UCT LIBRARIES

# Practical applications for an RDM system

## Getting on the road with Figshare

Monday, 03 July 2017: Cape Town, UCT
Tuesday, 04 July 2017: Pretoria, CSIR
Thursday, 06 July 2017: Durban, ICC


Niklas Zimmer

**Manager**

Digital Library Services, UCT Libraries
University of Cape Town



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

UCT LIBRARIES

# Outline

Drivers; Solutions; Features; Considerations; Implementation; References

UCT LIBRARIES

# Outline

- **Drivers** (*Why we are implementing OA IDRs at SA HEIs*)
  - Policies (international; national; institutional; departmental)
  - Publisher / journal requirements for open access
  - Benefits of Open Science and Open Data sharing (Open Data Accord, et al)
- **Solutions** (*Identifying and supporting a suitable mechanism for compliance*)
  - Identifying and supporting a suitable mechanism for compliance
  - Suggesting an Institutional Data Repository for the University of Cape Town (2016)
  - Supporting a federated, national approach
- **Features** (*Figshare use cases; UCT test site*)
  - What is Figshare?
  - Examples of Figshare applications
  - UCT test site
- **Considerations** (*Technical, operational, and conceptual deliberations*)
  - Initial QAs with the Researcher community
  - Some keywords towards decolonial concerns
- **Implementation** (*Setting up Figshare at our institution*)
  - Storage; Quota; Users and Groups; Org structure; Terms of deposit; Metadata; …
  - Next Steps
- **References**

Monday, 03 July 2017

UCT LIBRARIES

# Drivers

## Why we are implementing Open Access Institutional Data Repositories at South African Higher Education Institutions

UCT LIBRARIES

# 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)
  - G20 Anti-Corruption Open Data Principles (2015)
  - 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
  - Intellectual Property Rights from Publicly Financed Research and Development Act (2008)
  - Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research (2015) [DLS summary available here]
- Institutional
  - UCT Intellectual Property (IP) Policy (2011)
  - 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)

UCT LIBRARIES

# Excerpts

## 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

Monday, 03 July 2017

UCT LIBRARIES

## Excerpts (contd.)

**Journal requirements (example 1):**

**PLOS**

I. '*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)

UCT LIBRARIES

# Open Data in a Big Data World (ICSU)

**Maintaining self-correction**
Openness of the evidence (the data) for scientific claims is the bedrock of scientific progress. It permits the logic of an argument to be scrutinised and the reproducibility of observations or experiments to be tested, thereby supporting or invalidating those claims. […] Data must be intelligently open, meaning that they should be: discoverable, accessible, intelligible, assessable and (re-)usable.

**Adapting scientific reasoning**
Many of the complex relationships that we now seek to capture through big- or broad-, linked data lie far beyond the analytical power of many classical statistical methods. […] The complexity of patterns that machines are able to identify are not easily grasped by human cognitive processes, posing profound issues about the human-machine interface and what it might mean to be a researcher in the 21st century.

**Ethical constraints**
The open data principle has ethical implications for researchers and research subjects. […] In a regime of open sharing in which data are passed on from their originators, there is loss of control over future usage, whilst anonymisation procedures have been demonstrated to be unable to guarantee the security of personal records.

**Open global participation**
Big data and open data have great potential to benefit less affluent countries, and especially least developed countries (LDCs). […] Thus, both emerging and developed nations have a clear, direct interest in helping to fully mobilize LDC science potential and thereby to contribute to achievement of the UN Sustainable Development Goals.

See: Open Data in a Big Data World. ICSU. Accessible: https://www.icsu.org/publications/open-data-in-a-big-data-world

Monday, 03 July 2017

UCT LIBRARIES

# Benefits of Open Science and Open Data sharing

**Quality and integrity**: wider evaluation and scrutiny by the scientific community (replication and validation of research results).

**Innovation and knowledge transfer**: reduces delays in re-use of the results of scientific research (swifter path from research to innovation).

**Efficiency**: improved effectiveness and productivity of the research system, and of (non-) government services (transparency and democratic control).

**Public disclosure and engagement**: promotes awareness among citizens (public participation; impact measurement of policies).

**Economic benefits**: fosters innovation, and increases awareness and conscious choices among consumers (self-empowerment; improved services).

**Global benefits:** promotes collaborative efforts and faster knowledge transfer on international concerns (e.g. climate change: combining data sources and 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. accessed: 21.06.2017

b. Open Data handbook: *Why Open Data?* Available: http://opendatahandbook.org/guide/en/why-open-data/. Accessed: 22.06.2017

UCT LIBRARIES

# Solutions

Identifying and supporting a suitable mechanism for compliance

## Identifying and supporting a suitable mechanism for compliance

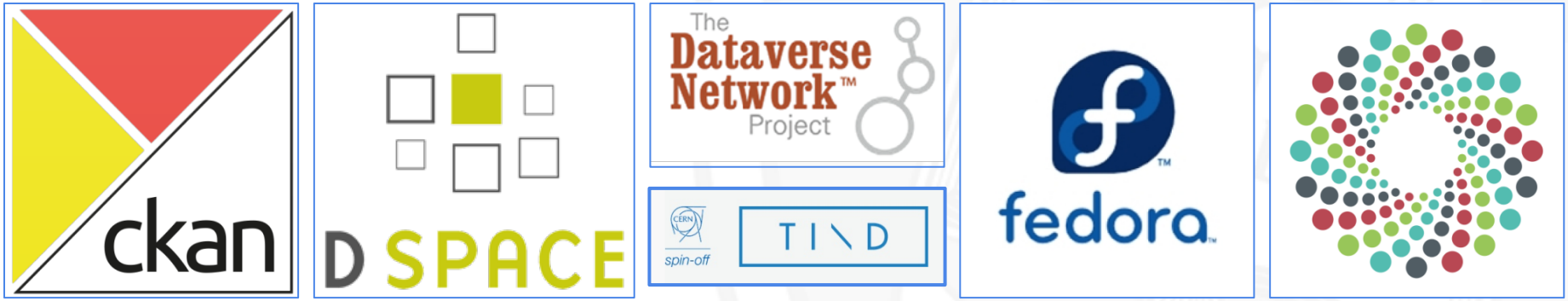
UCT LIBRARIES


Source: Pampel, Heinz (2013): How to find an appropriate research data repository? PLOS blogs, PLOS Tech. Available: http://blogs.plos.org/tech/how-to-find-an-appropriate-research-data-repository/

UCT LIBRARIES

## Suggesting an IDR for UCT (2016)

- An **extensive evaluation process** of data repository platforms was conducted by DLS for UCT eResearch, which included **CKAN, DSpace, DataVerse, DRYAD, Fedora, Figshare,** and **TIND**. It is available 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 took into consideration **local infrastructure support staffing costs**.
- The recommended (SaaS) solution was Figshare.

UCT LIBRARIES

## 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 'free', e.g. Open Source** platforms (requiring needs analysis, comparison, liaison, implementation, customisation, development, maintenance, upgrades, migration, …).
- **Individual institutional licences** for **SaaS** are **more expensive than federated licensing**, and Figshare's partnership with DataCite enables **minting of institution-specific dois**.
- Rather, 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; …)

UCT LIBRARIES

# Features

Figshare use cases; UCT test site

UCT LIBRARIES

# 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.’

Adapted from: (Hyndman, Alan (2017): figshare for institutions. figshare. https://doi.org/10.6084/m9.figshare.c.3582206.v5. Retrieved: 12:43, Jun 12, 2017)

UCT LIBRARIES

# Examples of Figshare 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

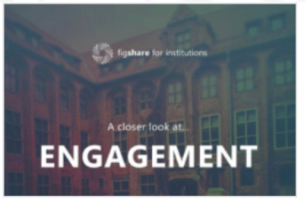
UCT LIBRARIES

# Examples of Figshare applications (contd.)

- **Figshare for institutions**:
  - Videos, webinars and brochures
- **Figshare case studies**:
  - Various use cases for different disciplines

Sharing ra
research |
Megan Harc

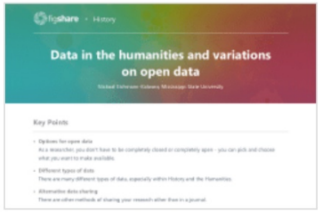

Using figshare's API to customise your data repository: A case study...
Jeremy Cope 12/05/2017



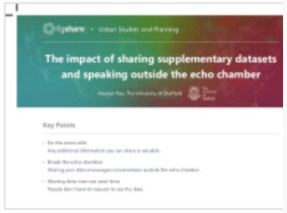

A Closer Look At Engagement
Georgina Parsons 05/05/2017

State of Open Data Poste
Alan Hyndman


Digital Scholarship and the future of cross-disciplinary work in researc...
Megan Hardeman 15/12/2016




Data in the humanities and variations on open data || History
Megan Hardeman 10/08/2016




Print-ready figshare case studies
Megan Hardeman 13/03/

figshare from the researcher perspective
Alan Hyndman 25/01/2017

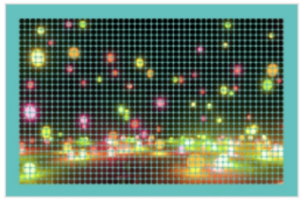
figshare & Oxford University live data project - Supporting research...
Alan Hyndman 23/02/2017


figmeta presentation
Simon Porter


Making research open and accessible || Physiological Science


Spreading the word about scientists' experience with science || Science ...


The importance of historical divergence time between languag



The State of Open Data Report
Jon Treadway 25/10/2016

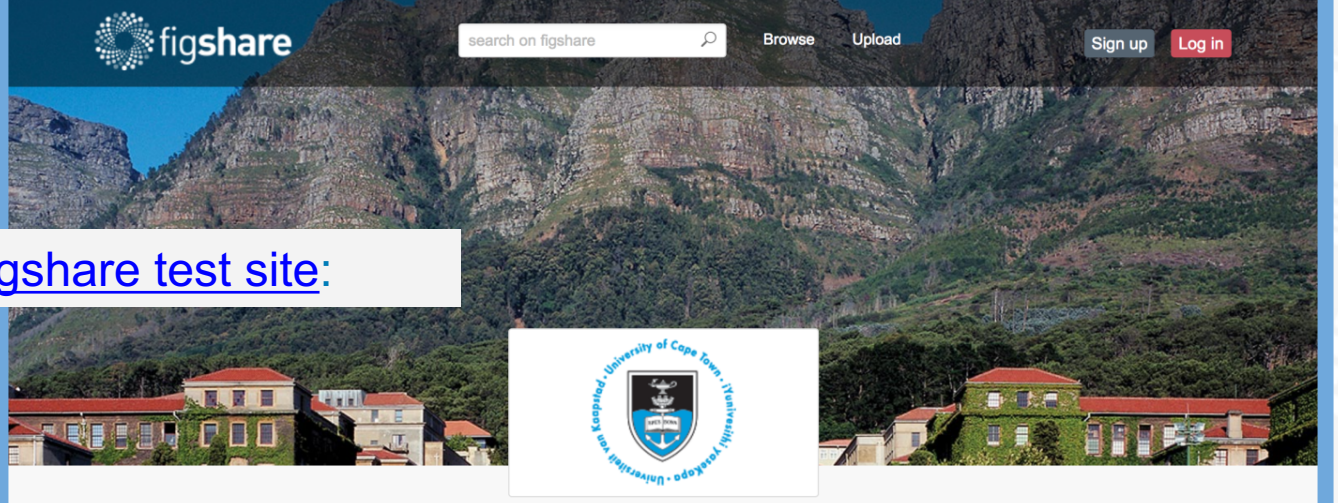


Introducing figshare for institutions
Adrian Harja 28/10/2016


figshare for institutions
Alan Hyndman

UCT LIBRARIES

[UCT Figshare test site](): 




Discover research from University of Cape Town

NEW POPULAR CATEGORIES SEARCH

## Astronomy, Astrophysics, Space Science

DATASET

**Summary of BLAST search query to identify intimin subtypes.**
Katie S. Viljoen 09/03/2015

DATASET

**Quantification of bacteria in colorectal cancer and adjacent nor…**
Katie S. Viljoen 09/03/2015

DATASET

**Clinicopathological characteristics of the cohort of fresh-frozen tissue…**
Katie S. Viljoen 09/03/2015

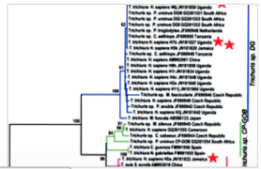
DATASET

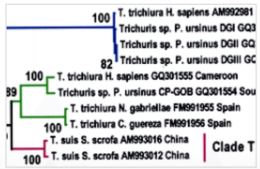
**Estimates of bacterial genome copies per nanogram of bacterial D…**
Katie S. Viljoen 09/03/2015

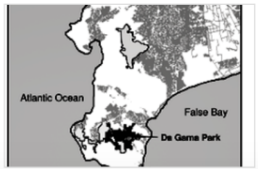
More items from Astronomy, Astrophysics, Space Science

## Biological Sciences

DATASET





UCT LIBRARIES

# Considerations

Technical, operational, and conceptual deliberations

UCT LIBRARIES

## Initial Q&As with the researcher community

- *What constitutes (supplementary) 'data'?*
  - investigate & communicate 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.

- *What are the consequences when requirements (policy, funder mandate, journal etc.) are not met?*
  - withholding of funding, publication, etc.

UCT LIBRARIES

## Initial 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)

- *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?)

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

UCT LIBRARIES

## Some keywords towards decolonial concerns

**Accessibility; Agency; Authority; Consent; Curriculum; Decentering; Delinking; Digital divide; Eurocentrism; Epistemological violence; Globalism; Hegemony; Ideology; Imperialism; Intersectionality; Inclusivity; Perception; Representation; Transnationalism;**

- In this age where data is the new oil, what role might this OA data sharing platform play in addressing the massively complex, **contemporary challenge for African institutions** to '*undo centuries of knowledge extraction*'?
- Since every tool [**NB: institutional archive**] is invested (more or less obviously) with hegemonic assumptions and motives, how do we envisage the plurality of our users' access to and control over **(self-)representation in terms of gender, language, class, race**, etc?
- While we are involved in institutionalizing / operationalizing the means to deposit and share our data worldwide, are we yet sufficiently aware of our comparative limitations in interrogating this data ourselves? How do we **grow and sustain our resources to create added value from (our) data** [analysis, reuse, prototype applications, ...]?
- How do we **establish a meaningful, decolonial dialogue** between: a) Protection (IP, patenting, monetization, grants), and: b) Openness (Open Access, Open Scholarship)?

Monday, 03 July 2017

UCT LIBRARIES

# Implementation

## Setting up Figshare at our institution

UCT LIBRARIES

## Implementation topics

- **Storage**:
  a. Swift connection to ARC, using Keystone authentication.
     Two containers: public and private access to objects (up-/download systems).
  **b. Data sovereignty** concerns:
     i. Now: (UCT) Test & Pilot in Cloud (Figshare)
     ii. Future: Live implementations in national stack (DIRISA)

- Set up **Users** and **Groups**:
  a. SAFIRE integration (rather than harvesting from HR admin system feed):
     i. what detail to include, e.g. organizational affiliation, projects?
     ii. multidisciplinary, or inter-departmental affiliation on Figshare roadmap

- Mimic **Organizational structure**:
  a. Nesting and linking *Groups* and *Collections* into virtual org structure
  b. Departmental branding: banners & logos (embedded info-box, e.g.: CORD)

UCT LIBRARIES

## Implementation topics (contd.)

- Draft **UCT Terms of [Data] Deposit** (*Draft: Andrew Bailey: 2017-06-05*):
  a. Enable compulsory acceptance of specific (CC) terms for upload to repository, relevant/specific to institution: Research Contracts and Innovation; Research Office; Ethics committees; further stakeholders
  b. Customise default CC licenses (CC-O default under discussion)

- Customise/add **Batch upload:**
  a. Possible via the figshare API, see: docs.figshare.com/api.
  b. Adding multiple items from a large collection (with a metadata spreadsheet and link the metadata to each file via the identifiers and DC title fields?):
     i. one-by-one, or
     ii. as an entire dataset with a single generic metadata description. Is there functionality for an admin to batch ingest a collection of files
  c. Investigate API for batch uploads on back-end? One file at-a-time, or multiple files with one metadata record (MAMU example)

UCT LIBRARIES

## Implementation topics (contd.)

- **Administration of Data** per department:
  a. Workshop, describe and share new workflows and documents
  b. Identify curators across campus:
     i. create and maintain subject-specific metadata schemas; metadata QA and feedback; approval of deposits; linking to thesis repository; …

- **Metadata fields / schema:**
  Figshare OAI-PMH endpoint currently only exposes the default figshare fields as DC fields (e.g. UCT Libraries' collections require full Dublin Core metadata fields).
  a. Collaborate?: **Update to harvestable extra fields in groups & collections**
  b. Investigate: Figmeta - Community Driven Metadata for Figshare.
  c. Investigate: VIVO open source tool for describing research outputs (Figshare OAI-PMH API now also includes VIVO-RDF)

UCT LIBRARIES

# Next steps

- **Listing, description** and **indexing**:
  a. Registry of Research Data Repositories (re3data.org)
  b. Investigate ((S.) African) inclusion in research Graph (Research Data Switchboard) via VIVO
- **Standardisation**:
  a. DIN 31644
  b. ISO 16363
- **Certification** and **auditing**:
  a. Data Seal of Approval (DSA)
  b. World Data System (WDS)
- **Cooperation with Open Science initiatives**:
  a. DataCite
  b. OpenAIRE
- **Advocacy, outreach**:
  a. Community news feed with 'latest uploads' (e.g.: UK Data Service, ANDS)
  b. Commission and contribute Open Data Science examples to Figshare Case Studies

UCT LIBRARIES

# References

Resources cited; further reading

UCT LIBRARIES

# References


- International Council for Science. (2015?). *Open Data in a Big Data World: an International Accord*. Abbreviated version. Available: https://www.icsu.org/publications/open-data-in-a-big-data-world [2017, July 6].
- De la Fuente, G.B. (n.d.). *What are the Benefits of Open Science?*. Available: https://www.fosteropenscience.eu/content/what-are-benefits-open-science [2017, July 6].
- Open Knowledge International. (n.d.). Why Open Data? In: *Open Data Handbook*: Available: http://opendatahandbook.org/guide/en/why-open-data/ [2017, July 6].
- Pampel, H. (2013). How to find an appropriate research data repository? [Blog 4 November]. *PLOS Blogs Tech.* Available: http://blogs.plos.org/tech/how-to-find-an-appropriate-research-data-repository/ [2017, July 6].
- Hyndman, A. (2017). Figshare for Institutions. *Figshare.* https://doi.org/10.6084/m9.figshare.c.3582206.v5
- University of Cape Town. (2017). *Digital Library Services at UCT Libraries*. Available: http://www.digitalservices.lib.uct.ac.za/ [2017, July 6].
- *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*.(2003). Available: https://openaccess.mpg.de/Berlin-Declaration [2017, July 6].
- Cape Town Open Education Declaration (2007). Available: http://www.capetowndeclaration.org/read-the-declaration [2017, July 6].
- G20 Anti-corruption Open Data Principles (2015). Avaliable: http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf [2017, July 6].
- Open data and the fight against corruption in South Africa (2017). Available: https://www.transparency.org/whatwedo/publication/open_data_and_the_fight_against_corruption_in_south_africa [2017, July 6].


Monday, 03 July 2017

UCT LIBRARIES

# References


- OECD Principles and Guidelines for Access to Research Data from Public Funding. Available: http://www.oecdilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-publicfunding_9789264034020-en-fr [2017, July 6].
- Intellectual Property Rights from Publicly Financed Research and Development Act 2008 (Act No. 51 of 2008). Available: http://www.wipo.int/wipolex/en/details.jsp?id=6663 [2017, July 6].
- Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research. Available: https://updatalibrarian.files.wordpress.com/2015/02/nrf_open_access_statement_19jan2015.pdf [2017, July 6].
- UCT IP Policy. Available: http://www.uct.ac.za/downloads/uct.ac.za/about/policies/intellect_property.pdf [2017, July 6].
- University of Cape Town. (2014). University of Cape Town: Open Access Policy. Available: http://www.uct.ac.za/downloads/uct.ac.za/about/policies/UCTOpenAccessPolicy.pdf [2017, July 6].
- University of Cape Town. (2016a). *Distinguishing UCT: The University of Cape Town's Strategic Planning Framework 2016 - 2020:* Available: http://www.dsa.uct.ac.za/usr/about/governance/council/2016-12-10_UCTCouncil_StrategicPlanningFramework.pd [2017, July 6].
- University of Cape Town. (2016b). *University of Cape Town: Research Data Management Policy.* Draft Policy Document Version 4. Available: http://www.digitalservices.lib.uct.ac.za/dls/rdm-policy [2017, July 6].
- The effects of the IP regime on generating value from publicly funded research: an exploratory study of two South African universities" https://www.slideshare.net/CarolBN/2nd-globalcongresscbn-02 [2017, July 6].
- Hyndman, A. (2017). Figshare for Institutions. *Figshare.* https://doi.org/10.6084/m9.figshare.c.3582206.v5


Monday, 03 July 2017

UCT LIBRARIES

# References


- Hardeman, M. (2017). Figshare Case Studies. *Figshare.* https://doi.org/10.6084/m9.figshare.c.3282239.v16.
- Available: http://opendatahandbook.org/guide/en/why-open-data/ [2017, July 6].
- Available: http://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf [2017, July 6].
- Available: https://kicsforum.wordpress.com/2016/02/27/the-benefits-of-open-science/ [2017, July 6].
- Available: https://www.fosteropenscience.eu/content/what-are-benefits-open-science [2017, July 6].
- Available: http://www.huffingtonpost.com/ariel-smilowitz/unlocking-the-benefits-of_b_4844336.html [2017, July 6].
- Auerbach, J. (2017). *What a New University in Africa Is Doing to Decolonise Social Sciences.* Available: https://theconversation.com/what-a-new-university-in-africa-is-doing-to-decolonise-social-sciences-77181?platform=hootsuite [2017, July 6].
- Wolski, M.; Howard, L.; Richardson, J. (2017). A Trust Framework for Online Research Data Services. *Publications*. 5(2): 14.doi:10.3390/publications5020014.
- AUTHOR? (2003). *Protecting our Cultural Capital A research plan for the heritage sector Harriet Deacon, Sephai Mngqolo, Sandra. Prosalendis.* Pretoria: HRSC Press.

UCT LIBRARIES

# Thank You