Libraries enabling Open Data: UCT's new Institutional Data Repository

powered by Figshare for Institutions

Wednesday, 6th December 2017
CHPC annual meeting, Pretoria

Niklas Zimmer

Manager
Digital Library Services, UCT Libraries
University of Cape Town
(some) Germans on this day, Dec 6\textsuperscript{th}:
Outline

Advocacy; Drivers; Solutions; Features; Considerations
Outline

- **Advocacy** *(some of our public engagements with some of the stakeholders)*
  - Roadshow, Info-sessions, Naming competition, Launch event, Presentations

- **Drivers** *(Why we are implementing OA IDRs at SA HEIs)*
  - Mandates, Requirements and Policies for Open Access Data
  - Benefits of Open Data for Open Science

- **Solutions** *(Identifying and supporting a suitable mechanism for compliance)*
  - Suggesting an Institutional Data Repository for UCT (2016)
  - Supporting a federated, national approach
  - UCT IDR statement

- **Features** *(Figshare use cases; UCT site)*
  - What is Figshare? Some examples of Figshare applications
  - UCT IDR online repository (live), and support pages

- **Considerations** *(Technical, operational, and conceptual deliberations)*
  - General clarifications – Figshare as a tool, a service, and a company
  - Initial QAs with the Researcher community

Wednesday, 6th December 2017
Advocacy

Looking back at some of our public engagements with stakeholders
FIGSHARE Pilot in Pretoria and Durban

Posted on July 19, 2017

Data Intensive Research Initiative of South Africa (DIRISA), aims to ensure the long persistence and preservation of the data on local infrastructure. As part of this, DIRISA enrolls on regional meetings to introduce Figshare to research institutions and academia. Figshare were organised by DIRISA on 4 July 2017 in Pretoria and 6 July 2017 in Durban. The aim of these roadshows was to show South African researchers and their institutions how Figshare works. Figshare is a web-based platform to help academic institutions, publishers and researchers to disseminate and measure the public attention of all their research outputs. All research output on Figshare are made available in a citable, shareable and discoverable manner.

The following presentations were made:

- Mr Anwar Vahed (DIRISA) – Introduction to the proposed national approach to research data management
- Mr Niklas Zimmer (University of Cape Town) – Practical applications for a research data management system
- Mr Isak Van de Walt (University of Pretoria) – Criteria and evaluation of repository platforms
- Mr Mark Hahnel (Figshare) – Introduction To Figshare
FIGSHARE FOR INSTITUTIONS: AN INTRODUCTION TO A NEW INSTITUTIONAL DATA REPOSITORY

NIKLAS ZIMMER

Digital Library Services, University of Cape Town Libraries

20 September 2017 Wednesday at 13:00
Beattie 240
Upper Campus, UCT

Join the history crowd at

Department of historical Studies
University of Cape Town
September 2017: IDR Naming competition

Help UCT Libraries NAME the University of Cape Town’s new Institutional Data Repository and stand a chance to win a UCT goodie bag with items up to the value of R1000.

Some of its key benefits:

- Get DOIs for your research outputs
- Comply with NRF and other funder requirements
- Control how your research outputs are accessed
- Validate and authenticate your research outputs
- Increase citations and boost your research metrics

For more information on how to enter, go to the link below, or scan the QR code:


Have any questions?
Contact us: dls@uct.ac.za
Visit our website: digitalservices.lib.uct.ac.za
Niklas Zimmer @NikasZimmer · Oct 2
Presenting to Chelsa (S.A. university libraries directors) today, on enabling Open Data & Science with Institutional Data Repositories.
The UCT IDR is to be called…

ZivaHub
Open Data UCT

Submitted by:
Thandeka Chehore

CONGRATULATIONS!

Tuesday, 24 October 2017

October 2017: Announcing the winner

Help UCT Libraries NAME The University of Cape Town’s new Institutional Data Repository and stand a chance to win a UCT goodie bag with items up to the value of R1000.

Some key benefits:
- Get DOIs for your research outputs
- Comply with NRF and other funder requirements
- Control how your research outputs are accessed
- Validate and authenticate your research outputs
- Increase citations and boost your research metrics

For more information on how to enter, go to the link below, or scan the QR code:


Have any questions?
Contact us: info@uct.ac.za
Visit our website: digitalservices.lib.uct.ac.za
November 2017: Official launch of ZivaHub: Open Data UCT

University of Cape Town
The Institute of Infectious Disease & Molecular Medicine
invites you to attend a

Seminar on **Wednesday, 22nd November at 13h00-14h00**
Wolfson Pavilion Lecture Theatre

---

**Nov 22nd: Presentation to Health Sciences, UCT**

**Niklas Zimmer**
UCT Libraries

"**Introducing ZivaHub: Open Data UCT**"

“Learn to make your research data citable, shareable and compliant”

**Host: Dr Yolande Harley [IDM]**

**ZivaHub** is an Open Access Data repository available to all students and staff at UCT. It provides access to the supplementary research data that inform scholarly outputs hosted on other platforms, such as OpenUCT. ZivaHub allows you to make your research data citable, shareable, discoverable, and reusable. Some of its key features are:

- a. Get DOIs for your research outputs
- b. Comply with NRF and other funder requirements
- c. Control how your research outputs are accessed
- d. Validate and authenticate your research outputs
- e. Increase citations and boost your research metrics

"OUR MISSION is to be an outstanding teaching and research university, educating for life and addressing the challenges facing society"
ZivaHub: Open Data UCT - User session with Figshare representatives

Digital Library Services, in collaboration with Figshare, invites you to a practical workshop session on using the new institutional data repository, ZivaHub: Open Data UCT, powered by Figshare for Institutions.

Figshare's Engagement Manager, Megan Hardeman, and Implementation Manager, Mimi Keshani, will be leading the session, and are visiting UCT for one day only, so this is a great opportunity for you to discuss your research data requirements and find out more about how ZivaHub might support your data management and sharing needs.
We warmly invite you to bring your own research data along for assistance with uploading to the platform.

Date: Tuesday 7 November, 2017
Time: 10am-12pm
Place: Ulwazi, Vincent Kolbe Knowledge Commons, UCT Chancellor Oppenheimer Library
Requirements: Please bring a lap top with you if you have one
RSVP: http://uct.ac.za.libcal.com/event/3717948
Drivers

Why we are implementing Open Access Institutional Data Repositories at South African Higher Education Institutions
“Carrots and Sticks” …

Compliance with mandates & policies

Benefits to researchers & research

Wednesday, 6th December 2017
Why we are implementing OA IDRs at SA HEIs

Increase in mandates, requirements, and policies

- **Funders:**
  1) NRF statement (2015)
  2) Gates Foundation, Mellon et al
don’t want to repeat-fund research

- **Journals:**
  1) pre-publication: peer-review of data
  2) post-review: data-as-publication

- **Institutions:**
  1) OA and RDM policies
  2) New IP developments (‘first to publish’)
  3) Supervision process, including data

Wednesday, 6th December 2017
Drivers for implementing OA IDRs at SA HEIs

- **International:**
  - Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)
  - Cape Town Open Education Declaration (2007)
  - Journal requirements (e.g.: Springer Nature, PLOS journals, IOP journals, et al)
  - Open Data Accord (Open Data in a Big Data World)
  - The FAIR Guiding Principles for scientific data management and stewardship (2016)

- **National:**
  - Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research (2015) [DLS summary available here]

- **Institutional:**
  - UCT Open Access (OA) Policy (2014)
  - UCT Research Data Management (RDM) Policy (draft, for UCT senate executive end 2017)

- **Departmental:**
  - UCT Libraries ‘Horizon 2019’ Strategic Plan (2014)
  - Suggesting an Institutional Data Repository for the University of Cape Town (2016)
Why we are implementing OA IDRs at SA HEIs

Benefits of Open Data sharing / Open Science (researcher perspective)

- **Get DOIs for research outputs**
  1) *Institution-specific* dois [minted by DataCite through *Figshare for Institutions*]
  2) proof of IP (i.e. ‘first to publish’ rulings)

- **Control how research outputs are accessed**
  1) Non-public deposit (embargo) provides record of existence; can stipulate access modes/details
  2) Various licencing options (CC-BY et al)

- **Validate and authenticate research outputs**

- **Increase citations and boost research metrics**
  1) Figshare part of Digital Science, like Altmetrics

Wednesday, 6th December 2017
Benefits of Open Data sharing / Open Science (research perspective)

- **Quality and integrity**
  replication and validation of research results

- **Innovation and knowledge transfer**
  swifter path from research to innovation

- **Efficiency**
  transparency and democratic control

- **Public disclosure and engagement**
  public participation; impact measurement of policies

- **Economic benefits**
  self-empowerment; improved services

- **Global benefits**
  patterns in large data volumes

adapted from:
  a. Open Science at the core of Libraries: What are the benefits of Open Science?
     Available: [https://www.fosteropenscience.eu/content/what-are-benefits-open-science](https://www.fosteropenscience.eu/content/what-are-benefits-open-science). accessed: 21.06.2017
  b. Open Data handbook: Why Open Data?
Examples & Excerpts (funder)

Bill & Melinda Gates Foundation

‘Data Underlying Published Research Results Will Be Accessible and Open Immediately. The foundation will require that data underlying the published research results be immediately accessible and open. This too is subject to the transition period and a 12-month embargo may be applied.’

(Bill & Melinda Gates Foundation Open Access Policy)

> DLS outreach Q&A: Which funders mandate data sharing in your specific discipline?

Wednesday, 6th December 2017
Examples & Excerpts (journal)

PLOS

‘Data Availability: The data underlying the findings of research published in PLOS journals must be made publicly available. Rare exceptions may apply and must be agreed to with the Editor. Data should be de-identified where appropriate (see Human Subjects and Animal Research).’

(PLOS journals Editorial and Publishing Policies)

> DLS outreach Q&A: Which journals mandate data sharing in your specific discipline?
Examples & Excerpts (policy)

UCT RDM Policy (2017, draft for UCT Senate executive end 2017)

The University of Cape Town is currently drafting a Research Data Management (RDM) Policy. Read the latest draft. At UCT, the drivers and principles for managing research data are emerging in response to a number of policies published by funders of research, which include:

- ensuring the validation of research results
- providing research opportunities in data reuse
- enabling actionable and socially-beneficial science to address global research challenges.

Making data resulting from publicly-funded research open access requires consideration of the necessary limits on openness. The UCT RDM Policy will assist researchers in complying with legal requirements and emerging terms of funding and scholarly publishing regarding personal information and commercial considerations.

Full policy information pack:

- Advocacy & Outreach Programme: agenda outline
- Current UCT RDM Policy draft
- Open scholarship and funder mandates
- NRF Open Access Statement (2015)
- NRF Requirements for processing grantholder-linked and free standing student and postdoctoral support
- Requirements for archiving NRF funded theses/dissertations
- RDM clauses for inclusion in faculty MoU
An Open Access IDR enables researchers & institutions to make research data

- Discoverable
- Accessible
- Intelligible
- Assessable
- Interoperable
- Citable
- Shareable
- Re-usable

... in other words: ‘FAIR’

See:

Wednesday, 6th December 2017
The FAIR Guiding Principles for scientific data management and stewardship

- **To be Findable:**
  - F1. (meta)data are assigned a globally unique and persistent identifier
  - F2. data are described with rich metadata (defined by R1 below)
  - F3. metadata clearly and explicitly include the identifier of the data it describes
  - F4. (meta)data are registered or indexed in a searchable resource

- **To be Accessible:**
  - A1. (meta)data are retrievable by their identifier using a standardized comms protocol
  - A1.1. protocol is open, free, and universally implementable
  - A1.2. protocol allows for an authentication and authorization procedure, when necessary
  - A2. metadata are accessible, even when the data are no longer available

- **To be Interoperable:**
  - I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - I2. (meta)data use vocabularies that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data

- **To be Reusable:**
  - R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards
Solutions

Identifying and supporting a suitable mechanism for compliance, and a platform to provide the benefits of OA Data
Identifying and supporting a suitable mechanism for compliance, and a platform to provide the benefits of OA Data

Suggesting an IDR for UCT (2016)

- An extensive review of data repository platforms was conducted by DLS for UCT eResearch, which included CKAN, DSpace, Dataverse, DRYAD, Fedora, Figshare, and TIND. It is openly accessible on the UCT Zenodo community: Suggesting an Institutional Data Repository for the University of Cape Town (2016)

- The evaluation compared open source and licensed options, and (NB) took into consideration local infrastructure support staffing costs.
- The cost of Figshare at consortium pricing is similar to hosting any alternative Open Source option, so the recommended solution became Figshare for Institutions.
Supporting a federated, national approach

- **Budgets** are limited (austerity measures), and there is **not enough IT staff** at most SA HEIs (Faculties, Libraries, IT services) to host independent, local solutions, including **Open Source platforms** (NB: these are **not free**, requiring ongoing needs analysis, comparison, liaison, customisation, development, maintenance, upgrades, migration, …).

- **Individual institutional licenses** for **SaaS** are more expensive than federated licensing, and Figshare’s partnership with **DataCite** enables **minting of institution-specific dois**.

- Instead, org structures require urgent review to embed **new, specialised skills**:
  - **Research Data Management** (Funder mandates; Advocacy; Support services, …)
  - **Digital Curation** (Data archiving & access; Metadata schemas; …)
  - **(Research) Data Librarianship** (Data locating and acquisition; Re-use; Citation; …)
  - **Data analysis, mining, and visualisation** support (Digital Humanities; GIS; …)
  - **Repository management** (coding; Digital library infrastructure; Semantic web; …)
  - **Digital Scholarship** (Liaison; Open Scholarship; R&D; Digital Humanities; …)

Wednesday, 6th December 2017
The UCT Institutional Data Repository (IDR) serves principal investigators at the University of Cape Town who are in need of a repository to store and openly disseminate the data that support their published research findings. The repository service is provided in terms of the UCT Research Data Management Policy. It provides open access to supplementary research data files and links to their respective scholarly publications (e.g. theses, dissertations, papers et al) hosted on other platforms, such as OpenUCT.

UCT IDR mission statement

Wednesday, 6th December 2017
Features

Figshare use cases; UCT instance
What is Figshare?

Figshare for institutions:

‘Figshare is a web-based platform to help academic institutions manage, disseminate and measure the public attention of all their research outputs. The light-touch and user-friendly approach focuses on four key areas: research data management, reporting and statistics, research data dissemination and administrative control. Figshare [helps institutions] meet key funder recommendations and to provide world-leading tools to support an open culture of data sharing and collaboration.’

See:
1. DCC. Resources: Figshare. Accessible: http://www.dcc.ac.uk/resources/external/figshare
Why not just use the free *Figshare.com* platform?

1. **More storage:**
   
   figshare.com users only get 20GBs, without the ability to purchase extra space; the same applies to project spaces.

2. **Curation (at various levels, e.g. groups):**
   
   a) ensures that sensitive data isn't accidentally published  
   b) offers review to ensure the metadata is thorough  
   c) provides additional, optional RDM support (from Libraries)

3. **Showcase department data with group-specific branding** (see: [MAMU](#))
4. Admin functionality for impersonations:
   PIs as admins can manage a project on behalf of their head

5. Institutional control and curation of data:
   Data is owned by the institution after researchers leave

6. Custom licenses (figshare.com only offers CC-BY and CC-0)

7. Influence product development:
   - custom metadata fields
   - feedback from end users goes directly to product manager

Why not just use the free Figshare.com platform? (contd.)
Examples of different *Figshare for Institutions* applications

- **Thesis repository:**
  - Monash University
  - Loughborough University

- **Customised researcher engagement:**
  - Cranfield University

- **Special Collections dissemination platform:**
  - Music Archive of Monash University

- **Custom interfaces and data visualisation services:**
  - Oxford University
  - St Edwards University

- **Journal OA repository for (supplementary) data:**
  - PLOS One (world’s first multidisciplinary Open Access journal)
  - IOP (Institute of Physics) Publishing
Examples of *Figshare for Institutions* applications (contd.)

- **Figshare for institutions:**
  - Videos, webinars and brochures

- **Example case studies:**
  - The impact of sharing supplementary datasets (...)
  - The value of a DOI and an open platform for sharing data
  - Visualising research and the addition of supplementary material as a complete research output
  - Tips for early career researchers on making data available (...)

---

*UCT Libraries*
**UCT IDR:** ZivaHub [Open Data UCT]

**UCT ZivaHub:** start-up guide

**DLS info:** ‘Sharing my data’
Considerations

Technical, operational, and conceptual deliberations
General clarifications – Figshare as a tool, a service, and a company

- Storage and management of so-called ‘raw’, e.g. unprocessed data is the responsibility of the researcher, not of the Libraries (DLS: ‘Data Storage’)

- Researchers upload (self-publish) their processed data and/or links to it on Figshare, accepting the terms provided (UCT terms | Figshare user guide)

- With the ‘curation’ function enabled in Figshare (the tool), institutions can dedicate admin staff to check upload-submissions for metadata, licensing etc. (Figshare: ‘Curation and Review’)

Wednesday, 6th December 2017
General clarifications – Figshare as a tool, a service, and a company

• As part of the pilot offering, Figshare (the company) provides some ‘start up’ cloud storage on Amazon cloud services. (UCT: 100TB overall, 20GB per user/group).

• Data storage and backup is the responsibility of each institution. Figshare (the service) merely manage the metadata – where the data is stored is irrelevant to the service (> APIs).

• Provision of national storage is envisaged by DIRISA within the next 3 years.

• ‘Nesting’ / filtering of Figshare instances allows institution-, discipline-and researcher- specific (meta)data sets across all tiers (GRID id)
Q&As with the researcher community

- **What constitutes (supplementary) ‘data’?**
  - investigate & communicate (NB: discipline-specific) differences between ‘raw’ and ‘processed’ data

- **Is deposit applicable to publicly funded research only?**
  - encourage rather than enforce: make new offerings in data management, sharing and discovery, e.g.: RAiD, Yewno, etc.

- **How is compliance measured?**
  - develop integrations with GMS (PeopleSoft), RMS (Converis), ORCiD, DMPTool, etc., i.e. ‘IDs for everything’

- **What are the consequences when requirements (policy, funder mandate, journal etc.) are not met?**
  - withholding of funding, publication, et al.
Technical, operational, and conceptual deliberations

Q&As with the researcher community (contd.)

- **How is the ethics approval managed?**
  - arrange with respective ethics committees to own / administer relevant data management policy types (examples and support: Springer Nature Data Policies)
  - RDM policy <> data sharing <> ethics clearance forms and process

- **How are data shared that were generated using (licensed) software?**
  - curation (normalization); containerisation; integration with GitHub, etc.

- **Where to find relevant examples of metadata schemas to describe data (sets)?**
  - training of (sub-)admins (feedback / ratification: a new metadata working group?)
  - Community-driven co-development in this area: Figmeta

- **What is the envisaged typical life cycle of data (the practical meaning of ‘in perpetuity’)?**
  - ...

Wednesday, 6th December 2017
What types of research data do you produce? e.g.
- Datasets? [NB: GitHub integration]
- Code?
- Posters, Presentations?
- Images, 3D files, ...?
- Video, Audio, ...?

How may you need to describe them?
Questions from DLS | UCT Libraries for researchers | groups:

- How may you need to **affiliate** ('nest') your data?
  
  1. Faculty
  2. Department
  3. Project
Thank You