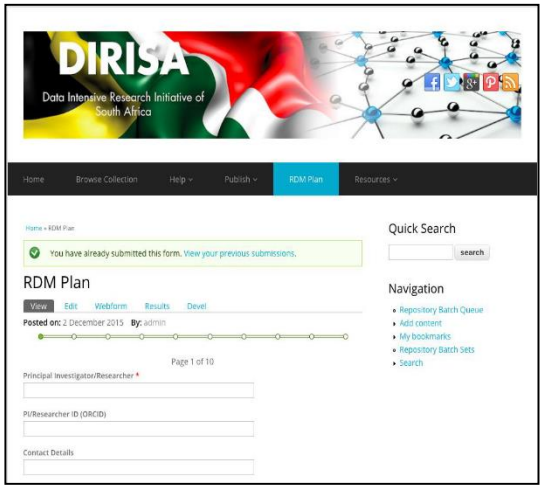
Date	Activity	Topic
26 January 2017 14:00 – 16:00	CoP Meeting	The implementation of an Institutional Data Repository (UCT: Kayleigh Roos, Erika Mias) View the report View the presentation slides View the cost evaluation
15 February 2017	NeDICC Workshop	Long-lived Data: Tools to Preserve Research Data (UP: Johann van Wyk and Isak van der Walt)
14 June 2017	NeDICC Workshop	Data Information Literacy (US: Wouter Klapwijk)
19 July 2017	NeDICC Workshop	Data Discovery and Metadata (UNISA: Makaba Macanda)
16 August 2017 14:00 – 16:00	CoP Meeting	
13 September 2017 9:00 – 14:00	NeDICC Workshop	
18 October 2017	NeDICC Workshop	Trustworthy Data Repository Certification (CSIR, HSRC)
16 November 2017 14:00 – 16:00	CoP Meeting	

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

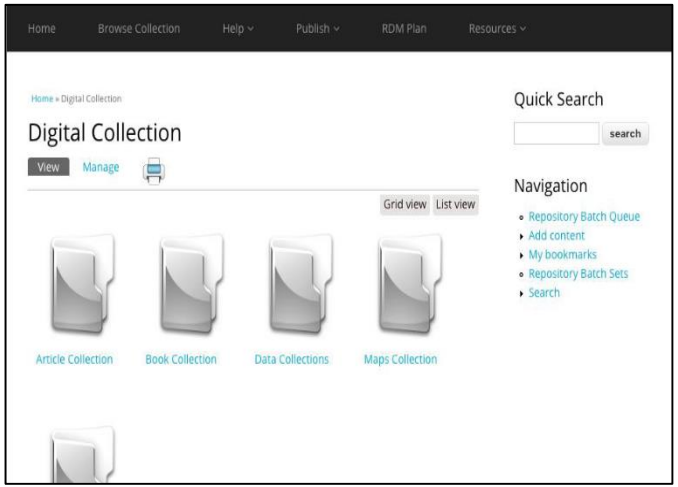
Collaborative Projects

Collaborative Projects between NeDICC and DIRISA

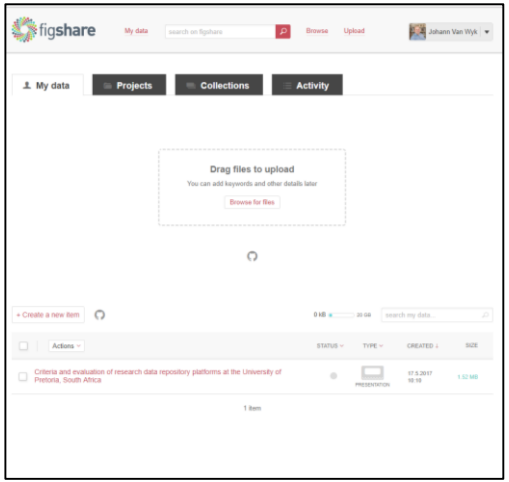
Data Management Planning Tool



Islandora Data Repository



Figshare pilot project



Thank You

Johann van Wyk
E-mail: johann.vanwyk@up.ac.za
Tel: 012-4205651