



Research Data Management (RDM) with Digital Library Services (DLS)

an introduction

Tuesday, 7th May 2019, 13:00 - 14:00
Pathology Learning Centre

Niklas Zimmer

Manager
Digital Library Services

Sanjin Muftić

Digital Scholarship Specialist
Digital Library Services

ORCID
Connecting Research
and Researchers

Let's change
what we value
in research.



Sign
DORA



We provide **open**, online access to primary resources for teaching, learning and research at the University of Cape Town (UCT) through digitisation, **digital scholarship**, data curation and preservation services.

We subscribe to and support the practice of **Open Science**.

Source: DLS website: <http://www.digitalservices.lib.uct.ac.za/>

Open Science

Open Science is the movement to make scientific *research* (including publications, data, physical samples, and software) and its *dissemination* **accessible to all levels** of an inquiring society, amateur or professional.

Open Science is arguably simply doing proper science. Others are enabled to **collaborate and contribute**, since research data [...] and other research processes are **freely available**, under terms that enable **reuse, redistribution and reproduction** of the research and its underlying data and methods.

Source: Foster Open Science: (<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>)

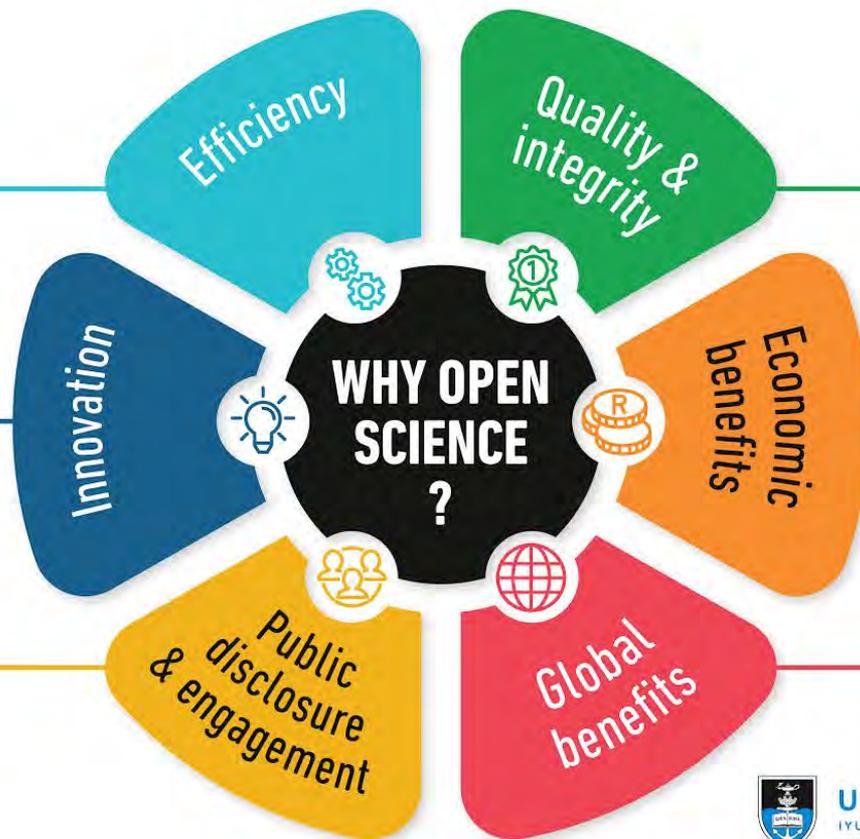
Adapted from: Woelfle, M.; Olliaro, P.; Todd, M. H. (2011). "Open science is a research accelerator". Nature Chemistry. 3 (10): 745–748. <https://doi.org/10.1038%2Fncchem.1149>

Open Science at UCT

Greater access to scientific inputs and outputs can increase scientific productivity through reducing duplication, allowing **more research from the same data** and multiplying opportunities for domestic and global participation in the research process.

Open science can **reduce delays in the re-use of scientific research** including articles and data, and promote a swifter path from research to innovation to produce new products and services.

Science, often publicly funded, should be publicly accessible to **promote a greater awareness** among citizens and to build public trust and support for public policies and investments in research. Open science also promotes citizen science in experiments and data collection.



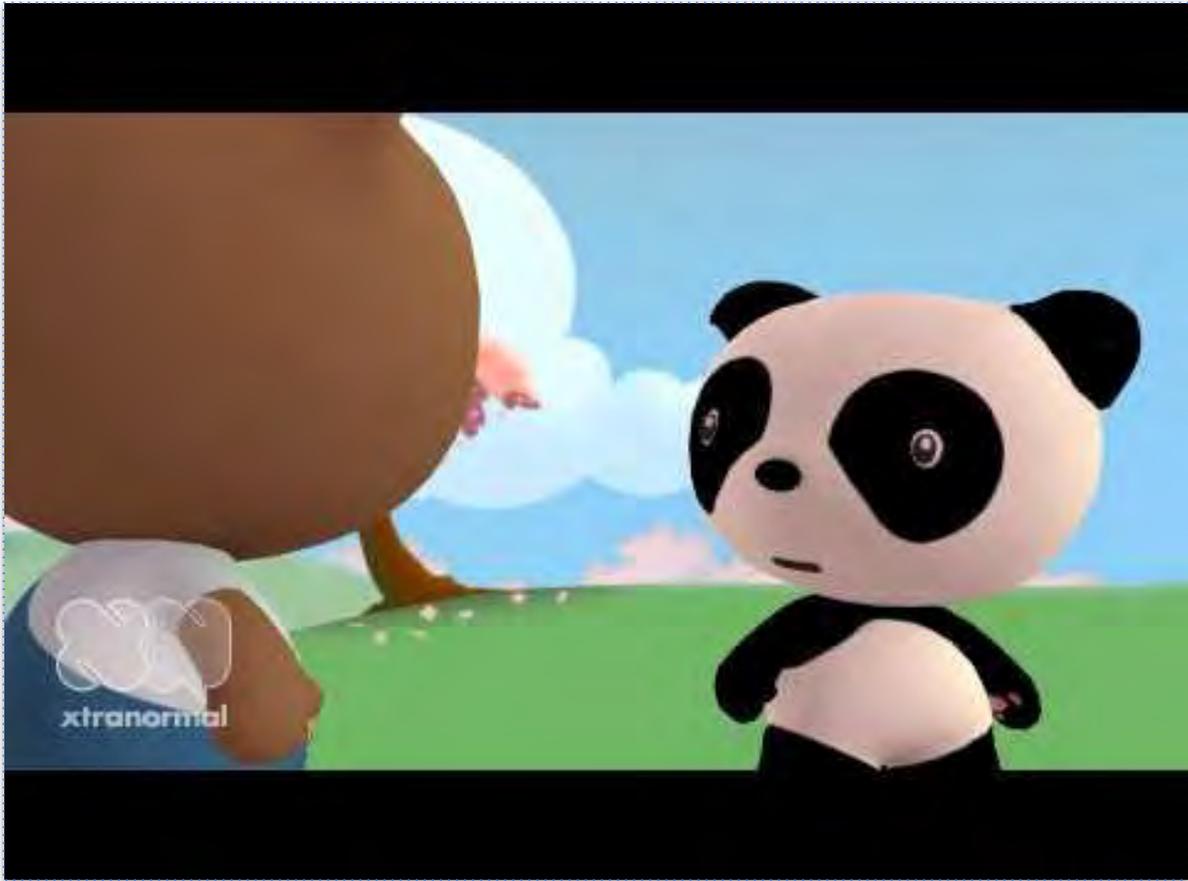
Open access to scientific outputs allow for **greater evaluation and scrutiny** by the scientific community which means more accurate replication and validation of research results. Openness to data contributes to maintain science's self-correction principle.

Science plays a key role in **today's knowledge economies** and increased access to research results, including data, can positive impact not only scientific systems but also innovation.

Open science promotes **collaborative efforts** and faster knowledge transfer for a better understanding of global challenges and wicked problems.



What is **closed** science?



Source: NYU Health Sciences Libraries. <https://youtu.be/N2zK3sAtr-4>



Research Data Management (RDM) towards Open Science *a brief overview*



Research data management (**RDM**)

- The **organisation and documentation** of the data processes (collection, description, de-identification, curation, archiving and publication) within a research project
- Already practised by researchers, but generally for internal use, and to varying degrees of professionalism
- International push towards **Open Science**, to professionalise data management practices and make them more coherent and shareable
 - Journals, institutions and funders increasingly stipulate that data be published alongside research outputs (reports, working papers, journal articles)

RDM at UCT

In March 2018, UCT implemented its **RDM policy** which – in line with an increasing number of funder demands – requires the (open) publication of research data produced by UCT researchers from master's level upwards.

Following this in 2019, a **new student MoU** (**Memorandum of Understanding**) was implemented for all postgraduate researchers that requires them to **create a DMP** (**Data Management Plan**) as part of their registration process.

Resources:

- [UCT RDM policy](#)
- [UCT DMPonline platform](#)

The research data management (RDM) lifecycle

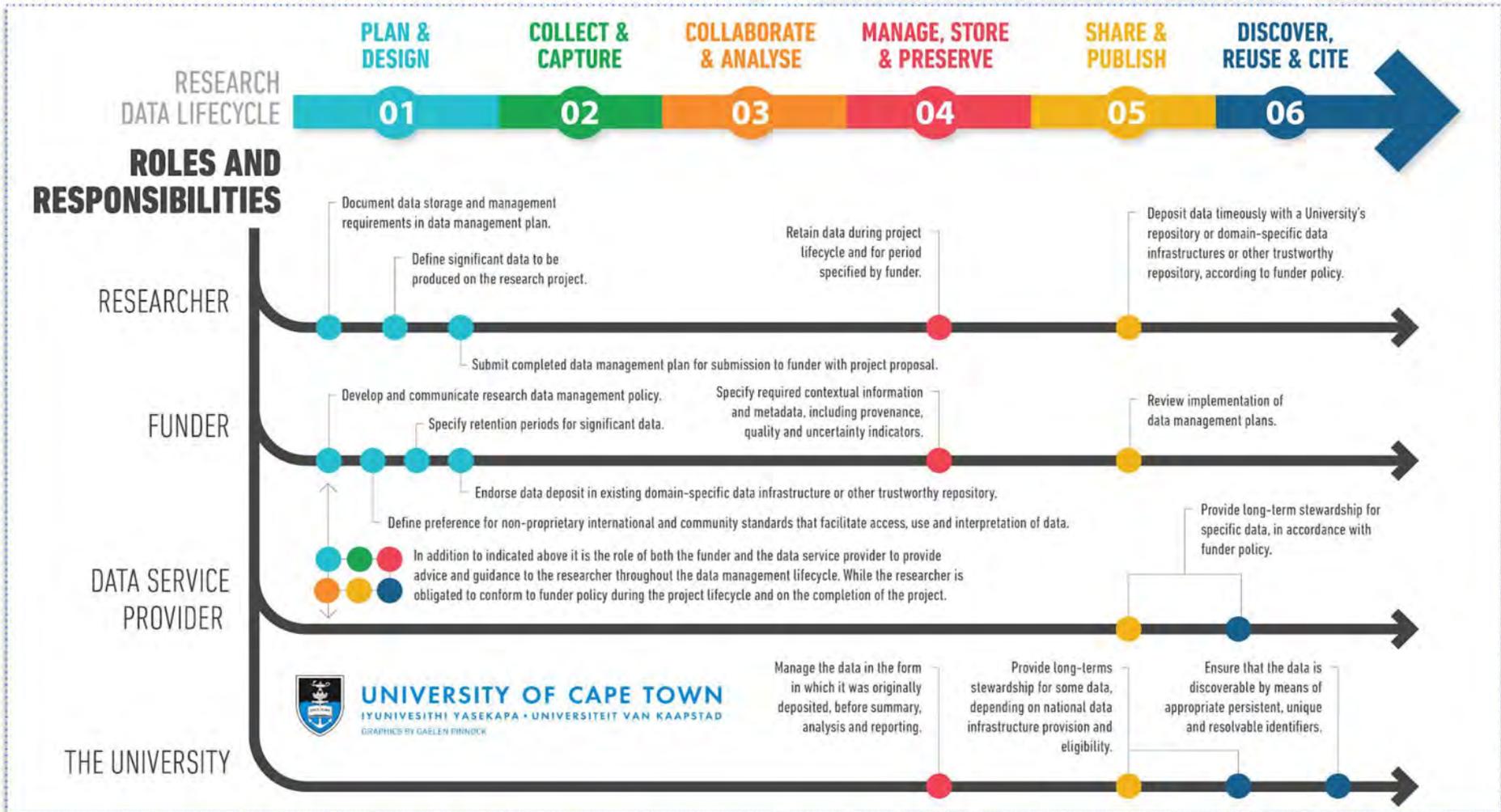


The research data management (RDM) lifecycle

How this presentation works



The RDM lifecycle at UCT



Source: UCT RDM Why Open Science: https://commons.wikimedia.org/wiki/File:UCT_RDM_Roles-and-Responsibilities.png



What is (your) data and why should you make it **reusable**?



“What (is my) data?”

- On-site server systems
- Storage protocols
- Cloud storage
- Hot, warm, cold storage
- Network transfer data
- Software
- HPC data
- Websites
- (...)

ICT data

Research data

- Microdata
- Unit record data
- Raw data
- Field data
- Experimental data
- Cleaned data
- Processed data
- Primary data
- Secondary data
- De-identified data
- (...)

“Data”

- Preservation data (SIPs, AIPs, DIPs)
- Library catalogues
- Metadata records
- (...)

Library data

Enterprise data

- HR data on Staff
- Student records
- Exam / test scores
- Student numbers
- Lists of researchers
- Financial data
- Budgets
- (...)

Data Types (examples)

Data Types

Community/Citizen

- SMS Mobile Data
- Email
- Surveys
- Social Media (Whatsapp, Facebook, Twitter, Instagram)
- Multimedia (Photos, Video, VR)

Government

- Census/Population
- Statistics
- Infrastructure
- Finance/Budgets/Spending
- Companies/Land Ownership
- Pollution Index/Water Quality

Physical

- Geographical
- Infrastructure

Sensor/New Tech

- Biometric
- Genetic (Crispr)
- Movement
- Meteorology
- Bitcoin

Aerial/Satellite

- Satellite
- Aerial/UAV
- Balloon Mapping



International Federation of Red Cross and Red Crescent Societies
 Fédération internationale des Sociétés de la Croix-Rouge et du Croissant-Rouge
 Federación Internacional de Sociedades de la Cruz Roja y de la Media Luna Roja
 الاتحاد الدولي لجمعيات الصليب الأحمر والهلال الأحمر

DATA PLAYBOOK: SLIDEDeck 1

Source: <https://www.oreparecenter.org/sites/default/files/slidedeck1whatisdata120618.pdf>

Data vs. Information

Data is an element which has not yet been interpreted and put into context. Examples:

- Fuel consumed by RC trucker on 2nd February: 60l
- Distance traveled by trucker on 3rd February: 85km

Information: An interpreted data.

- Ex: Number of cholera cases during February month

Putting data into context creates added value to constitute information

**Data
Playbook
(Beta)**



International Federation of Red Cross and Red Crescent Societies
Fédération Internationale des Sociétés de la Croix-Rouge et du Croissant-Rouge
Federación Internacional de Sociedades de la Cruz Roja y de la Media Luna Roja
الاتحاد الدولي لجمعيات الصليب الأحمر والهلال الأحمر

DATA PLAYBOOK: SLIDEDeck 2

Source: https://www.preparecenter.org/sites/default/files/slidedeck2data_a21061818.pdf

'Good RDM makes data reusable'



Source: [10 aspects of highly effective research data - Good research data management makes data reusable](#) By Anita de Waard, Helena Cousijn, PhD, and Usbrand Jan Aalbersberg, PhD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Some questions at each stage of the life cycle...

- **Which datasets** will be generated in your research?
- How will your datasets be **named and referenced**?
- Which **file formats** will be used for each dataset?
- What data and **metadata standards** will each dataset follow?
- Who will **have access** to your datasets? How and when will you share your datasets, if applicable?
- How will you **archive and preserve** your datasets?
- How will you **license** your datasets?
- How will you deal with **privacy or confidentiality**, if applicable?

Adapted from: OSF Guides > Best Practices > Handling Data > Creating a data management plan (DMP). Available: <http://help.osf.io/m/bestpractices/1/618674-creating-a-data-management-plan-dmp>



The Support-Your-Data RDM rubric

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I have created plans for managing my data that are designed to streamline its future use by myself or others.
Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it not necessarily standardized.	I organize my data so that others can navigate, understand, and use it without me being present.
Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed to maximize opportunities for re-use by myself and others.
Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.
Analyzing your data and handling the outputs	I often have to redo my analyses or examine their products to determine what procedures or parameters were applied.	After I finish my analysis, I document the specific parameters, procedures, and protocols applied.	I regularly document the specifics of both my analysis workflow and decision making process while I am analyzing my data.	I have ensured that the specifics of my analysis workflow and decision making process can be understood and put into action by others.
Sharing and publishing your data	I share the results of my research, but generally I do not share the underlying data.	I share my data only when I'm required to do so or in response to direct requests from other researchers.	I regularly share the data that underlies my results and conclusions in a form that enables use by others.	Because of my excellent data management practices, I am able to efficiently share my data whenever I need to with whomever I need to.

Adapted from: Borghi J, Abrams S, Lowenberg D, Simms S, Chodacki J (2018) **Support Your Data: A Research Data Management Guide for Researchers**. Research Ideas and Outcomes 4: e26439. <https://doi.org/10.3897/rio.4.e26439>





The Research Data Lifecycle

many tools to assist





PLAN & DESIGN



What is a DMP?

A **data management plan** (DMP) is a living, written document explaining what you intend to do with your data during and following the conclusion of your research project. A DMP is required by many funders.

Even if it is not a requirement, having made such a plan can save you time and effort during your research, as it assists you with organising your data, preparing it for the next step in its lifecycle, and clarifying who will have access to it, how, and when.

If you plan on sharing your data, a DMP can help you troubleshoot the issues you should address to make sharing possible. Finally, a DMP helps ensure that your data remains useable to yourself, your collaborators, and other researchers in future.

Adapted from: OSF Guides > Best Practices > Handling Data > Creating a data management plan (DMP). Available: <http://help.osf.io/m/bestpractices/1/618674-creating-a-data-management-plan-dmp>



DMPonline

<https://dmp.lib.uct.ac.za/>

My plan (Gender, Health and Justice Research Unit)

Plan details | CHIRU DMP | Share | Export

This page gives you an overview of your plan. It tells what your plan is based on and gives an overview of the questions that you will be asked.

Plan name	My plan (Gender, Health and Justice Research Unit)
ID	-
Grant number	-
Principal Investigator/Researcher	Ya'qub Ebrahim
Plan data contact	-
Description	-

This plan is based on:

Institution | University of Cape Town (UCT-GeneriC)

Pick from a variety of templates (funder-specific or generic, i.e. 'UCT') to assist you with planning how you will collect, store, manage and analyse your research data during your research project.

Sections	Questions
1. Project name	- Insert the name of your project proposal.
2. Introduction/type of study	- Provide a summary of the written description of the proposed study. Include the study's objectives, design, and methods.
3. Description of existing data	- Provide if possible a survey of previously existing data relevant to the project; the nature and scale of such data; and a brief discussion of whether and how these data will be integrated or the gaps in these datasets the new study will fill.
4. Data collection and generation	- TYPES OF DATA/DATA OUTPUTS - Describe what types of data will be collected. Indicate whether the data will be qualitative or quantitative and the likely file formats in which the data will be collected. Indicate if there is an intention to convert file formats for long-term accessibility and preservation. - METHODOLOGIES FOR DATA CREATION/GENERATION - Describe the how data will be collected for this study. - QUALITY MANAGEMENT - Describe the quality control (QC) measures and quality assurance (QA) measure you will implement.
5. Data management, documentation and curation	- MANAGING, STORING AND CURATING DATA - Indicate how you will be storing and curating your electronic and paper/hard copy data. Focus on principles and systems with brief examples, and avoid long lists. - DATA DOCUMENTATION - Indicate what additional documentation (aside from the DMP) if any will accompany the dataset to support future users. - FILE NAMING CONVENTIONS - Indicate the naming convention for your data files. - DATA ARCHIVING - Outline your plans for storage/archiving of the final datasets. - ETHICS AND PRIVACY - Indicate how informed consent will be handled in your project.

DMPonline

<https://dmp.lib.uct.ac.za/>

My plan (Gender, Health and Justice Research Unit)

0/18 questions answered

0/100,000 available space used

Plan details: **GHJRU DMP** Share Export

1. Project name (1 question, 0 answered) +

2. Introduction/type of study (1 question, 0 answered) +

3. Description of existing data (1 question, 0 answered) +

4. Data collection and generation (3 questions, 0 answered) -

TYPES OF DATA/DATA OUTPUTS - Describe what types of data will be collected. Indicate whether the data will be qualitative or quantitative and the likely file formats in which the data will be collected. Indicate if there is an intention to convert file formats for long-term accessibility and preservation.

B I [List] [List] [Link] [Grid]

Useful information is provided at every step.

Save

Not answered yet

METHODOLOGIES FOR DATA CREATION/GENERATION - Describe the how data will be collected for this study.

B I [List] [List] [Link] [Grid]

Guidance Add comment

UCT Guidance

Data collected and stored by the GHJRU typically includes the following:

- In-depth interview audio files (mp3) and transcripts (MS word documents)
- Focus group discussion audio files (mp3) and transcripts (MS word documents, Nvivo files)
- Notes from in-depth interviews and focus group discussions, and other fieldnotes (MS word documents, Nvivo files)
- Quantitative survey data: both electronic (CSV, STATA, SPSS) and paper
- Minutes of research meetings—to be considered "data" only if collected as the result of a research process (Microsoft word documents)

Accessibility and preservation

Open and machine-readable formats help preserve data in the long term. Consider converting text files into RTF, PDF or XML format, quantitative data into CSV, and audio files into WAV to ensure they are accessible for future users and software systems.

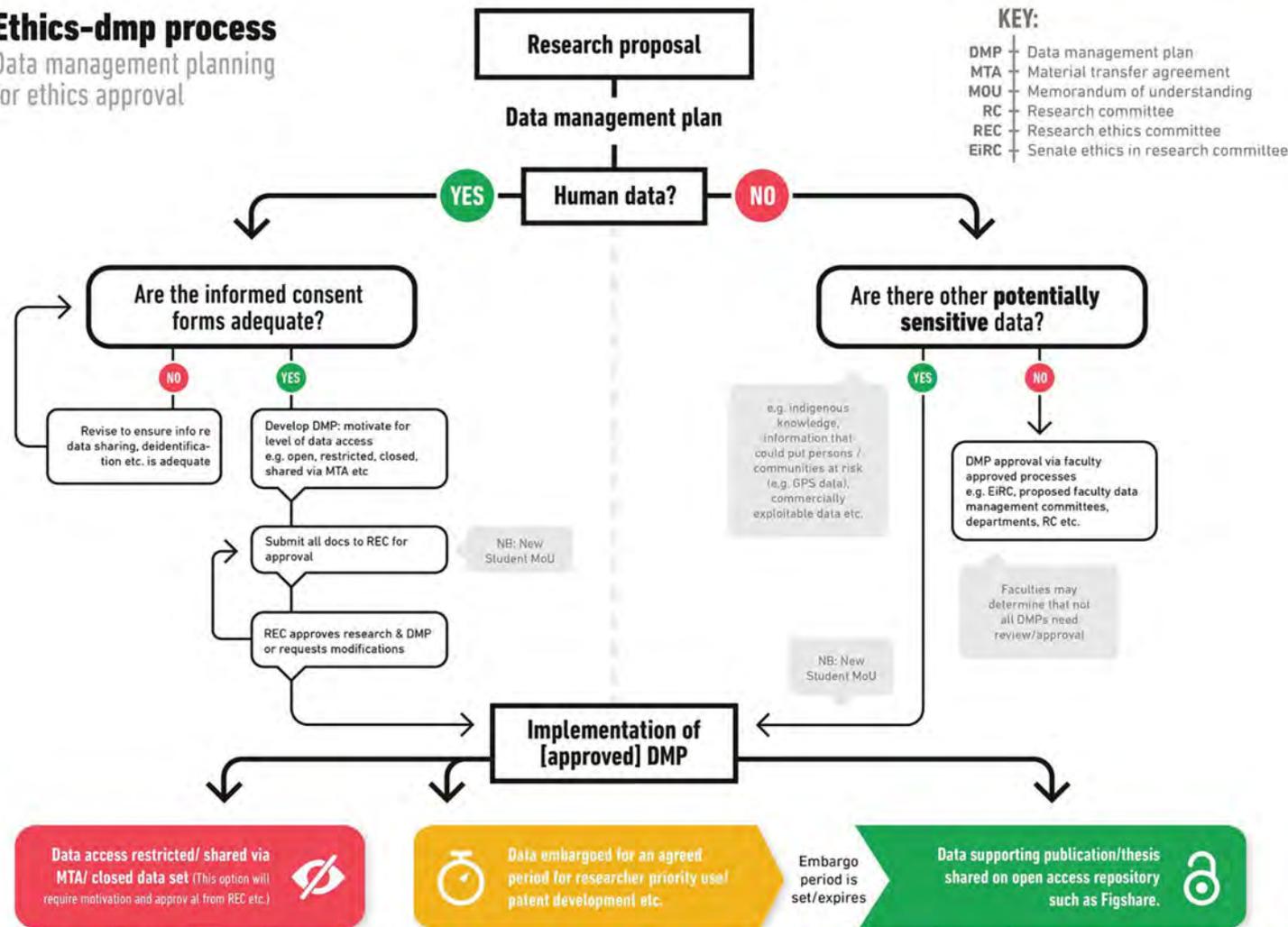
Guidance Add comment

UCT Guidance

New ethics clearance process, including DMPs

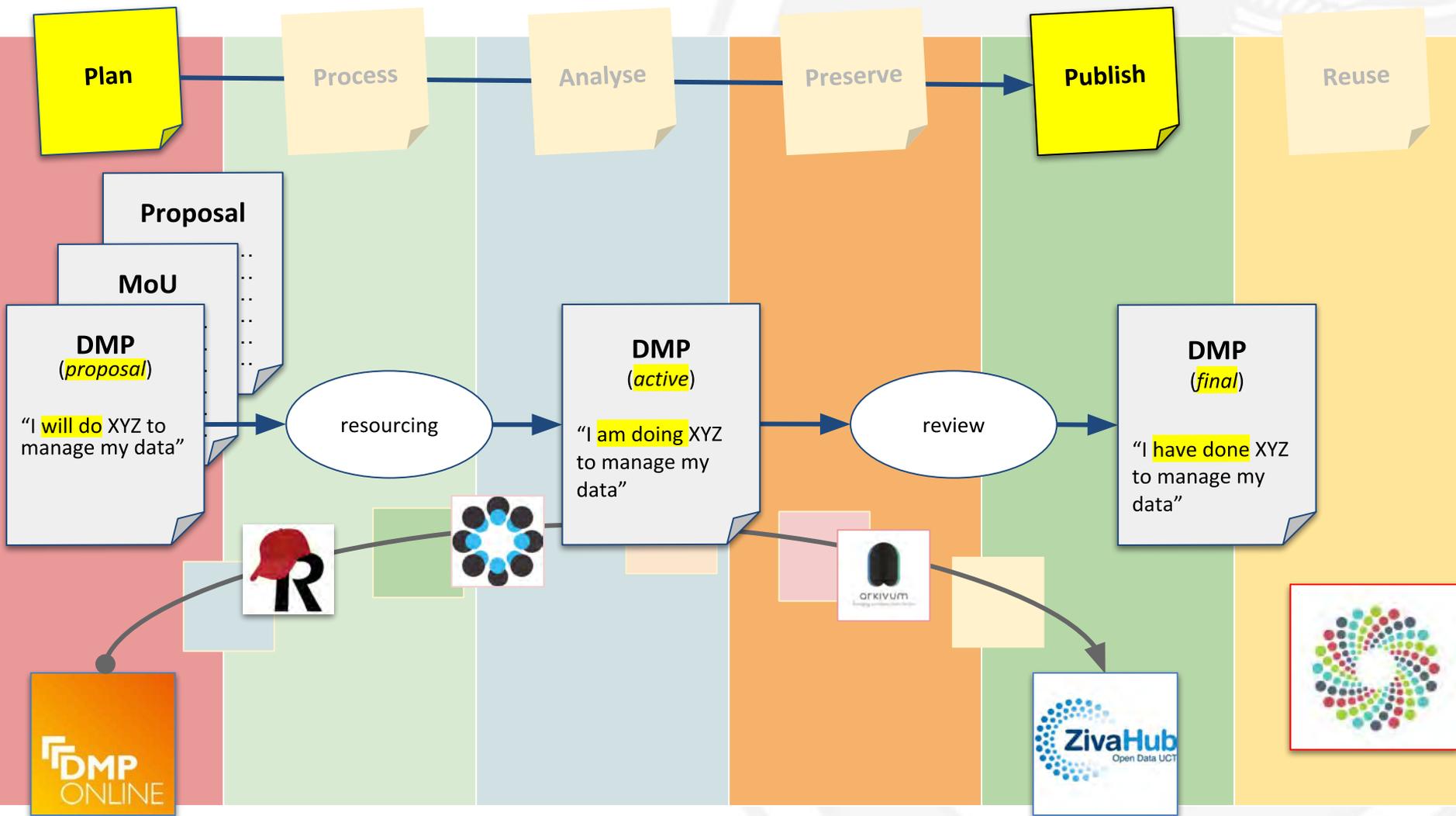
Ethics-dmp process

Data management planning for ethics approval



Source: https://commons.wikimedia.org/wiki/File:UCT_RDM_DMP-for-ethics-approval.png

The research journey (data focus)



Adapted from: Zimmer, Niklas; King, Thomas (2018): Research Data Management (RDM): What now for research administrators and PASS staff?. figshare. Presentation. <https://doi.org/10.25375/uct.7352495.v1>



COLLECT & CAPTURE



RedCap

<https://trn-redcap.uct.ac.za/>

The screenshot shows the REDCap interface for a project named 'DMPOnline Survey'. The user is logged in as '01401241'. The main navigation menu includes 'My Projects', 'Project Home', 'Project Setup', and 'Project Revision History'. The 'Project status' is 'Development'. The interface is divided into several sections:

- Main project settings:** Includes options to 'Use surveys in this project?' (checked), 'Use longitudinal data collection with defined events?' (unchecked), and a field to 'Modify project title, purpose, etc.'.
- Design your data collection instruments & enable your surveys:** Provides instructions on using the Online Designer or Data Dictionary, with links to download PDFs and a shared library.
- Enable optional modules and customizations:** Lists various modules like 'Repeatable instruments', 'Auto-numbering for records', 'Scheduling module', 'Randomization module', and 'Designate an email field'.
- Set up project bookmarks (optional):** Explains how to create custom bookmarks for navigation.
- User Rights and Permissions:** Describes how to grant access to other users or create Data Access Groups.

At the bottom left, there is a note: 'Adapted from: Harvard Catalyst <https://catalyst.harvard.edu/services/redcap/>'.

A secure web application for building and managing online surveys and databases, useful for collecting and tracking information and data from research studies, scheduling study events and conducting surveys.

Features:

- input data from anywhere in the world
- projects can be used by researchers from multiple sites and institutions
- total control of shaping your database or survey
- data may be imported from external data sources to begin a study or to provide mid-study data uploads
- export survey results to common data analysis packages
- generate a PDF version for printing in order to collect responses offline

Advice for the Collect & Capture Process

While collecting and capturing your data, make sure that you document it with correct, meaningful **metadata**:

- Describe the type of dataset generated
 - The form (*What kind of data does it hold?*)
 - The stability of each dataset (*How does the dataset change over time?*)
 - Create unique names for each of your datasets
- Document the data you are capturing and how you are identifying it within each data set by building a data dictionary
- Practice Good File Naming Conventions
- Document your process and store it alongside your data

Adapted from: OSF Guides > Best Practices > Handling Data > Creating a data management plan (DMP). Available: <http://help.osf.io/m/bestpractices/1/618674-creating-a-data-management-plan-dmp>



COLLABORATE & ANALYSE



Open Science Framework (OSF)

<https://osf.io/institutions/uct/>

Research Methods PRACTICAL in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring

Contributors: Tamas Nagy, Zoltan Kekecs
Date created: 2019-02-11 01:24 AM | Last Updated: 2019-04-30 02:46 PM
Category: Project

Wiki

Practical slides can be found here:
<https://drive.google.com/drives/folders/1brpFv87f0f1Uye6yad9jY5ajocFa7i?usp=sharing>

Files

- Name Modified
- Research Methods PRACTICAL in Clinical and Health Psychology - PS...
 - Dropbox: Readings and lecture slides to OSF
 - + Lecture slides to OSF
 - + Mini-exam questions and results
 - + readings
 - Google Drive: slides
 - Practical 1 - Managing research projects, introducing OSF.gslides 2019-03-18 12:41 PM
 - Practical 12 - Writing an abstract.gslides 2019-04-29 09:38 AM
 - Practical 2 - Creating online questionnaires.gslides 2019-02-18 04:06 AM
 - Practical 3 - Reading, writing, and citing research papers.gslides 2019-02-25 02:14 PM
 - Practical 4 - Ethical issues in conducting and publishing resear... 2019-03-18 03:48 AM
 - Practical 5 - Intervention studies and group design.gslides 2019-03-18 12:42 PM
 - Project evaluation rubric.gsheet 2019-04-28 09:18 PM
 - OSF Storage (Germany - Frankfurt)

Citation

Recent Activity

- Tamas Nagy updated file Research Methods in Clinical in Nagy Tamas.docx in OSF Storage in Research Methods PRACTICAL in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.
- Tamas Nagy updated file Research Methods in Clinical in Nagy Tamas.docx in OSF Storage in Research Methods PRACTICAL in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.
- Tamas Nagy updated wiki page Home to version 2 of Research Methods in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.
- Tamas Nagy updated wiki page Home to version 1 of Research Methods in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.
- Zoltan Kekecs linked Dropbox folder to Research Methods in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.
- Zoltan Kekecs authorized the Dropbox addon for Research Methods in Clinical and Health Psychology - PSYM17-CH-107 - 2019 Spring.

- a free, online platform that allows you to register your project, manage stakeholders, and centralise data that might be stored at different locations with different collaborators
- allows integrations with Google Drive, Dropbox, OneDrive, *figshare*, and many more
- provides unlimited, free storage
- helps with creating versions of your project at different stages ('forking')
- includes wiki-components for ease of documentation and description, including the development of a data dictionary

Show rows with cells including:

Variable	Variable name	Measurement unit	Allowed values	Description
Participant ID number	ID	Numeric	001-999	ID number assigned to participant in sequential order
Group number	GROUP	Numeric	1-30	Group assigned to participant based on ID number
Age in years	AGE	Numeric	18.0-65.0	Age of participant in years
Date of birth	DOB	mm/dd/yyyy	1-12/1-31/1951-1998	Participant's date of birth
Gender	SEX	Numeric	1 = male 2 = female	Participant's gender
Date of survey	SURVEY	mm/dd/yyyy	01/01/2015 - 01/01/2016	When the participant completed the survey
Self-reported consumer spending	SPEND	Numeric	0-100,000,000	Self-reported average yearly expenditure
Market sentiment	SENTIMENT	Numeric	1 = negative 2 = neutral 3 = positive	Sentiment towards US domestic economy
Actual GDP growth	GDP	Numeric	-5.0-5.0	Average US yearly GDP growth



Digital Scholarship

Data Anonymization:

Tools and processes for [data de-identification](#), to safeguard privacy of patients.

Data Analysis and Mining:

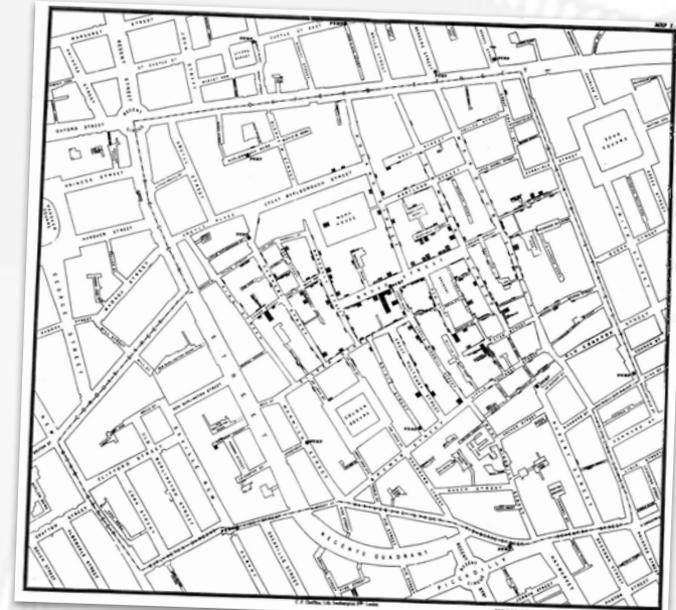
Tools that help you identify patterns in large volumes of data, combining statistics, AI and machine learning.

Data Visualization:

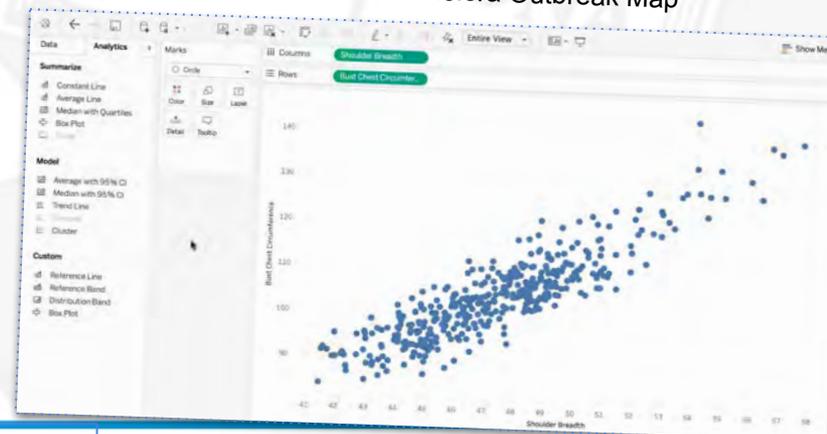
Tools that develop a graphical presentation of data and information through visual means.

Geographic Information Systems:

GIS tools that assist with public health, including tracking child immunizations, conducting health policy research, establishing service areas and districts, as well as epidemiological mapping. Geographical and other health data can be disseminated in a visually accessible manner that is readily understood.



Early Data Visualization
1854 Broad Street Cholera Outbreak Map





DISCOVER, REUSE & CITE



Generic view of the scientific workflow



Source: The Scholarly Kitchen: Lettie Y. Conrad: Mapping Open Science Tools. <https://scholarlykitchen.ssonnet.org/2018/08/30/mapping-open-science-tools/?informz=1>

Data catalogues, registries and repositories



Data catalogues, registries and repositories

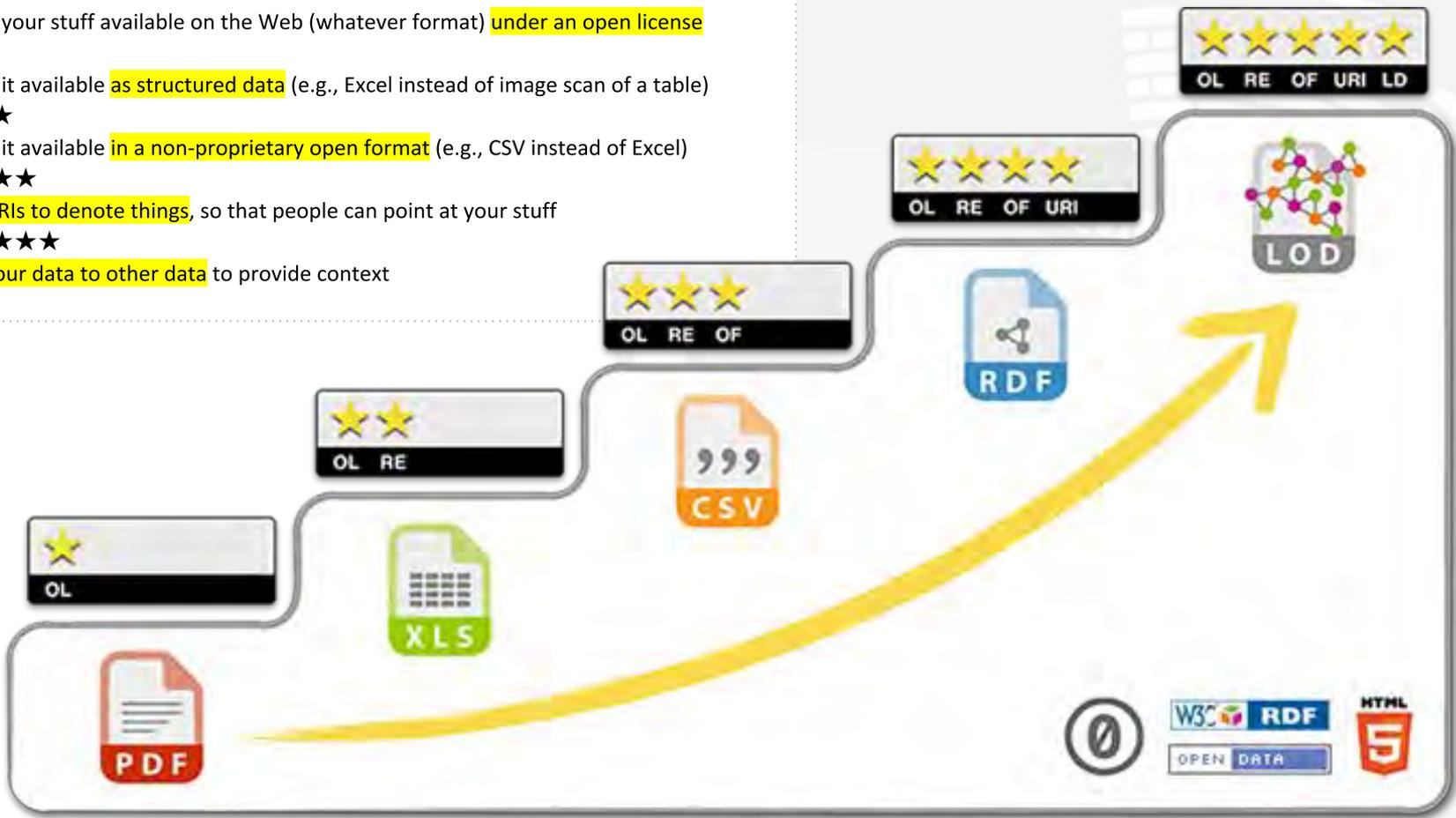
directly UCT - relevant

- [BioLINCC](#) – Clinical specimen database.
- [Code Ocean](#) – Cloud-based computational platform which provides a way to share, discover and run published code.
- [ContentMine](#) – Uses machines to liberate 100,000,000 facts from the scientific literature.
- [DataBank](#) – Analysis and visualisation tool that contains collections of time series data on a variety of topics.
- ✗ [DataCite](#) – Establish easier access to research data by providing persistent identifiers for data.
- [Datahub](#) – Publish or register datasets, create and manage groups and communities
- [Dataverse Network](#) – Harvard-based tool to share, cite, reuse and archive research data.
- [Delvehealth](#) – A data collection of global clinical trials, clinical trial investigator profiles, publications and drug development pipelines.
- [Deveo](#) – Free, private Git, Mercurial, and SVN repository management platform.
- [Dryad](#) – Data repository for any files associated with any published article in the sciences or medicine.
- ✗ [Figshare](#) – Manage your research in the cloud and control who you share it with or make it publicly available and citable.
- [GenBank](#) – Gene sequence database provided by the National Center for Biotechnology Information.
- [GitHub](#) – Online software project hosting using the Git revision control system.
- [How Can I Share It](#) – Find information and tools to ensure your articles can be shared with your colleagues quickly and easily.
- ✗ [Open Science Framework](#) – Gathers a network of research documents, a version control system, and a collaboration software.
- [Quip](#) – Combines chat, documents, spreadsheets, checklist, and more to collaborate on any device.
- ✗ [re3data](#) – Global registry of research data repositories.
- [Research Compendia](#) – Tools for researchers to connect their data, code and computational methods to their published research
- [SlideShare](#) – Community for sharing presentations and other professional content
- ✗ [Zenodo](#) – A home for the long-tail of science, enabling researchers to share and preserve any research outputs.

Source: <http://connectedresearchers.com/online-tools-for-researchers/>

5 ★ Open Data

- ★ make your stuff available on the Web (whatever format) **under an open license**
- ★★ make it available **as structured data** (e.g., Excel instead of image scan of a table)
- ★★★ make it available **in a non-proprietary open format** (e.g., CSV instead of Excel)
- ★★★★ **use URIs to denote things**, so that people can point at your stuff
- ★★★★★ **link your data to other data** to provide context

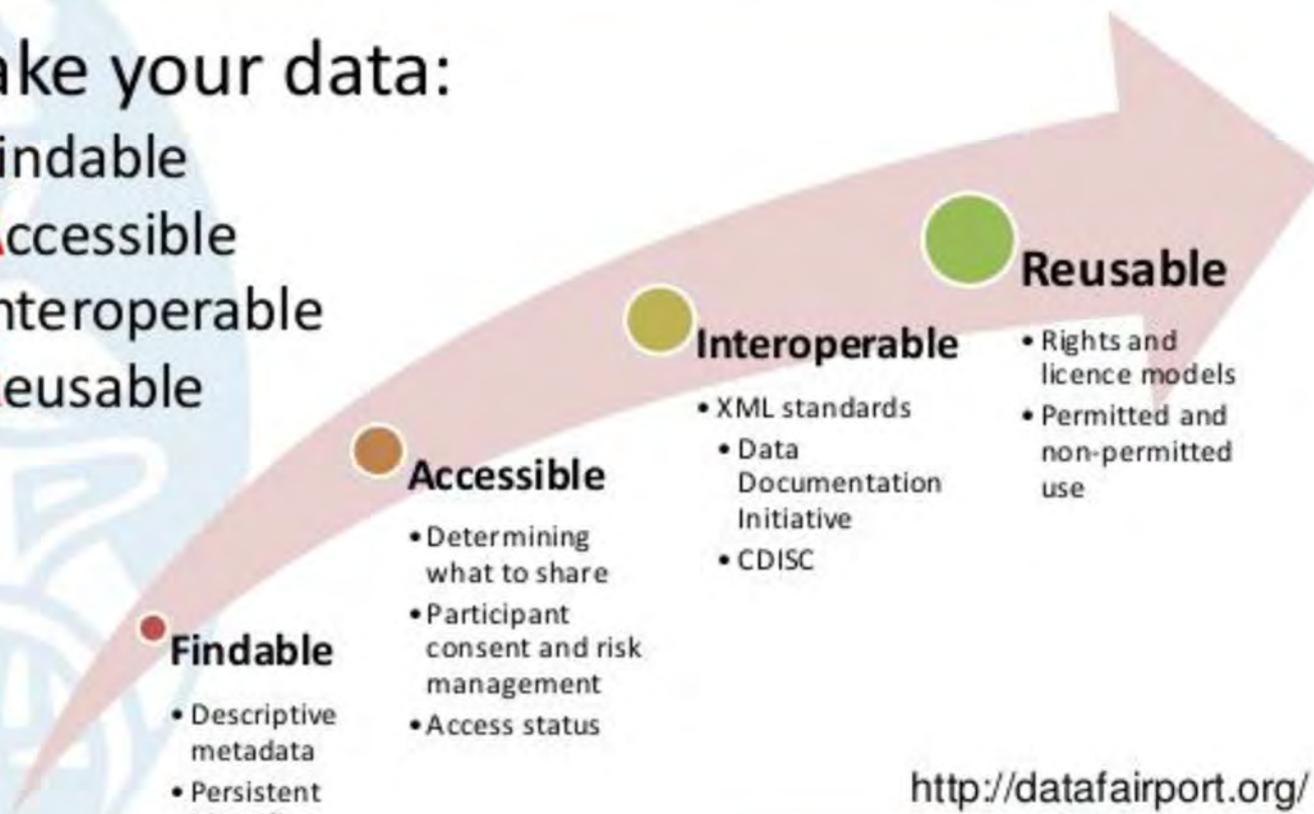


Source: <https://5stardata.info/en/>

The FAIR guiding principles

Make your data:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable



Findable

- Descriptive metadata
- Persistent Identifiers

Accessible

- Determining what to share
- Participant consent and risk management
- Access status

Interoperable

- XML standards
- Data Documentation Initiative
- CDISC

Reusable

- Rights and licence models
- Permitted and non-permitted use

<http://datafairport.org/>



Source: <https://www.slideshare.net/lshmt/preparing-data-for-sharing-the-fair-principles>

Some things **you can do** to make your data **FAIR**

- Describe your data in a data repository
- Receive a persistent identifiers (e.g. uct doi provided by ZivaHub)

Findable

- Consider what can be published
- Obtain participant consent
- Perform de-identification / anonymisation

Accessible

- Use open formats
- Apply consistent vocabulary
- Use common/disciplinary metadata standards

Interoperable

- Consider permitted use
- Apply machine-readable open licenses (e.g. CC-BY etc.)

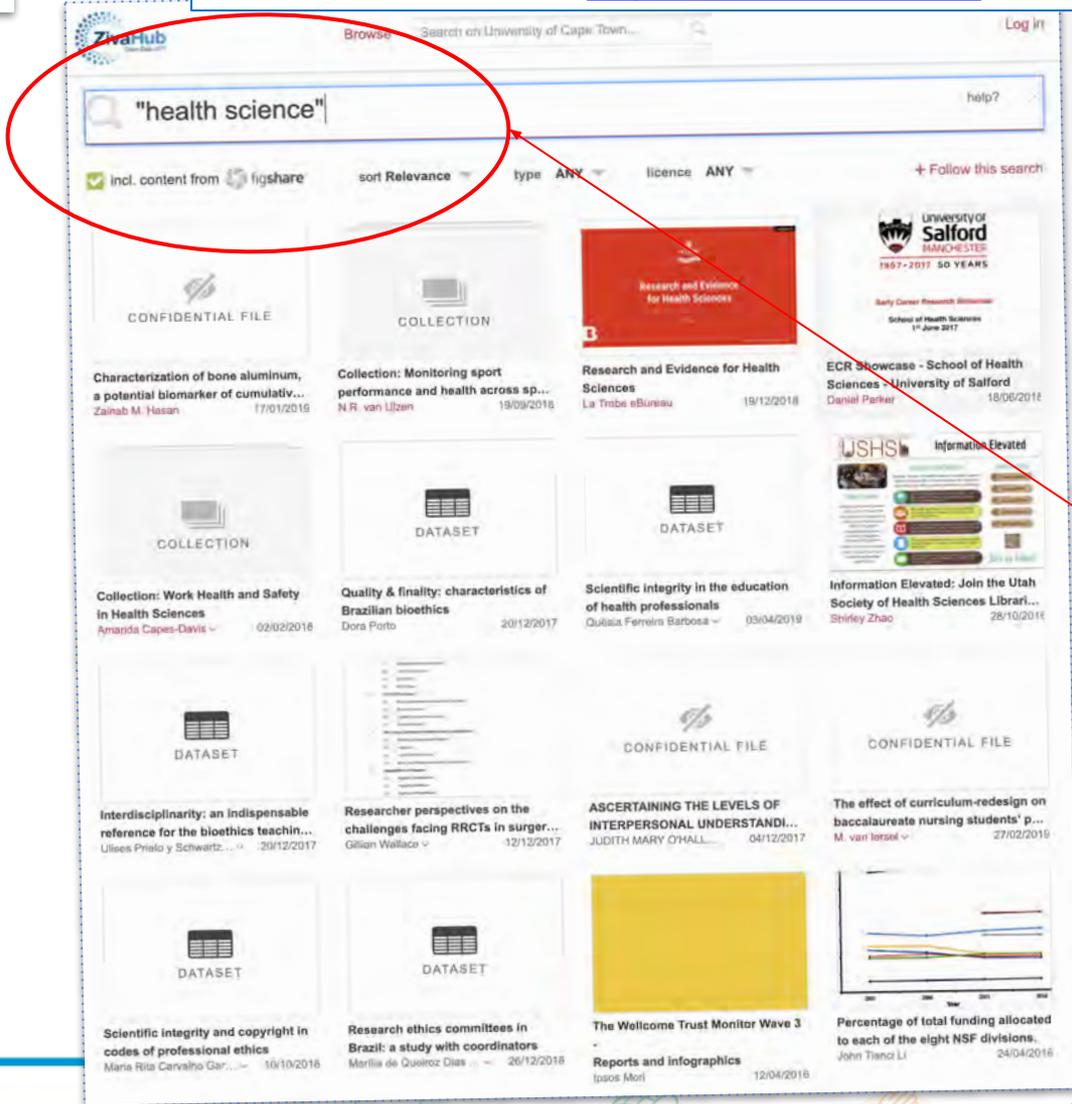
Reusable



SHARE & PUBLISH

ZivaHub | Open Data UCT

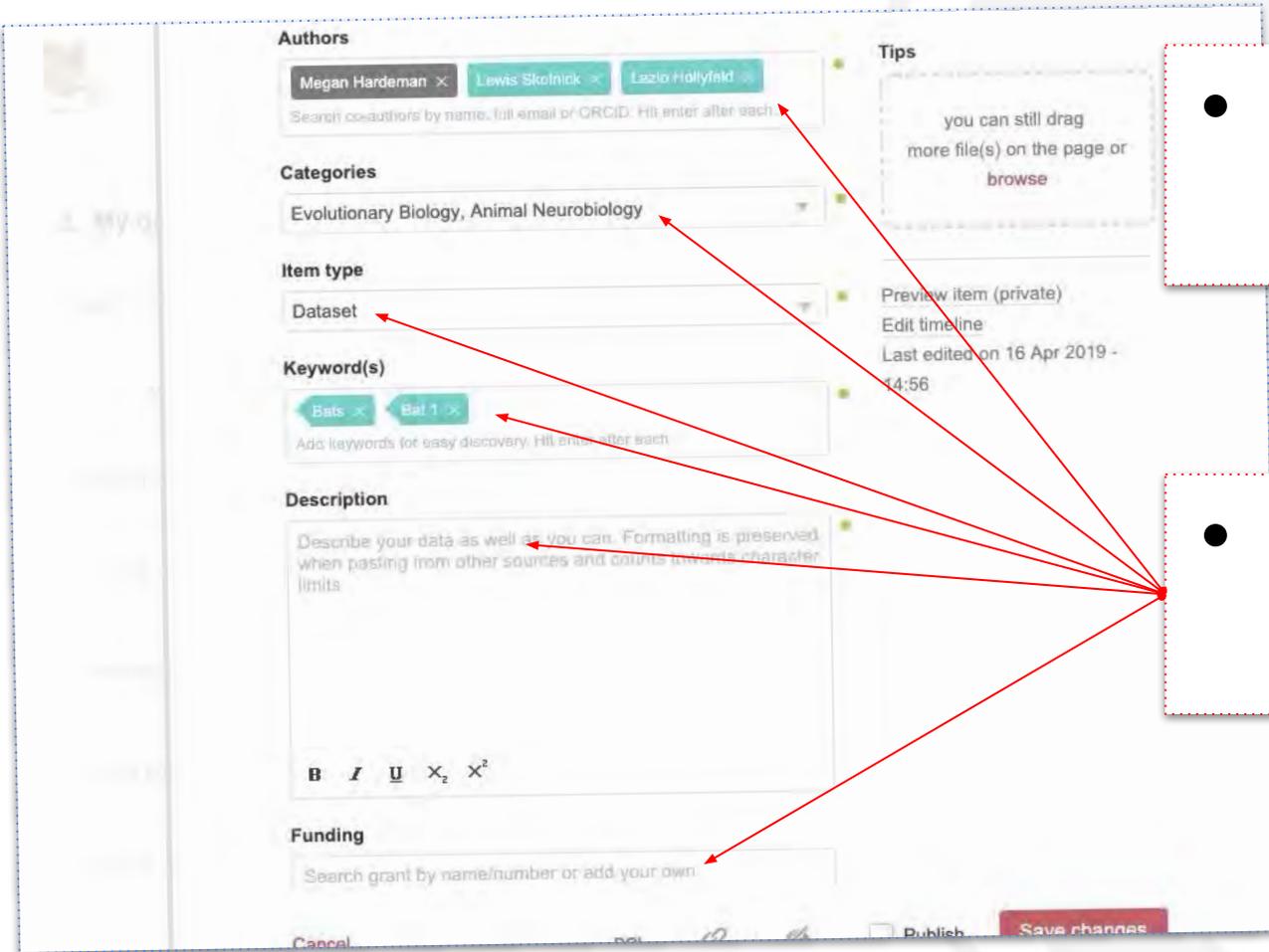
<https://zivahub.uct.ac.za/>



- a repository to store and openly disseminate data
- powered by *Figshare* for institutions
- keeps track of views, downloads and citations
- integrates with [ORCID](#)
- allows search across all Figshare platforms

ZivaHub | Open Data UCT

<https://zivahub.uct.ac.za/>



The screenshot shows a metadata form with several sections: **Authors** (Megan Hardeman, Lewis Skolnick, Lizzie Hollyfeld), **Categories** (Evolutionary Biology, Animal Neurobiology), **Item type** (Dataset), **Keyword(s)** (Bats, Bat 1), **Description** (text area with rich text editor), and **Funding** (search for grant). A 'Tips' box indicates that more files can be added. At the bottom, there are 'Cancel', 'Publish', and 'Save changes' buttons.

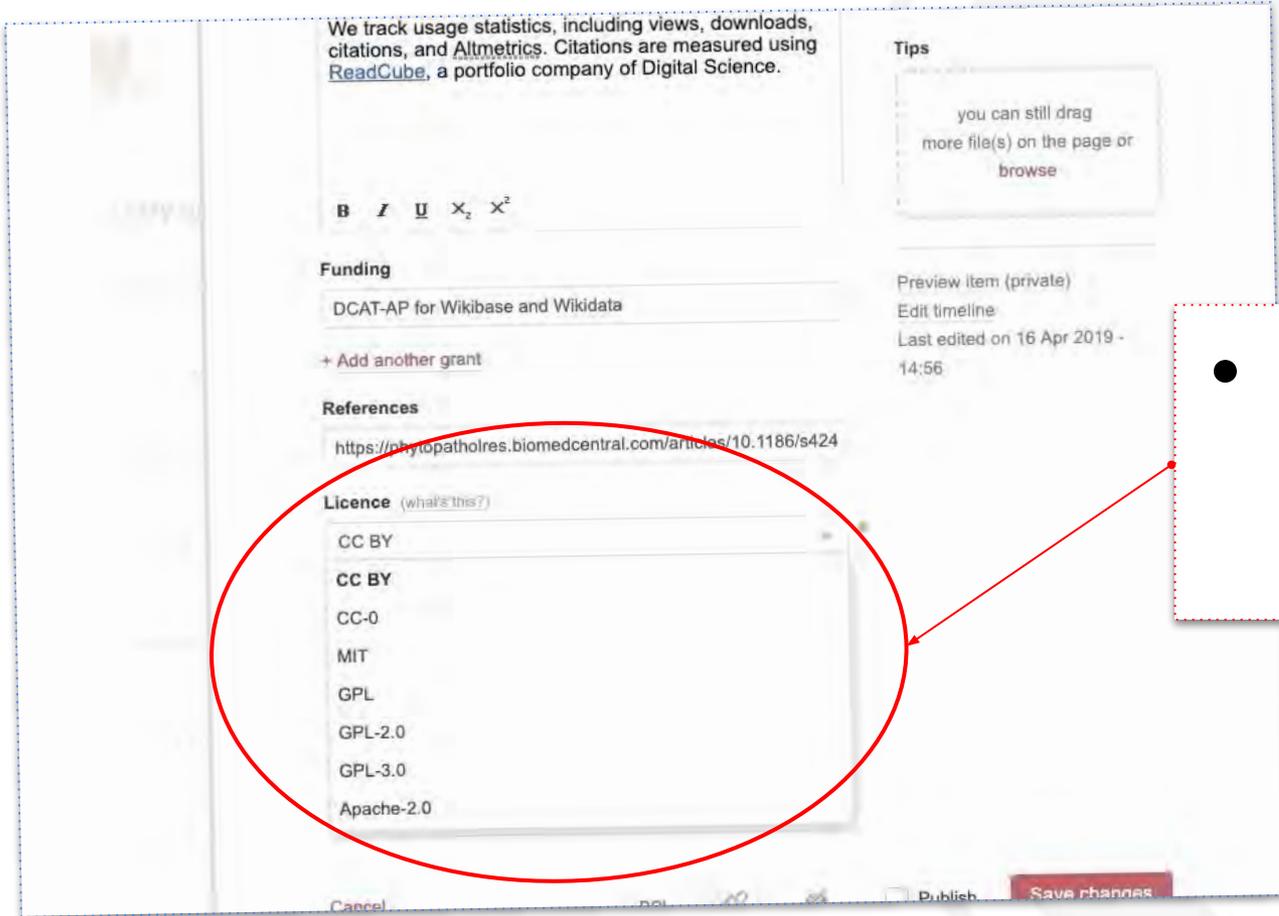
- supports the upload of *any* file format, and aims to visualise all of them

- embeds relevant metadata to make data FAIR

Source: Figshare. End User Guide GIFs. Available: <https://figshare.com/s/b3600c85f576d88d067b>

ZivaHub | Open Data UCT

<https://zivahub.uct.ac.za/>



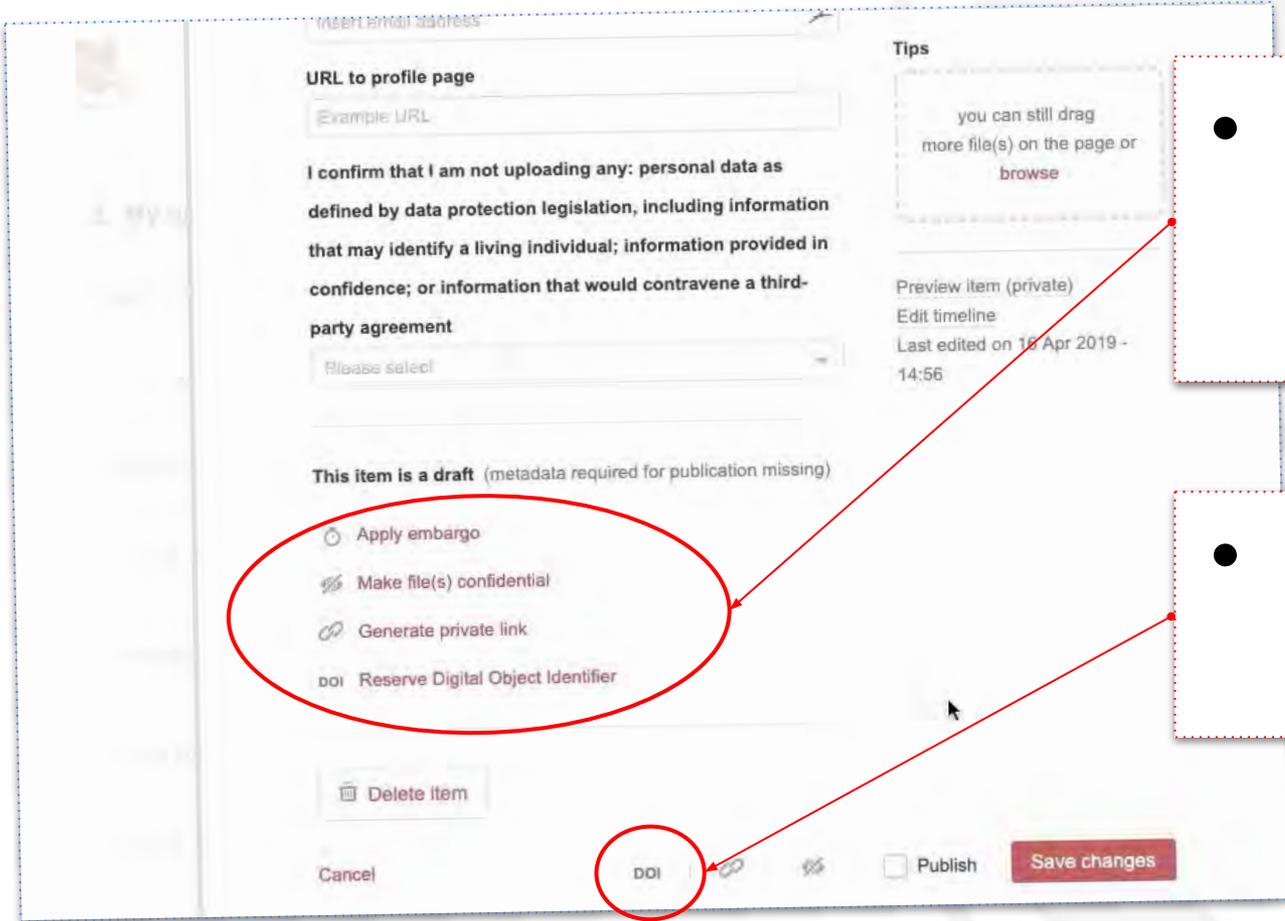
The screenshot shows the ZivaHub data upload interface. It includes a text area for a description, a 'Funding' field with 'DCAT-AP for Wikibase and Wikidata', a 'References' field with a URL, and a 'Licence' dropdown menu. The 'Licence' menu is highlighted with a red circle and contains the following options: CC BY, CC BY, CC-0, MIT, GPL, GPL-2.0, GPL-3.0, and Apache-2.0. A 'Tips' box on the right says 'you can still drag more file(s) on the page or browse'. At the bottom, there are 'Cancel', 'Publish', and 'Save changes' buttons.

- choose from a range of licensing options when publishing your data openly

Source: Figshare. End User Guide GIFs. Available: <https://figshare.com/s/b3600c85f576d88d067b>

ZivaHub | Open Data UCT

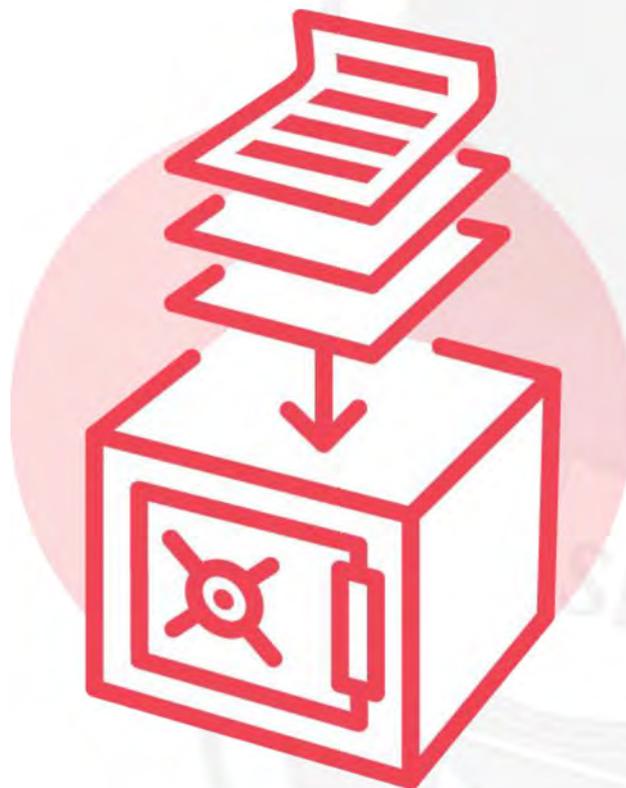
<https://zivahub.uct.ac.za/>



- share data privately with your funders, reviewers, or supervisors before publishing it later

- all data items get assigned a Digital Object Identifier (DOI)

Source: Figshare. End User Guide GIFs. Available: <https://figshare.com/s/b3600c85f576d88d067b>



MANAGE, STORE, PRESERVE

Manage, Store, Preserve

Unfortunately, digital files aren't guaranteed to be accessible forever. Certain file types, particularly rare types or those produced under proprietary formats, are more likely to become unreadable over time as technology advances and new software is developed. Where possible, it is advisable that you produce or export files in **open formats** as these tend to be the most accessible over time.

Likewise, storage locations change, and links, even so-called *persistent identifiers* tend to 'rot' due to this.

A future problem: where is my data?

I know where it is but...

It's in an unsupported file format

It's in a legacy system

It's not well described so it's irretrievable

It's corrupted

I don't even know where it is...

It was on destroyed hardware

A third party has it

It's on a hard drive in a vault

I expected it to be just where I left it

Adapted from: Arkivum: **Webinar Recording - Making the case for digital preservation.** Available:

<http://sites.arkivum.com/webinar-recording-making-the-case-for-digital-preservation-how-to-engage-your-internal-stakeholders-20-sept?hsCtaTracking=afd562aa-7fef-4f16-a1de-0958a8d68dce%7C277de3d6-6467-4c10-a387-8931548403fe>



Closing remarks & questions

Moving Towards **Open Science**

Open Science is a *practice* which with the objective of making data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. It aims to help make research more efficient, accountable, collaborative, and of good quality.

The **end** of one research project is the **beginning** of another one.

How can we assist you towards *practicing* **Open Science**?

How to Reach Us

RESEARCH DATA MANAGEMENT TRAINING



ALL SESSIONS @ 10AM IN ULWAZI TRAINING ROOM

Discover how you can become a more **EFFICIENT** researcher in today's digital world. Start managing your **DATA** and your **RESEARCH** process with guidance from the **DLS TEAM**.

RESEARCH DATA MANAGEMENT WITH DMPONLINE



The new Student MoU as well as the NRF require students to outline their data plans for their research projects in a Data Management Plan (DMP). This talk/workshop takes you through the reasons for creating a DMP, as well as guiding you through using the DMPonline website.



WEDNESDAY
12 JUN | 14 AUG

DOING DIGITAL SCHOLARSHIP



COLLABORATE & ANALYSE



COLLECT & CAPTURE

Doing research requires interacting with a multitude of digital spaces. This talk outlines digital processes and tools that can increase efficiencies throughout a research project. It looks at collaborative tools for managing, analyzing, mapping and visualizing research data.

WEDNESDAY
15 MAY | 11 SEP | 13 NOV | 11 DEC

SHARING AND PUBLISHING WITH ZIVAHUB



DISCOVER, REUSE & CITE



SHARE & PUBLISH

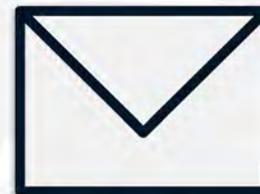


MANAGE, STORE, PRESERVE



UCT's open data repository is rapidly growing. Uploading your research outputs to *ZivaHub* makes them discoverable, citable, shareable and reusable. Learn about open data and *ZivaHub* which allow you to engage with researchers at UCT and the world.

WEDNESDAY
10 JUL | 9 OCT



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH

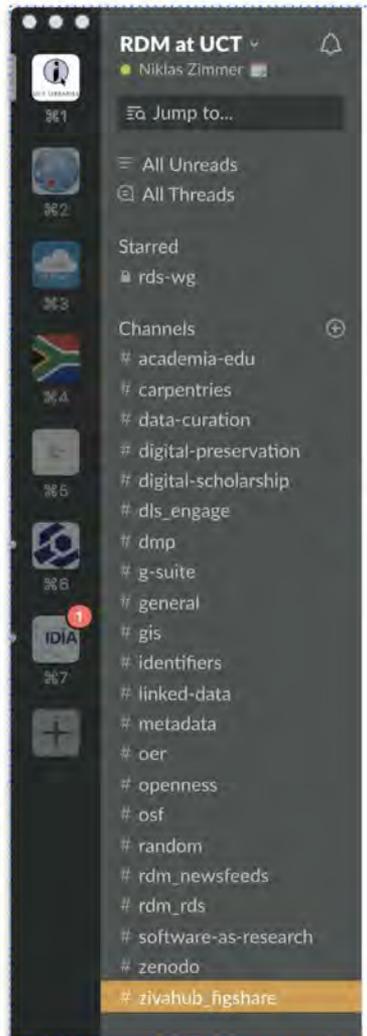


MANAGE, STORE, PRESERVE



'RDM at UCT' Slack workspace

Slack: *'Searchable Log of All Conversation and Knowledge'*



RDM at UCT

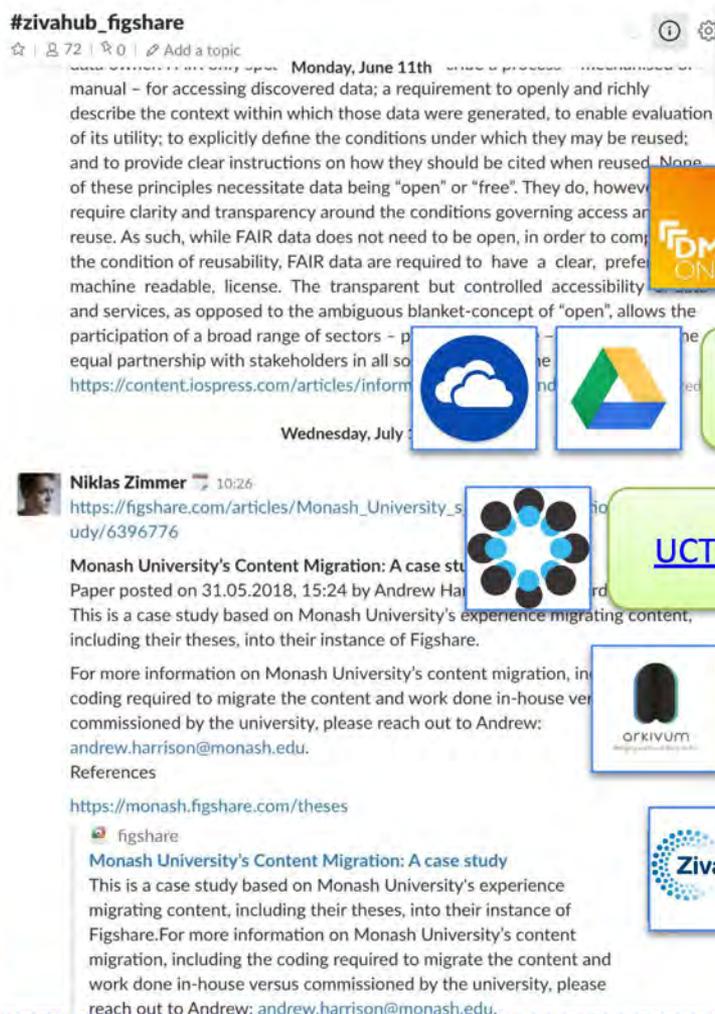
Niklas Zimmer

Jump to...

- All Unreads
- All Threads
- Starred
- rds-wg

Channels

- # academia-edu
- # carpentries
- # data-curation
- # digital-preservation
- # digital-scholarship
- # dls_engage
- # dmp
- # g-suite
- # general
- # gis
- # identifiers
- # linked-data
- # metadata
- # oer
- # openness
- # osf
- # random
- # rdm_newsfeeds
- # rdm_rds
- # software-as-research
- # zenodo
- # zivahub_figshare



#zivahub_figshare

Monday, June 11th

manual – for accessing discovered data; a requirement to openly and richly describe the context within which those data were generated, to enable evaluation of its utility; to explicitly define the conditions under which they may be reused; and to provide clear instructions on how they should be cited when reused. None of these principles necessitate data being “open” or “free”. They do, however, require clarity and transparency around the conditions governing access and reuse. As such, while FAIR data does not need to be open, in order to comply with the condition of reusability, FAIR data are required to have a clear, preferably machine readable, license. The transparent but controlled accessibility and services, as opposed to the ambiguous blanket-concept of “open”, allows the participation of a broad range of sectors – in an equal partnership with stakeholders in all sectors.

Wednesday, July

Niklas Zimmer 10:26

https://figshare.com/articles/Monash_University_study/6396776

Monash University's Content Migration: A case study

Paper posted on 31.05.2018, 15:24 by Andrew Harrison

This is a case study based on Monash University's experience migrating content, including their theses, into their instance of Figshare.

For more information on Monash University's content migration, including the coding required to migrate the content and work done in-house versus commissioned by the university, please reach out to Andrew: andrew.harrison@monash.edu.

References

<https://monash.figshare.com/theses>

figshare

Monash University's Content Migration: A case study

This is a case study based on Monash University's experience migrating content, including their theses, into their instance of Figshare. For more information on Monash University's content migration, including the coding required to migrate the content and work done in-house versus commissioned by the university, please reach out to Andrew: andrew.harrison@monash.edu.



[RDM at UCT \(Slack\)](#)



[UCT DMPonline](#)



[OneDrive / Google Drive etc.](#)



[UCT Open Science Framework \(OSF\)](#)



[Digital preservation](#)



[ZivaHub | Open Data UCT](#)



Declaration On Research Assessment

Improving how research is assessed

sfdora.org

@DORAssessment

now also
including
UCT!

Signed by >500 organizations and >12,500 individuals

Supporting organizations



Thank you!



dls@uct.ac.za



[@DigitalUct](https://twitter.com/DigitalUct)



rdm-at-uct.slack.com



DIGITAL LIBRARY
SERVICES

<http://www.digitalservices.lib.uct.ac.za/>