UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Facilitate Research Data Management (RDM) - Project

Strategic Feedback Meeting – 25 May 2017 - Department of Library Services

Johann van Wyk

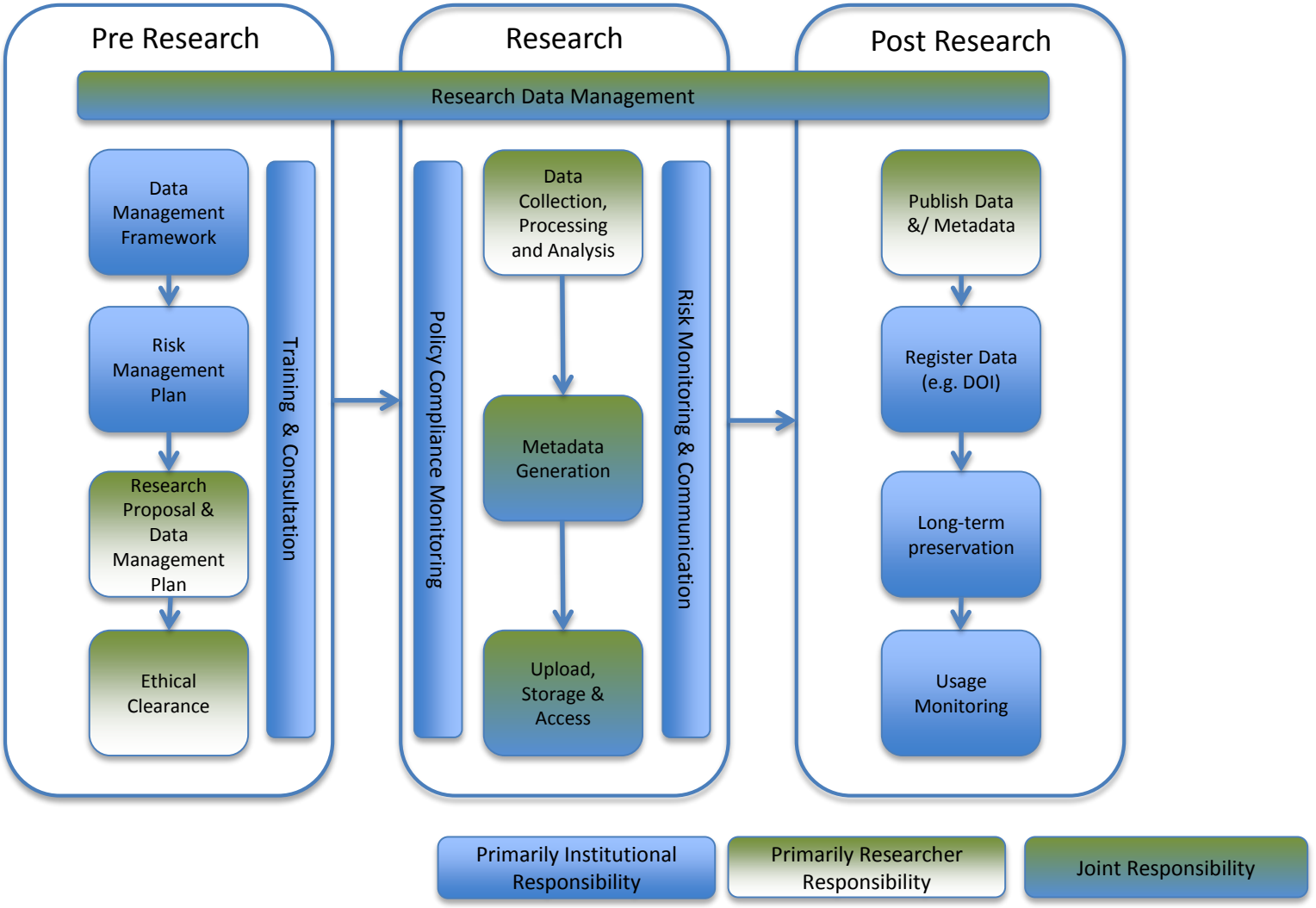
UP Library RDM road map 2016-

Objective / RDM services	Outcome	Actions	Milestones	Target date	Roles & responsibilities (who will do what)
An updated UP RDM policy and guidelines are available	RDM policy approved by UP Senate and Executive	• Investigate national and international RDM policies • Formulate and edit UP RDM policy • Finalise UP RDM policy	• Report completed • Implement DVC Research feedback • Approval by UP Executive	• Done • Done • Current	Johann van Wyk & Prof S. Burton
Training in dmp tool for researchers & librarians (In collaboration with DIRISA)	Librarians & researchers are able to use the dmp tool optimally	• Hands-on training sessions	• All Information Specialists are trained in DMP tool • Some researchers trained in DMP tool	• May 2016 • Ongoing	Johann van Wyk & Information Specialists
A platform for the publishing of data is available	Provide a platform for the publishing of data	• Investigate, test and evaluate different platforms • Pilot chosen platform (In collaboration with DIRISA) • Implement RDM platform for UP	• Done • Pilot successfully completed • Working RDM platform for UP implemented	• Done • ? • Currently in process	• Isak van der Walt • Anwar Vahed (DIRISA), Wim Hugo (SAEON, NRF) • RDM Project Team (UP IT & Library Services)

UP Library RDM road map 2016-

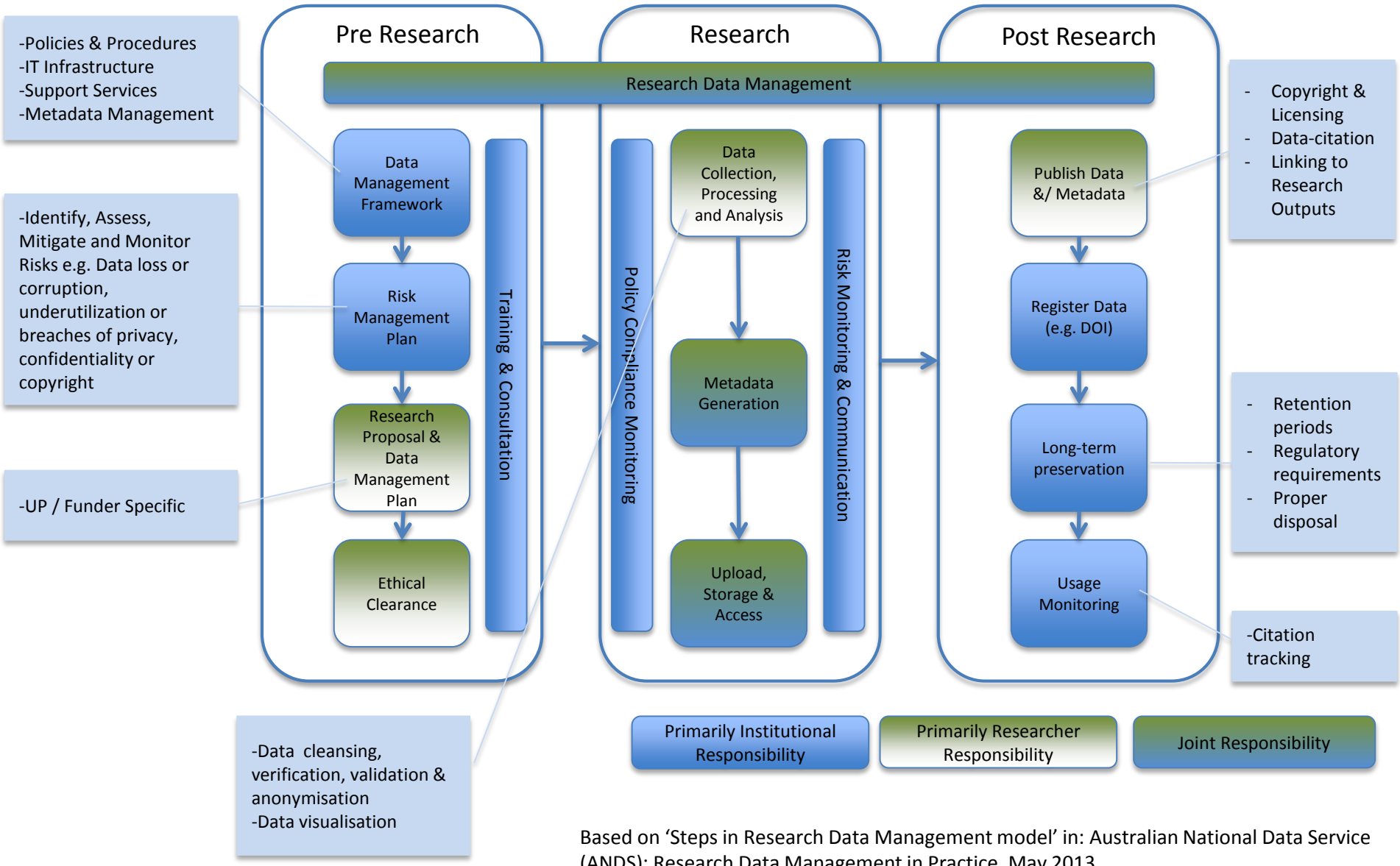
Provide meta data guidance					
Train researchers to make data ready for deposit					
Train researchers how to deposit their data					
Provide advice on data copyright					
Provide training on data citation	Librarians are able to cite data correctly	• Public UP lecture • Hands-on-training	• An expert will give a UP public lecture • Hands-on-training session will be developed	• May 2016 • End 2017	• Martie van Deventer (CSIR) • Information Specialists

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

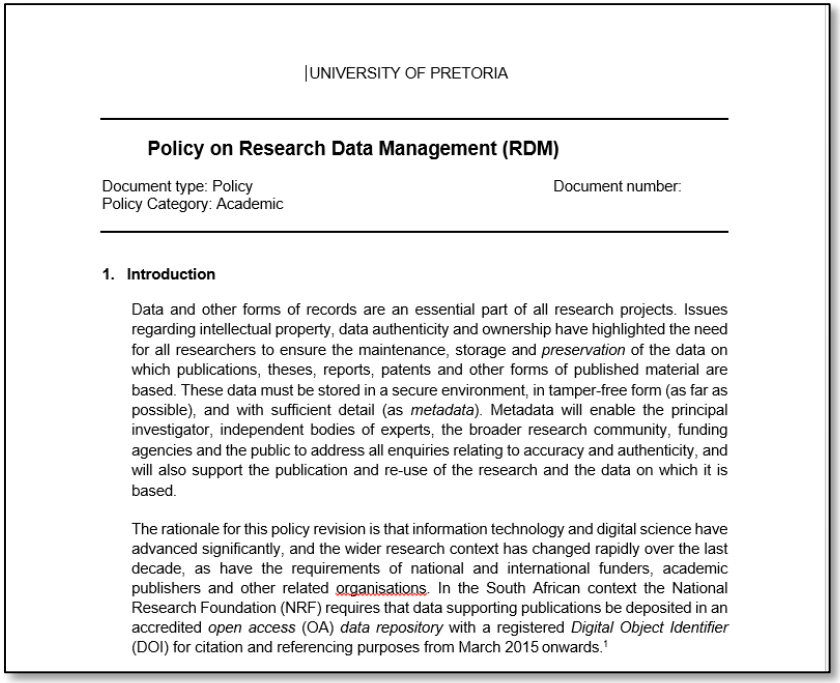
Key Steps in Research Data Management (more detail)



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

UP RDM policy and guidelines

- The final version of the updated UP RDM policy was sent through to the University Executive on 26 January 2017
- Waiting for approval.
- Approval is dependent on the roll-out of a data repository platform for the University



UNIVERSITY OF PRETORIA

Policy on Research Data Management (RDM)

Document type: Policy
Policy Category: Academic

Document number:

1. Introduction

Data and other forms of records are an essential part of all research projects. Issues regarding intellectual property, data authenticity and ownership have highlighted the need for all researchers to ensure the maintenance, storage and *preservation* of the data on which publications, theses, reports, patents and other forms of published material are based. These data must be stored in a secure environment, in tamper-free form (as far as possible), and with sufficient detail (as *metadata*). Metadata will enable the principal investigator, independent bodies of experts, the broader research community, funding agencies and the public to address all enquiries relating to accuracy and authenticity, and will also support the publication and re-use of the research and the data on which it is based.

The rationale for this policy revision is that information technology and digital science have advanced significantly, and the wider research context has changed rapidly over the last decade, as have the requirements of national and international funders, academic publishers and other related organisations. In the South African context the National Research Foundation (NRF) requires that data supporting publications be deposited in an accredited *open access* (OA) *data repository* with a registered *Digital Object Identifier* (DOI) for citation and referencing purposes from March 2015 onwards.[1]

Training of library staff and researchers

2016

- UP through NeDICC (Network of Data and Information Curation Communities) conducted a number of workshops:
  - The Life of research data and a roadmap to enable the implementation of services to support RDM
  - Data Management Planning (DMP)
  - Role of the information professional in RDM
  - Data Curation Profiles
  - Evaluation of data repository software
- In-house training
  - Library Carpentry initiative: Workshop on Data Cleaning using Openrefine (presented as part of OpenCon 2016)
  - Data Citation Principles Research Seminar presented by Dr Martie van Deventer (CSIR)
  - Role of the information professional in RDM (Merensky and Medical Library)
  - Data Management Plans (Merensky)

Training of library staff and researchers

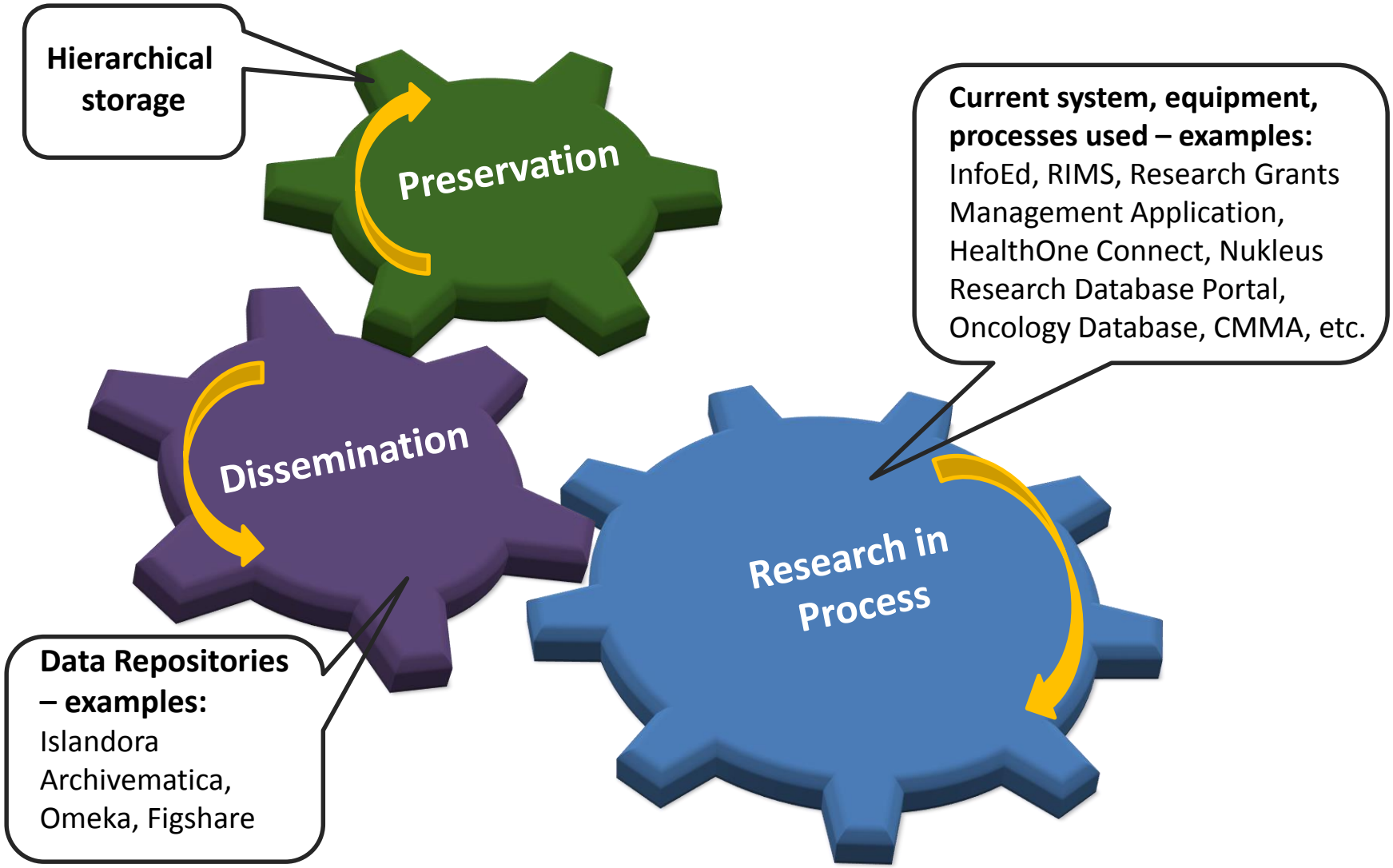
2017

- UP through NeDICC (Network of Data and Information Curation Communities) conducted a workshop:
  - Long-lived Data: Tools to Preserve Research Data
- Presentation at RDM Workshop at NWU University (Potch) - giving an overview of RDM at UP
- Presentations at eResearch Africa 2017:
  - Criteria and evaluation of research data repository platforms @ the University of Pretoria, South Africa
  - Mobilising a Nation: RDM Education & Training in South Africa
- Workshop presented at eResearch Africa 2017:
  - Library Carpentry Workshop in collaboration with UCT

Evaluation of research data repository platforms

- The scope of the project was to evaluate products (commercial and open source) which could be utilised as a **Research Data Repository Platform** as part of a total Research Data Management (RDM) solution at UP.
- A total RDM solution include all phases of the Research data life cycle, but for the repository solution, the focus was thus on identifying a potential solution for the “Dissemination” phase of the research data life cycle.

DATA FLOW within the RESEARCH DATA LIFE CYCLE



Product Investigation Methodology

Finalisation of product evaluation criteria

- Consulted with various UP stakeholders to obtain their input (Library and ITS staff)
- Consulted with external stakeholders at the NEDICC workshop held at the CSIR
- Consulted with peer Universities (e.g. Australia and SA), 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

Products were short listed based on the following:

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

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, and
- Products were scored and evaluated.

RDM Repository Project Team

Business Sponsor – Prof Stephanie Burton (VP: Research)

ITS Sponsor – Andre Kleynhans (Deputy Director: ITS)

Project Team members:

ITS Project Manager and Business Analyst – Karin Meyer

ITS Infrastructure Architect - Dr Yzelle Roets

ITS eResearch Support Manager – Herman Jacobs

Library Services: Senior IT Consultant – Isak van der Walt

Library Services: Assistant Director: RDM – Johann van Wyk

Library Services: Deputy Director: Strategic Innovation – Dr Heila Pienaar

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 & RFI Feedback

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**.

Implementation options

Options	Description
Option 1	Locally hosted (both application and storage are locally hosted at UP)
Option 2	Hybrid (application is cloud hosted, while the storage is locally hosted)
Option 3	Fully cloud-based (both the application and storage are cloud hosted through the vendor)

Product Evaluation Results

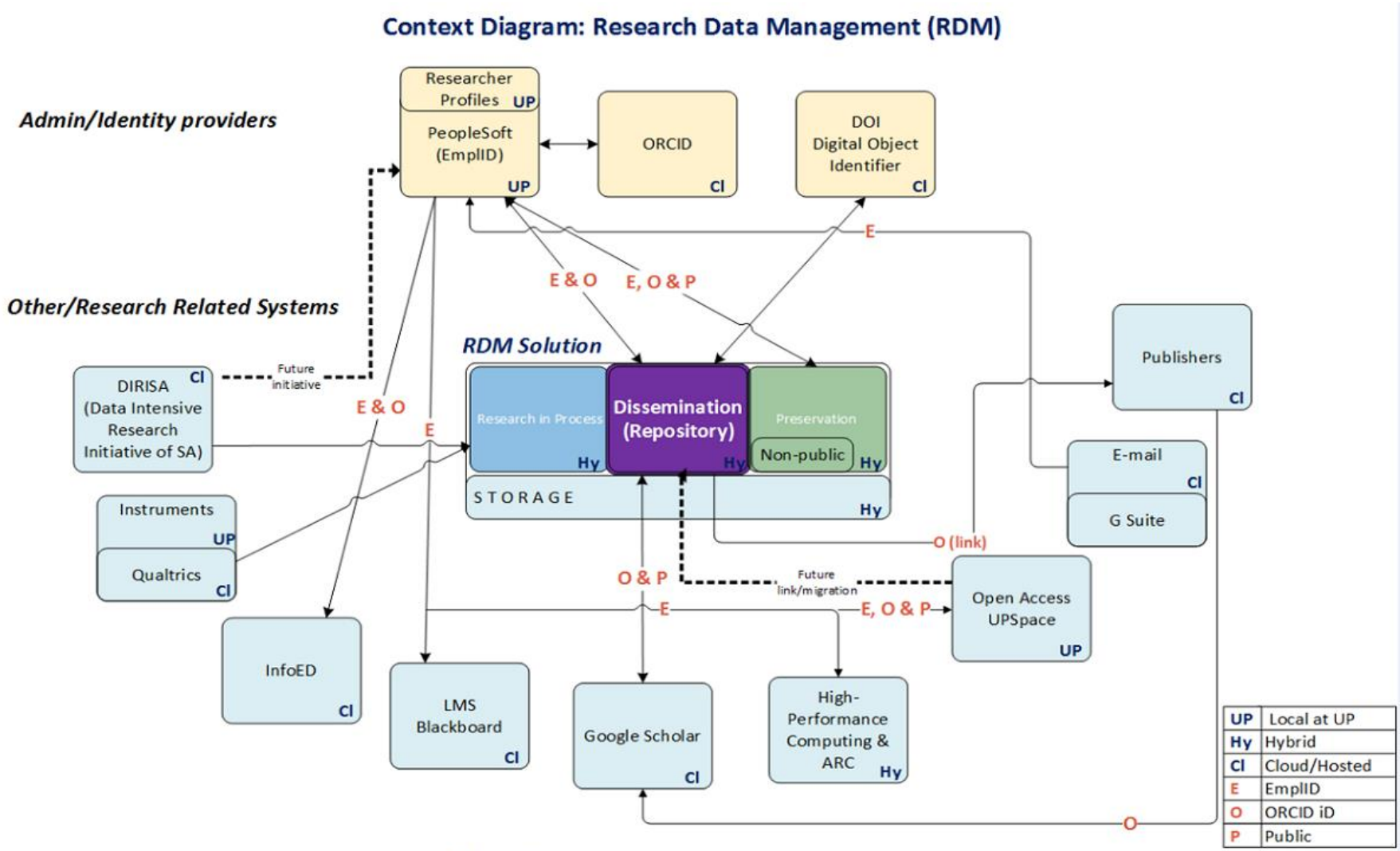
Criteria	Figshare	Islandora	DSpace
BEEEE	All products and associated vendors/implementation partners are internationally based , therefore no weight was assigned in the scoring exercise.		
Requirements Criteria (incl functional, non-functional, vendor)	85% fit	96% fit	65% fit
Pricing	CONFIDENTIAL		
Preferential 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

Recommendations

The following is recommended for implementing of a Research Data Repository platform) solution at UP:

- **Figshare** should be considered as the product of choice
- Implement the **Hybrid** implementation option with the application being cloud hosted and a local storage of 20Tb to start with
- Local storage can be supplemented in future with Cloud storage
- Storage should be investigated in line with the total eResearch initiative and framework of UP
- A business owner needs to be identified to be responsible for a total RDM implementation
- Implementation of a Research Data Repository platform requires a significant increase in Human and Infrastructure Resource components, and
- Consortial pricing can be kept in mind for the future and was not used as a determining selection criterion.

Context Diagram: Research Data Management



Data can be open for public access or embargoed for public use within the repository, preservation system, researcher profile, ORCID and DOI.

Documents sent through to UP Executive

The following documents were sent through to the UP executive on 24 February 2017:

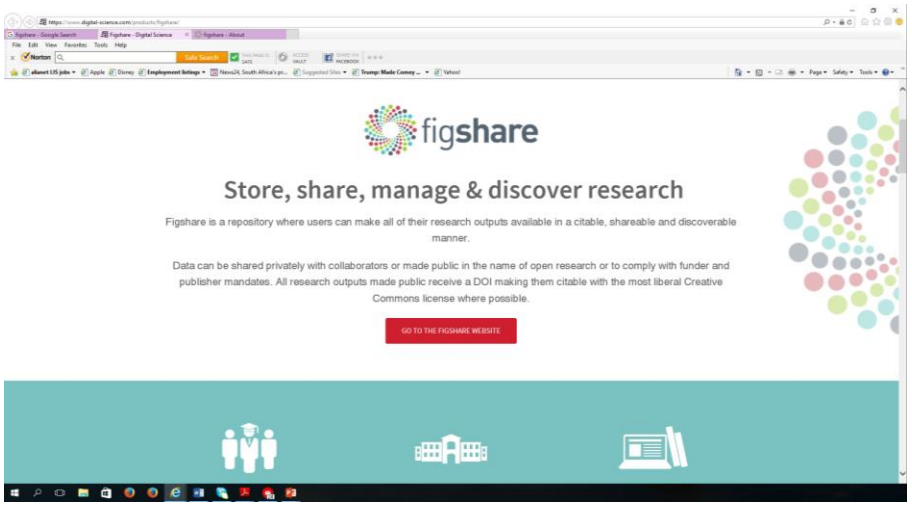
- UP Research Data Repository Evaluation
- UP Research Data Management Business Requirements Specification
- Executive summary
- RDM Project Progress Feedback
- Context Diagram for RDM
- Islandora, Figshare, Redbox, DSpace, Dataverse, PURR requirements criteria feedback documents

National Developments

- Investigation by DIRISA on possibility of implementing **Figshare** nationally

- Meeting held between DIRISA and Figshare

  Outcome:
  - Figshare will host a pilot Figshare for SA Institutions, using Amazon
  - SA Institutions are invited to trial this
  - 2 events will be held, one in Durban and one in Johannesburg where presentations will given by Figshare on the product. UP and UCT will also do presentations on the outcomes of their investigations of repository platforms. Representatives from HE institutions and councils will be invited.



National Developments

- After the 6 months trial period, the uptake will be assessed and a decision will be made on how Figshare will be implemented

- Several options on how Figshare can be implemented:
  (a) A large proportion of institutions adopt it: Tenet becomes the Figshare "service provider" and negotiates an additional cost recovery line item in the institution connectivity bill.
  (b) Some institutions adopt it and form a grouping that contracts Figshare.
  (c) No one... Unlikely

- DIRISA can provide storage

Still a lot of ground to cover