Let's change what we value in research.

Sign DORA

DIGITAL LIBRARY SERVICES

Winning at Data: Digital Scholarship - RDM - Open Science

eResearch Emerging Scholars

Thursday 31st October 2019

UCT Libraries

Digital Library Services

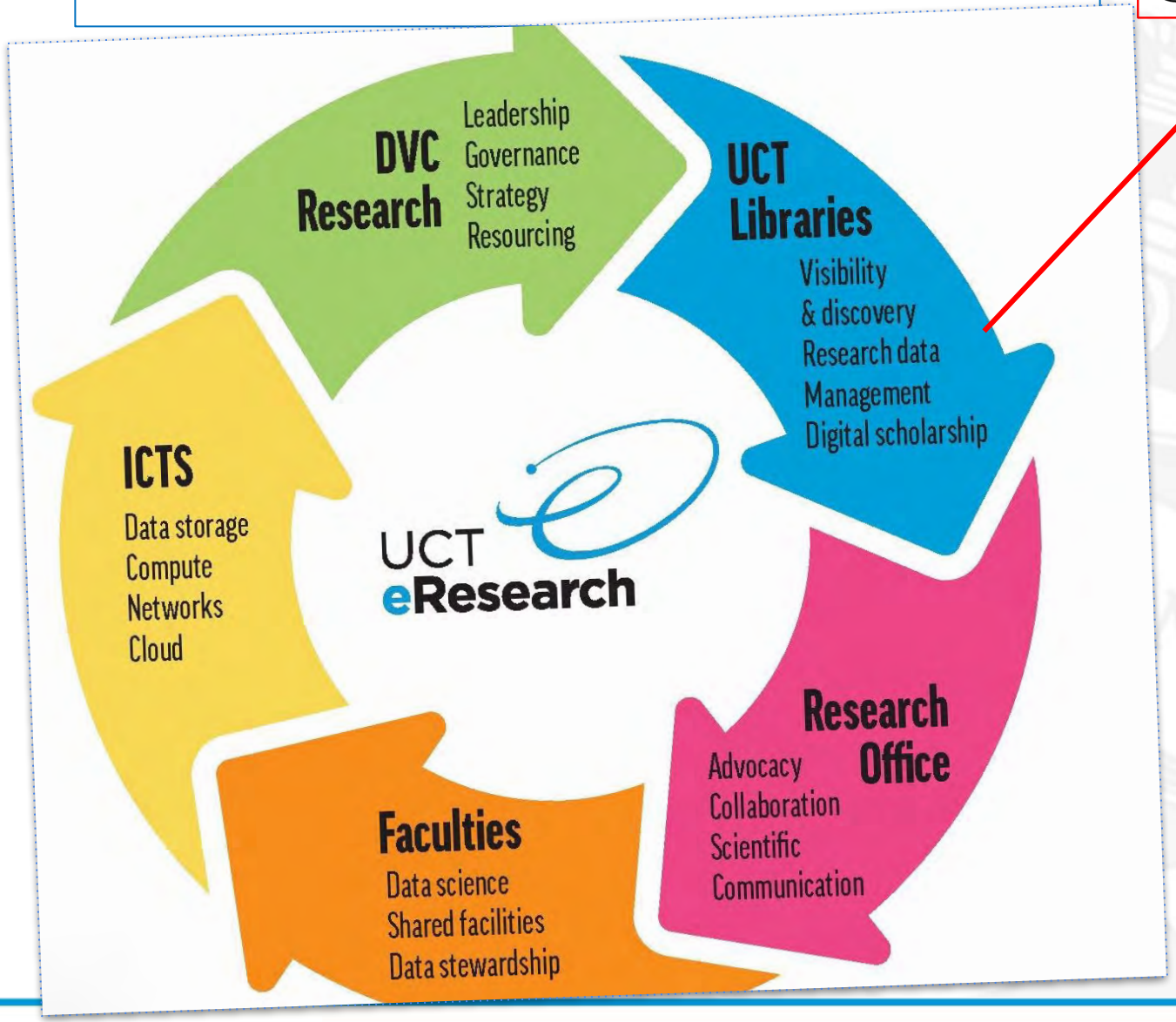
- Sanjin Muftić
  (digital scholarship specialist)

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Much of this presentation is adapted from / part of:
Muftić, S., Zimmer, N., Chikuni, P., Ebrahim, Y & Slingsby, T. (2019): **Doing Digital Scholarship with DLS: The Motivations**. Presentation. 10.25375/uct.8118014

https://osf.io/cz2sd/

Introduction



DIGITAL LIBRARY SERVICES

Digital Scholarship and DLS | **mission & vision**

We provide Digital Scholarship services to the University of Cape Town, including the following:

- Data Curation activities supporting best practices in **Research Data Management (RDM)**;
- specialist **Digitisation** services towards **Digital Preservation**;
- expertise in **Geographic Information Systems (GIS)**.

We advocate for **Open Science**, to make research done at UCT more *efficient, collaborative, accessible, findable and reusable*. We spearhead these practices as contributions to a more equitable and sustainable social order in the higher education landscape.

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

DIGITAL LIBRARY SERVICES

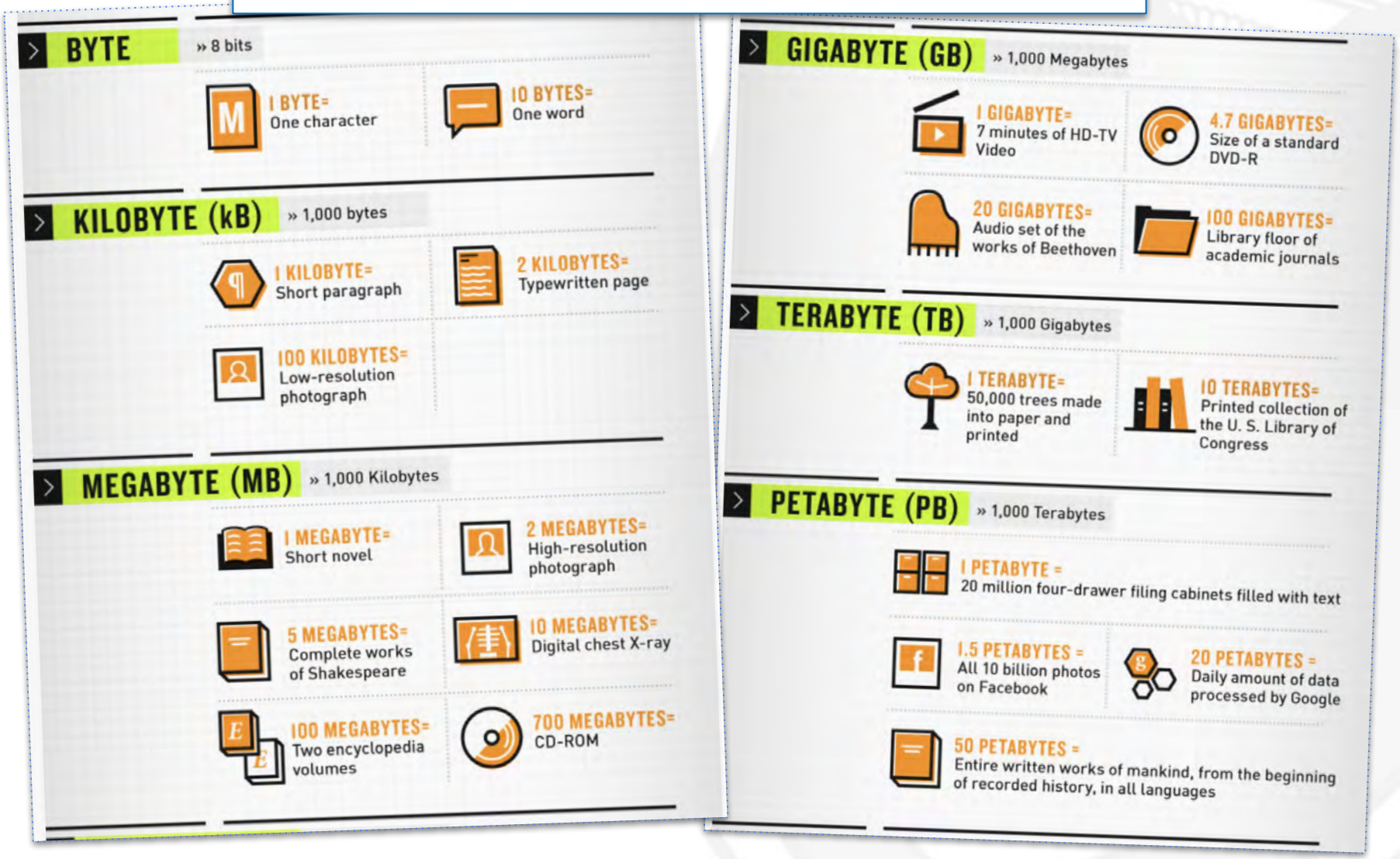
Your Data, Your Challenges

"**What** (are my) **research data?**"

… any information collected, stored, and processed to produce and validate original research results.

Compiled from: LibGuides@ Macalester University. Available at: https://libguides.macalester.edu/c.php?g=527786&p=3608583

Data sizes



Source Data Science Berkely: https://datascience.berkeley.edu/big-data-infographic/

"What are your data and what is tough about it?"

DIGITAL LIBRARY SERVICES

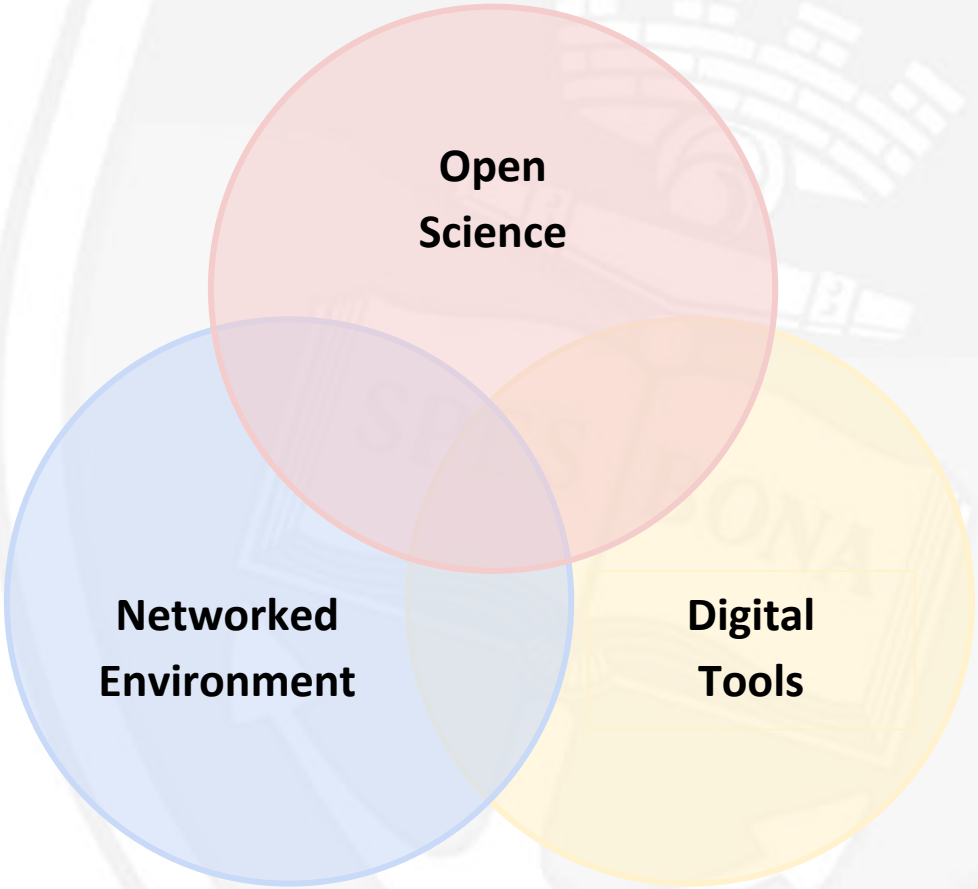
Digital Scholarship, Open Science and Research Data Management
a brief overview

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Thursday 31 October 2019

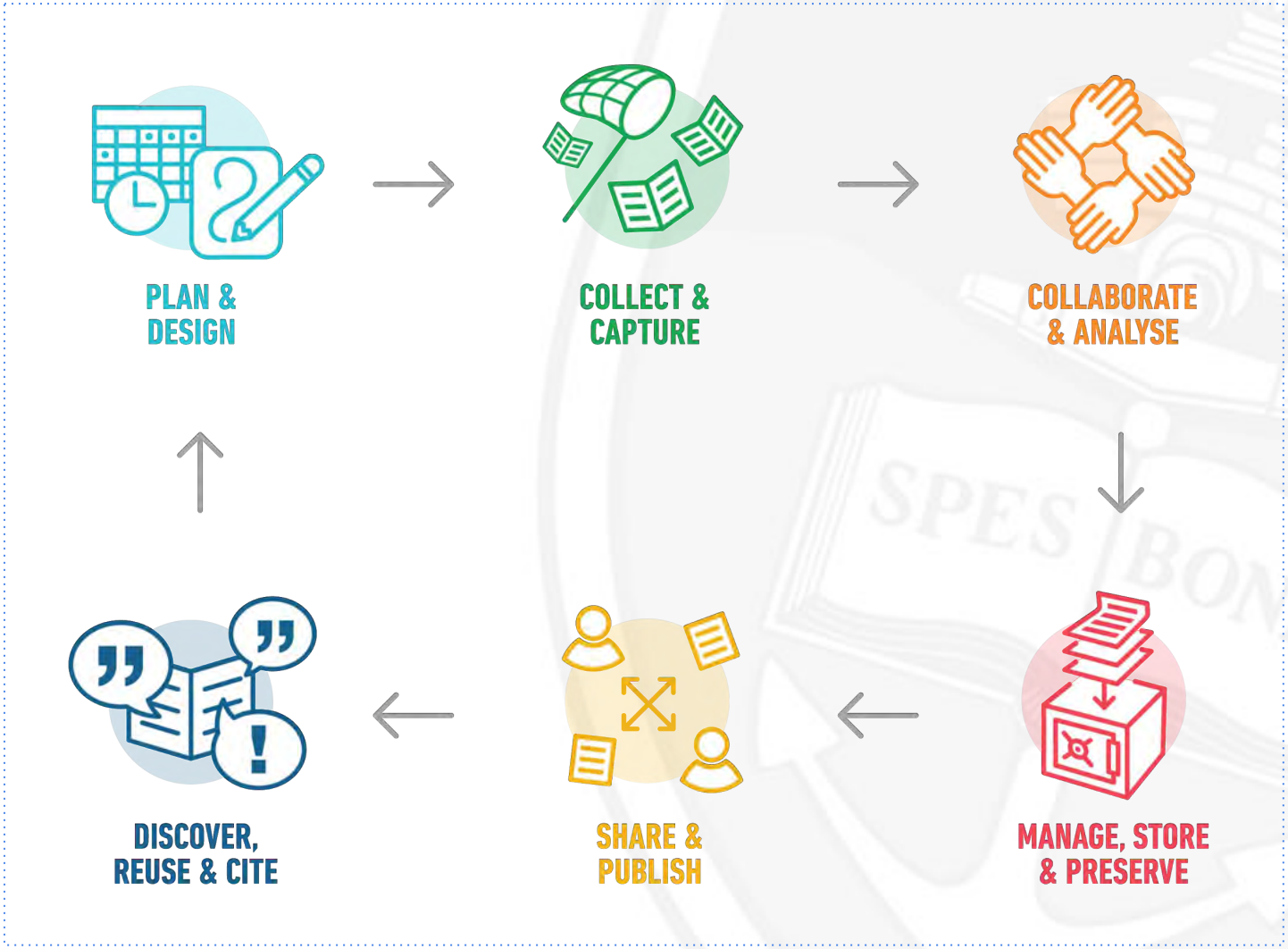
What is **Digital Scholarship**?

Digital Scholarship *is the application and integration of digital tools and methods to discovery, research, teaching and learning.*



Adapted from: Weller, M. 2011. *The Digital Scholar*: https://www.open.edu/openlearn/ocw/mod/oucontent/view.php?id=48677&section=2

The research project (& its data) lifecycle



Take you Back - What is Science?

- Doesn’t belong to a single discipline
- Ongoing Activity:
  - Investigating our world
  - Asking questions
- To better understand all that is around us and within us
- Science
  - Scholarship of asking questions and
    the knowledge that is obtained in the whole process

How do we practice science in our environment?

- Knowledge
  - The ones who fund can't access it
  - Restrictions on how it can be used and reproduced
- Process and Data
  - Poorly formatted if it all accessible
  - Hidden in propriariety software
- Researchers
  - Culture of fear of being scooped
  - Pressure of publishing or perishing
- Indicators of "Quality" Research
  - Pre -technological
  - Judged on how and where it is disseminated - high impact journals
- Business of Publishing

CLOSED SCIENCE?

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%2Fnchem.1149

How does science/knowledge circulate now?

- **For whom do we do science?**
  - Is this publishing model in the best interest of all humans?
- **Are we using research effectively?**
  - Are we building on the work of others?
    - *Are we learning from the "failures" of others?*
  - Are we allowing others to build on our work?
    - *Are we allowing others to learn from our "failures"?*
- **Do we subscribe to the primary human right to have access to knowledge?**
  - Should knowledge be locked behind paywalls?
  - *Should the process of asking questions be exclusive?*

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%2Fnchem.1149

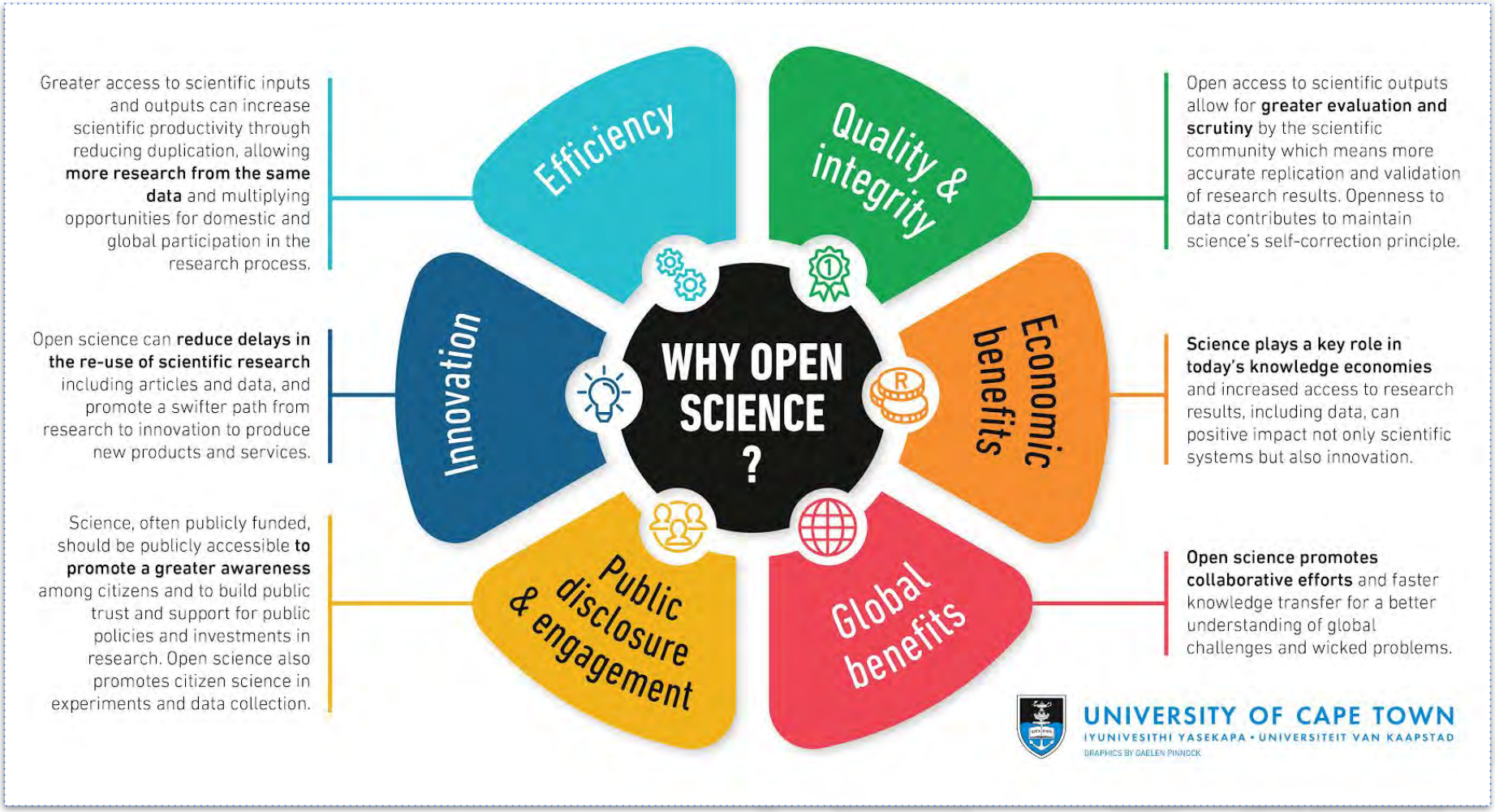
Open Science is ...

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 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 and subscribe to grounded ethical practices.

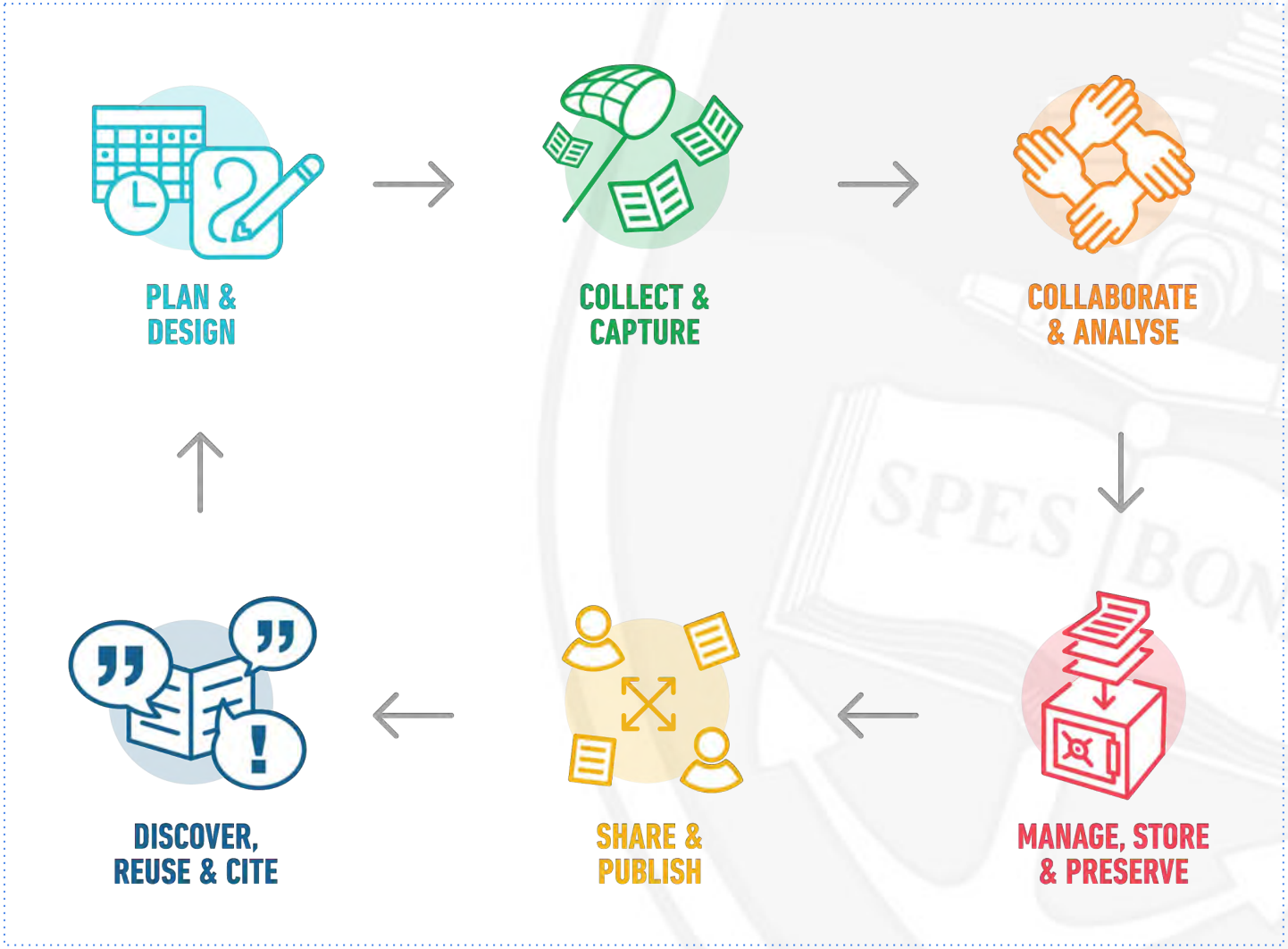
Source: Foster Open Science: (https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition
Adapted from: Woelfle, M.; Olliaro, P.; Todd, M. H. (2011). "Open science is a research accelerator". Nature Chemistry. 3 (10): 745–748. https://doi.org/10.1038%2Fnchem.1149

Open Science at UCT

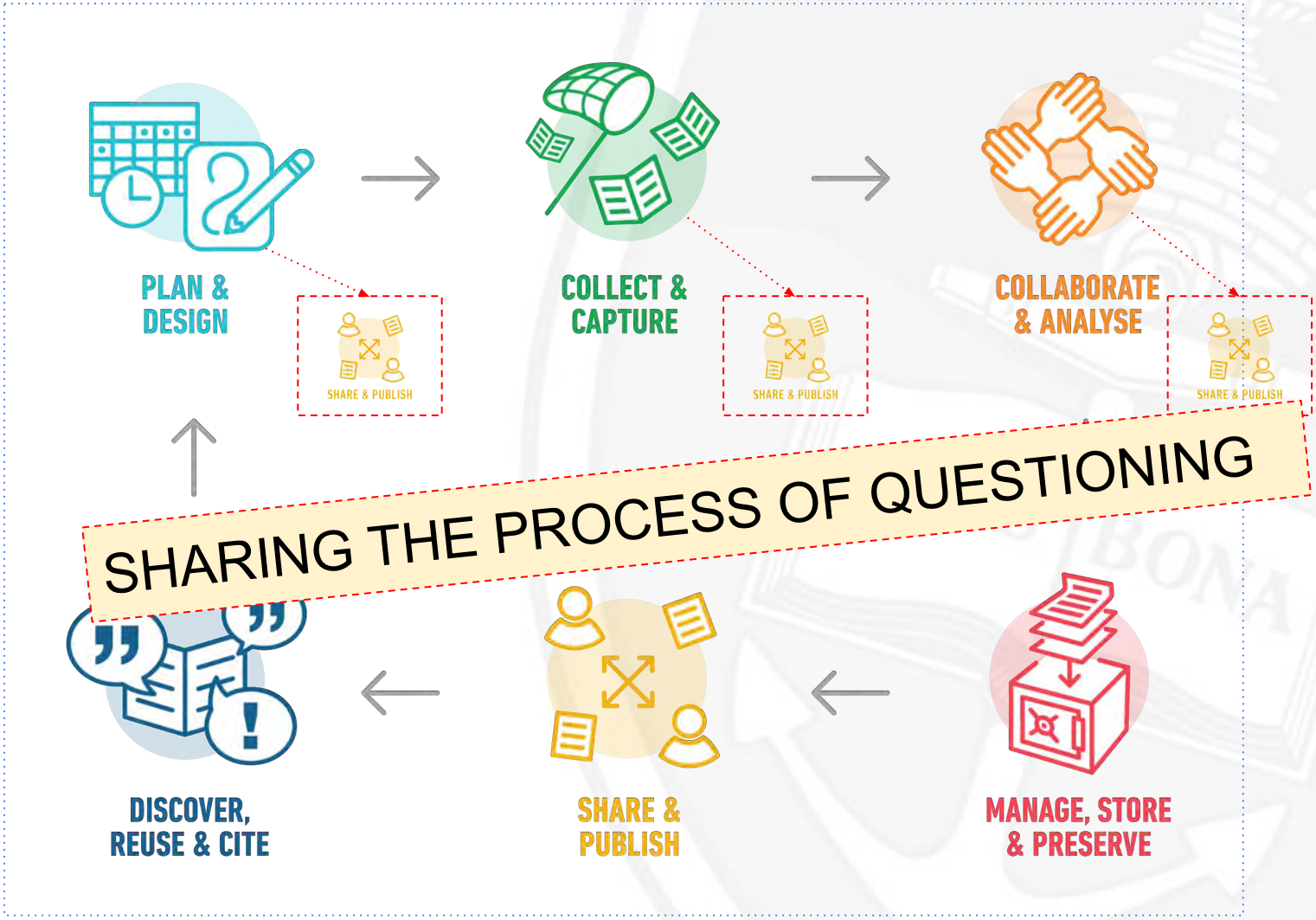


Source: UCT RDM **Why Open Science**: https://commons.wikimedia.org/wiki/File:UCT_RDM_Why-Open-Science.png

The research project (& its data) lifecycle



The OPEN SCIENCE research cycle



How far does Open go?

"As open as possible …

as closed as necessary."

H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Available at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

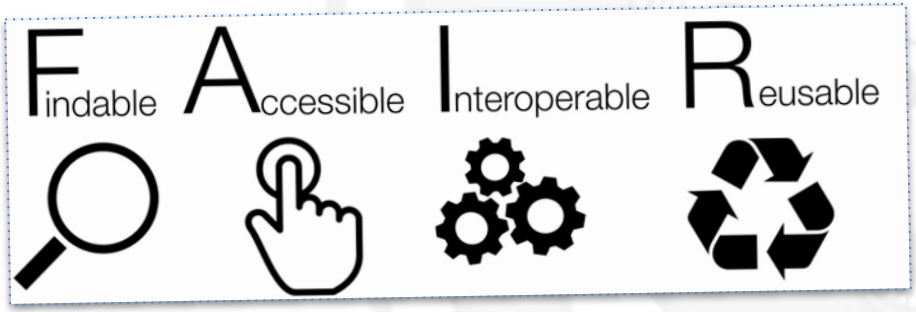
MANAGEMENT - Data about Data

- **Digital Scholarship** tools exist to help us work with our data - keep track, store, secure, cleanup, analyze, collect, collaborate, visualize and organize
- Part of management is having sufficient and clear data describing the data and what has happened to it - or ***metadata***
- In the digital and networked world, ***Metadata*** becomes the currency of exchange that enables data to link with other data and researchers. It contributes to sharing and Open Science.

Compiled from: LibGuides@ Macalester University. Available at: https://libguides.macalester.edu/c.php?g=527786&p=3608583

What is Research Data Management (**RDM**)?

- The **organisation and documentation** of the data processes (collection, description, de-identification, curation, archiving and publication) within a research project.
- Contributes to **Open Science**: professional data management practices make research more coherent and shareable.

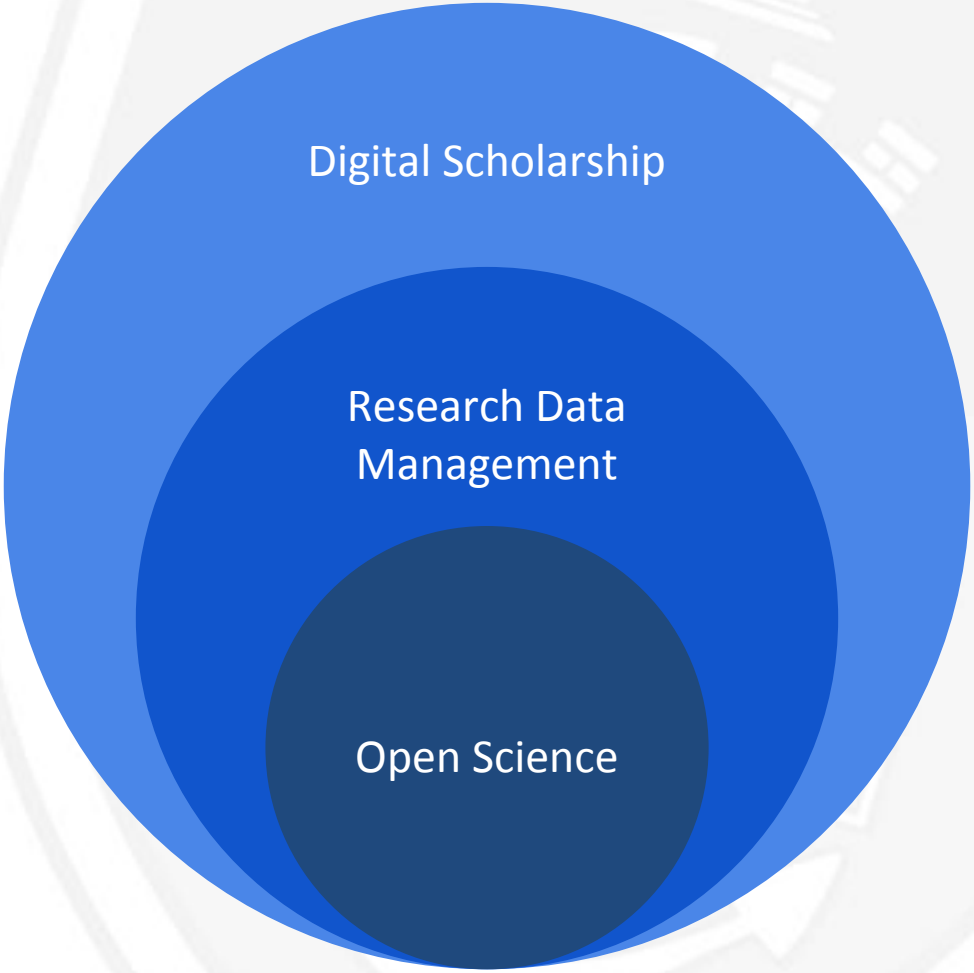


- Good **Digital Scholarship** practices along every step of the research lifecycle help data management and enable Open Science.

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

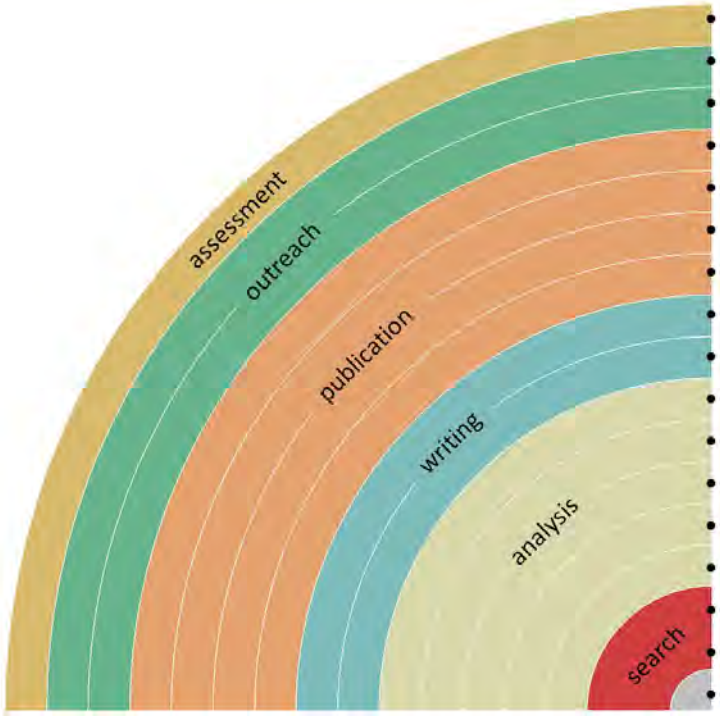
DS - RDM - OS

As a **digital scholar**, practicing **Research Data Management** helps you be more efficient with your research project, and enables you to contribute to **Open Science.**



Open science workflows & tools

You can make your workflow more open by …



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

 Bianca Kramer & Jeroen Bosman https://101innovations.wordpress.com

DOI: 10.5281/zenodo.1147025

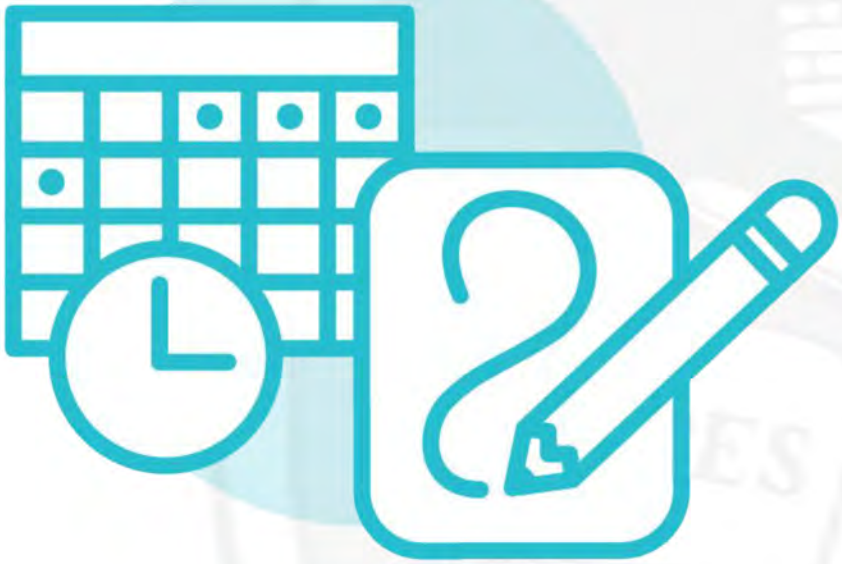
Source: *Foundations for Open Scholarship Development*. Available: https://open-scholarship-strategy.github.io/site/

some UCT-provided systems

DIGITAL LIBRARY SERVICES

The Research Data Lifecycle

Digital Scholarship tools and methods to assist you with Research Data



PLAN & DESIGN

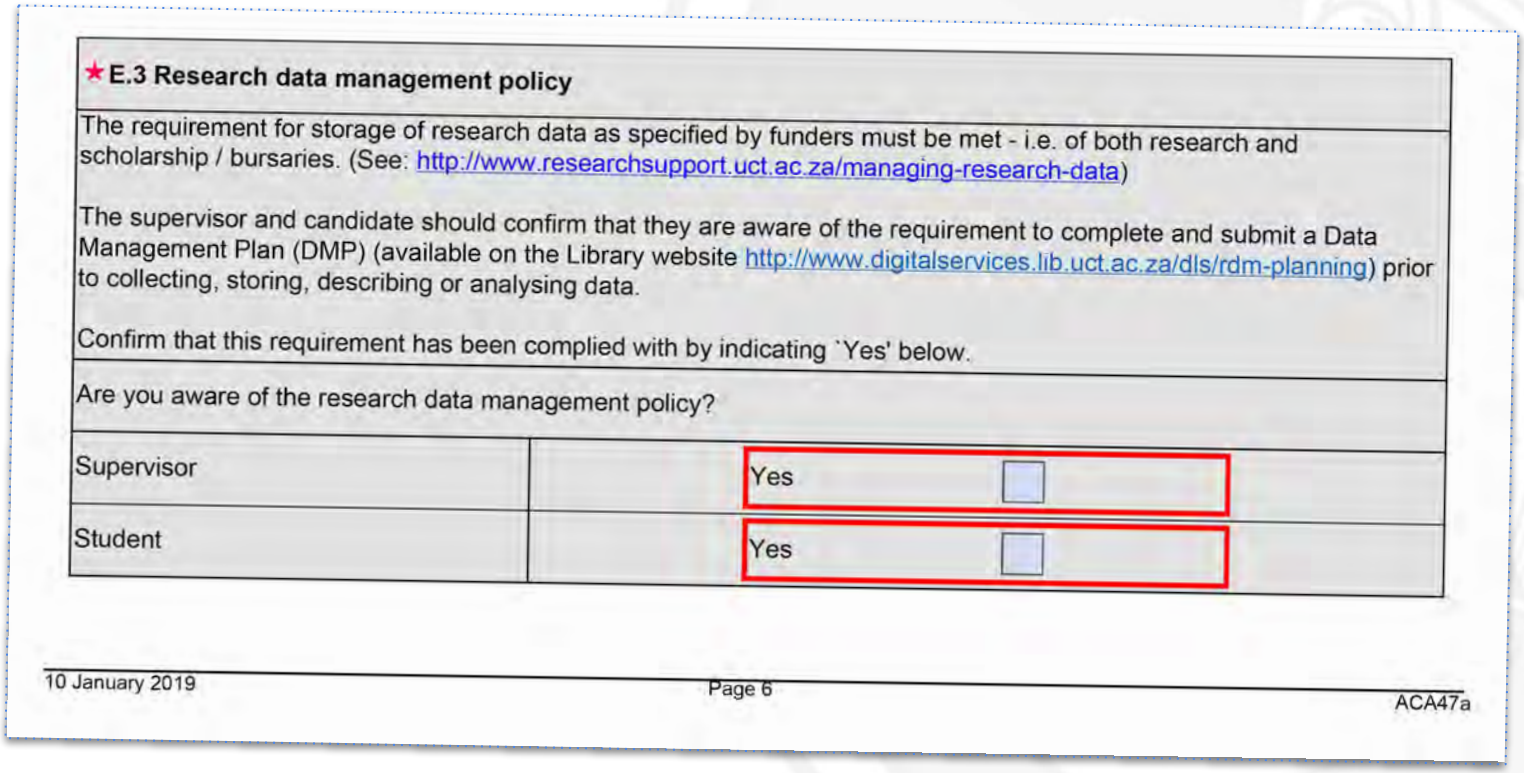
What is a DMP & why create one?

- 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 **requirement** by many **funders** (NIH, Wellcome Trust, NRF).
- 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.
- It provides **guidance for curation-specific activities**, such as file-naming, archiving, formats suitable for long-term preservation, etc.

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

The new **student MoU** at UCT

Institutional requirement: In 2019, a new **student MoU** (Memorandum of Understanding) was implemented for all postgraduate researchers, requiring the creation of a DMP as part of the registration process:

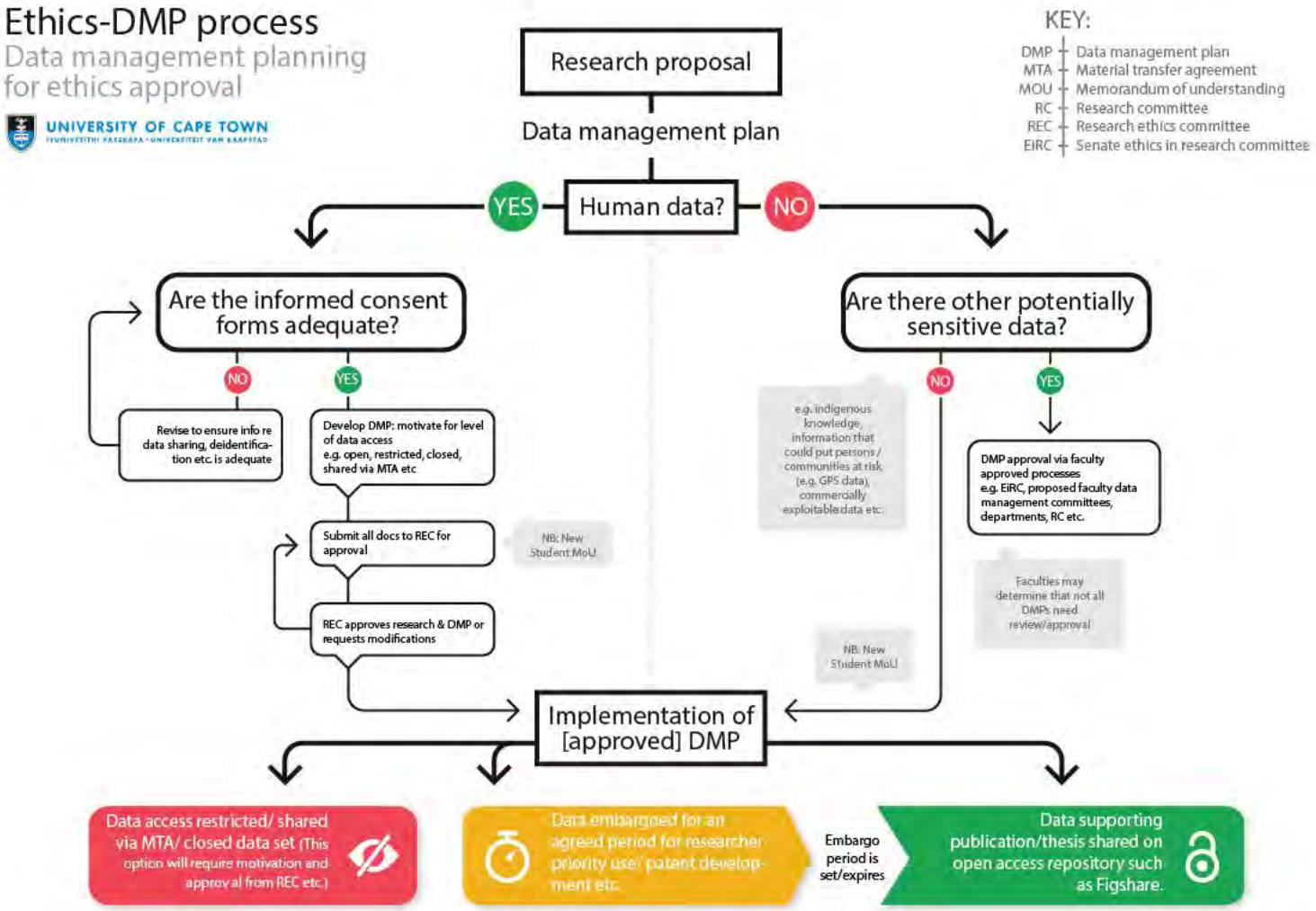


★E.3 Research data management policy	
The requirement for storage of research data as specified by funders must be met - i.e. of both research and scholarship / bursaries. (See: http://www.researchsupport.uct.ac.za/managing-research-data) The supervisor and candidate should confirm that they are aware of the requirement to complete and submit a Data Management Plan (DMP) (available on the Library website http://www.digitalservices.lib.uct.ac.za/dls/rdm-planning) prior to collecting, storing, describing or analysing data. Confirm that this requirement has been complied with by indicating `Yes' below.	
Are you aware of the research data management policy?	
Supervisor	Yes ☐
Student	Yes ☐

10 January 2019 Page 6 ACA47a

For more information, see: OSF Guides > Best Practices > Handling Data > **Creating a data management plan (DMP)**. Available: http://help.osf.io/m/bestpractices/l/618674-creating-a-data-management-plan-dmp

Data Management Planning & Ethics at UCT



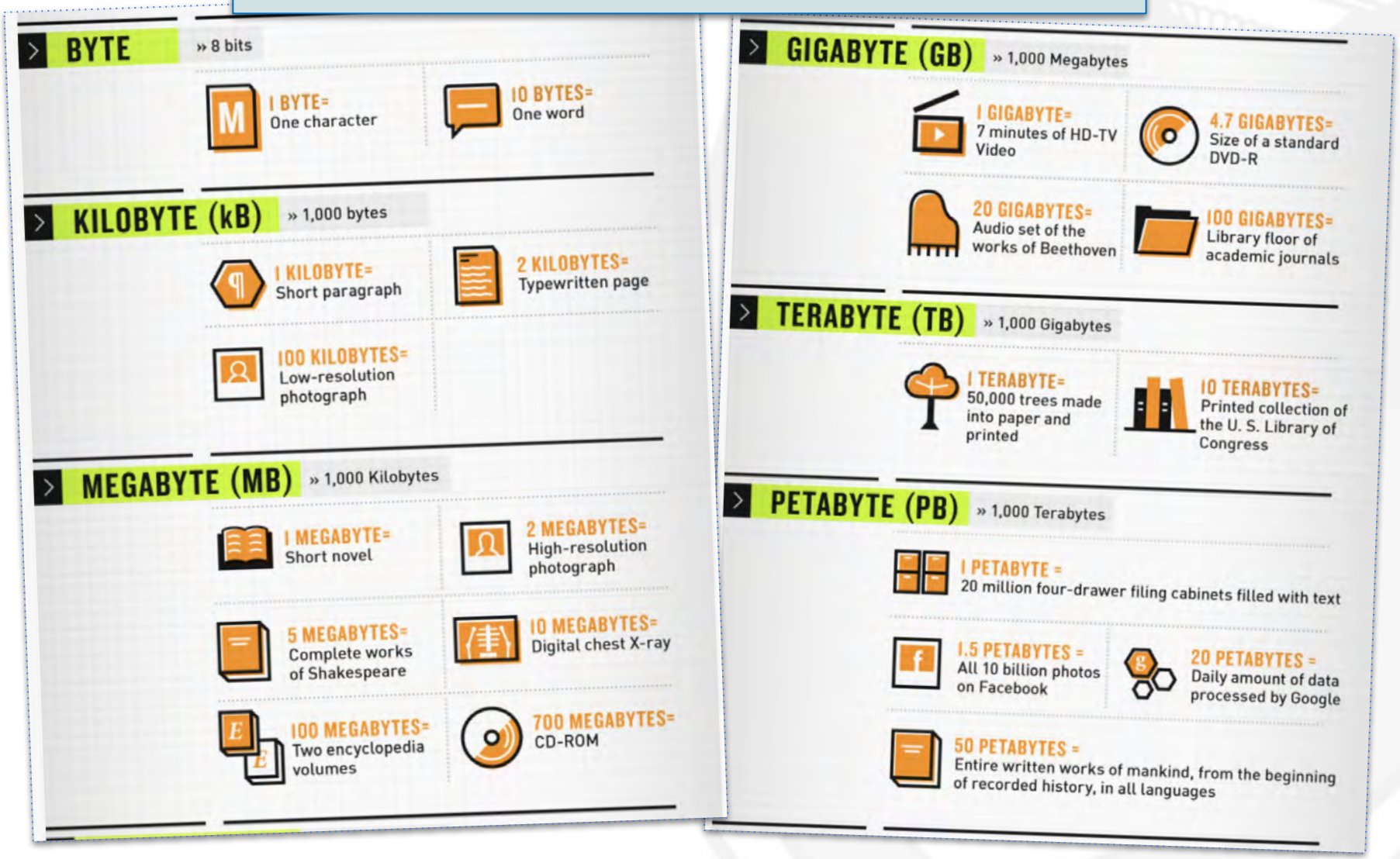
Source: https://commons.wikimedia.org/wiki/File:UCT_RDM_DMP-for-ethics-approval.png

"**What** (are my) **research data?**"

… any information collected, stored, and processed to produce and validate original research results.

Compiled from: LibGuides@ Macalester University. Available at: https://libguides.macalester.edu/c.php?g=527786&p=3608583

Data sizes



Source Data Science Berkely: https://datascience.berkeley.edu/big-data-infographic/

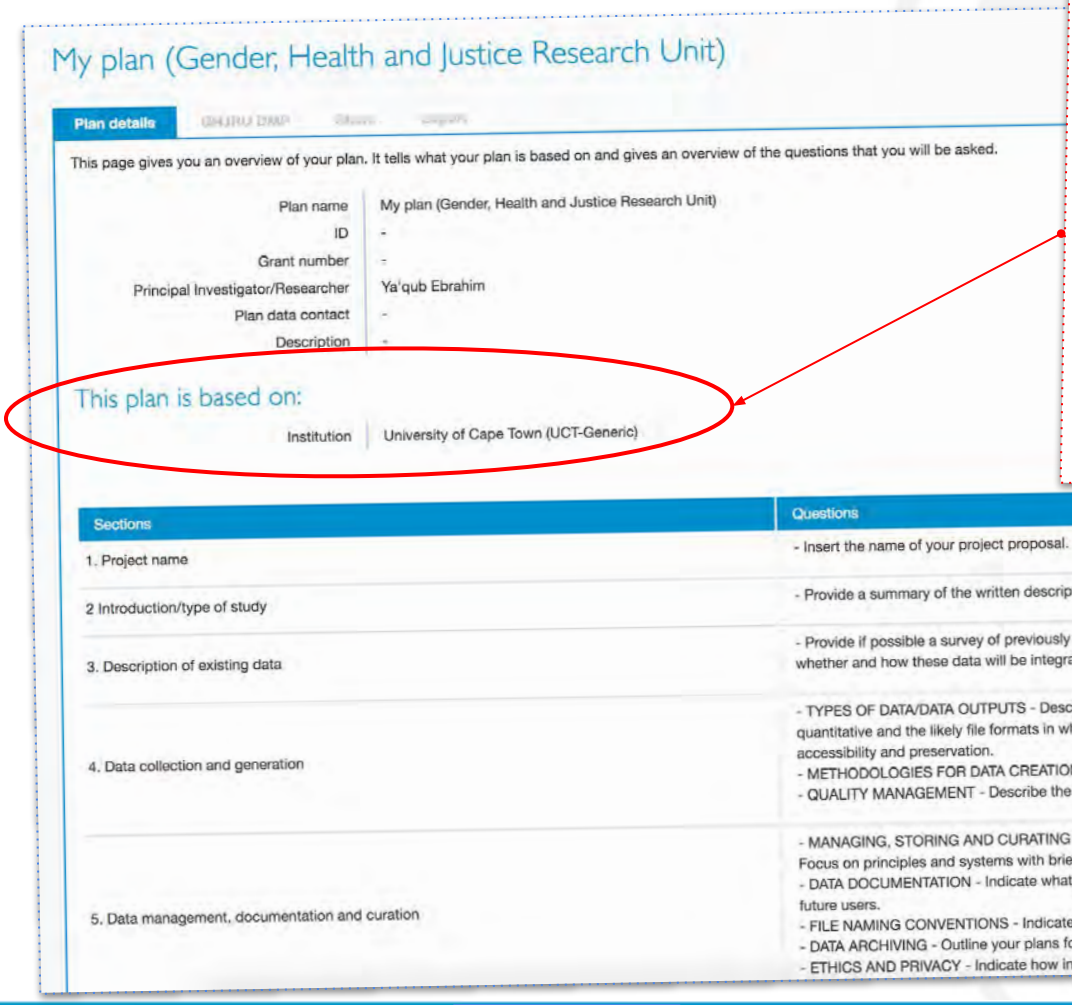
Typical DMP questions

- **What type of data** will be generated in your research?
- How will your data be **named and referenced**?
- What **file formats** are involved?
- What data and **metadata standards** will you follow?
- Who will **have access** to your data and **how**?
- How and when will you **share** your data, if applicable?
- Will you be **digitally preserving** your data? If yes, how so?
- How will you **license** your datasets?
- How will you ensure **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/l/618674-creating-a-data-management-plan-dmp

The **UCT DMP** platform

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



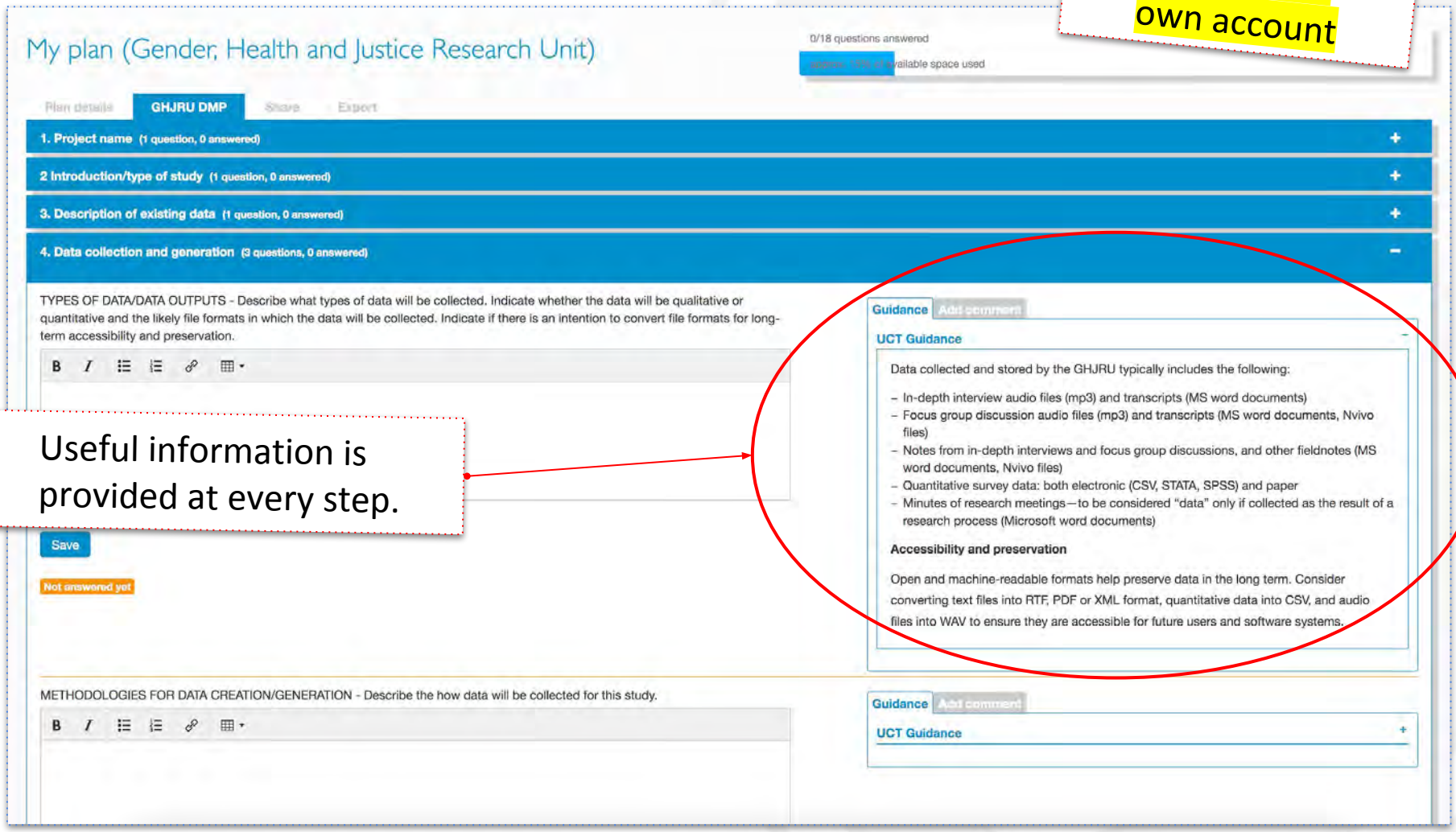
Pick from a variety of templates (funder-specific or generic, i.e. 'UCT') to assist you with planning how you will collect, store, manage and analyse your research data during your research project.

UCT DMP

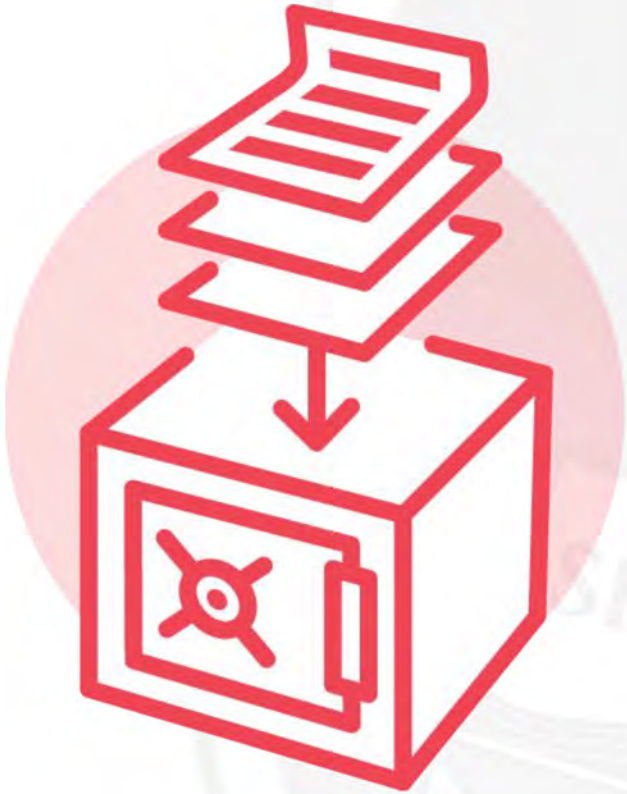
https://dmp.lib.uct.ac.za/

DIGITAL LIBRARY SERVICES

Create your own account



Useful information is provided at every step.



MANAGE, STORE, PRESERVE

Organizing your folders

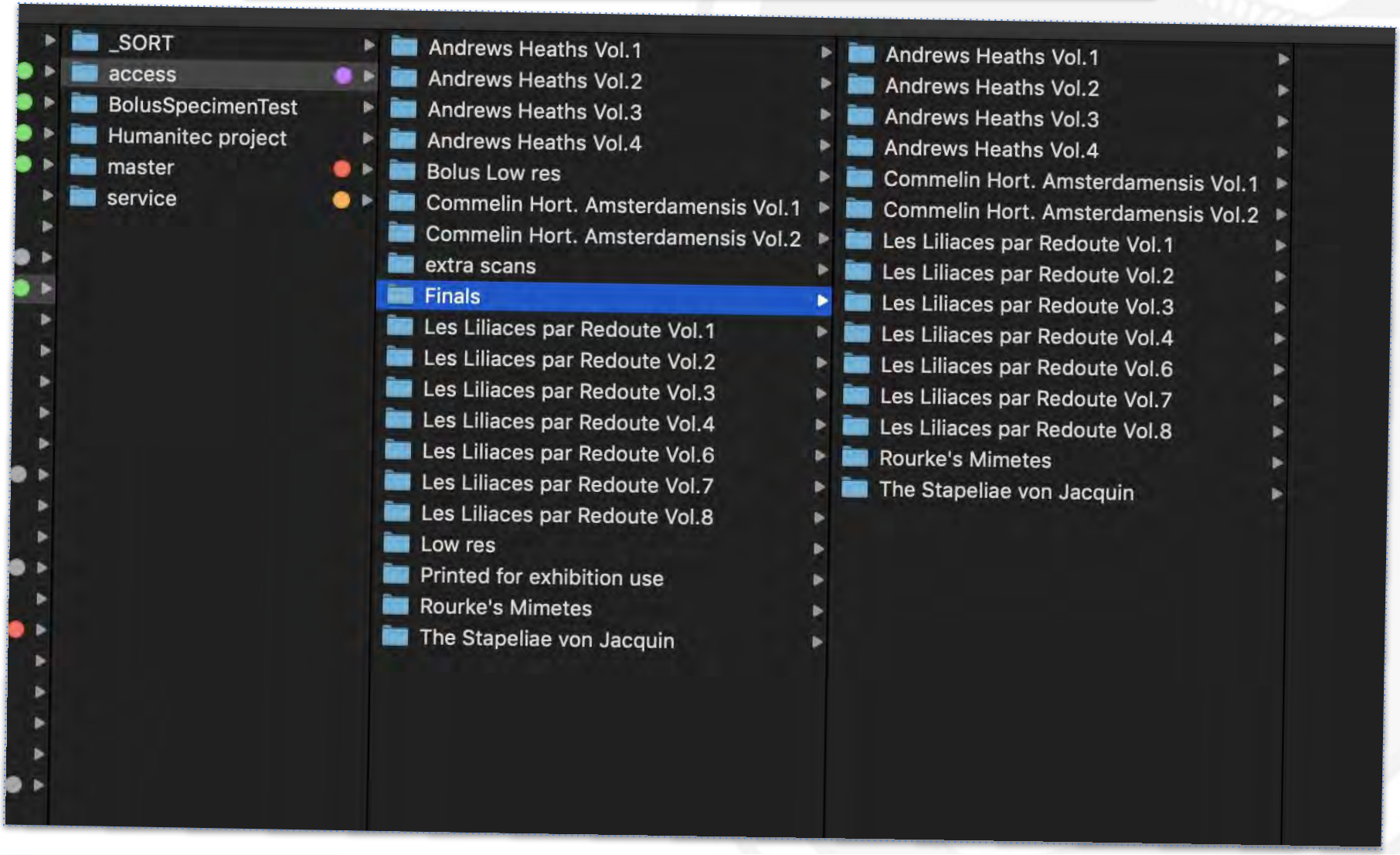
1. **Use folders** - group files within folders so information **on a particular topic is located in one place**
2. Adhere to existing procedures - check for **established approaches** in your team or department which you can adopt
3. **Name folders appropriately** - name folders after the areas of work to which they relate and not after individual researchers or students. This avoids confusion...
4. **Be consistent –** when developing a naming scheme for your folders it is important that **once you have decided on a method, you stick to it**. If you can, try to agree on a naming scheme from the outset of your research project
5. **Structure folders hierarchically** - start with a **limited number of folders for the broader topics**, and then create more specific folders within these
6. **Separate ongoing and completed work** - separate your older documents from those you are currently working on
7. **Review records** - assess materials regularly or at the end of a project to ensure **files are not kept needlessly**.

Adapted from University of Cambridge: https://www.data.cam.ac.uk/data-management-guide/organising-your-data

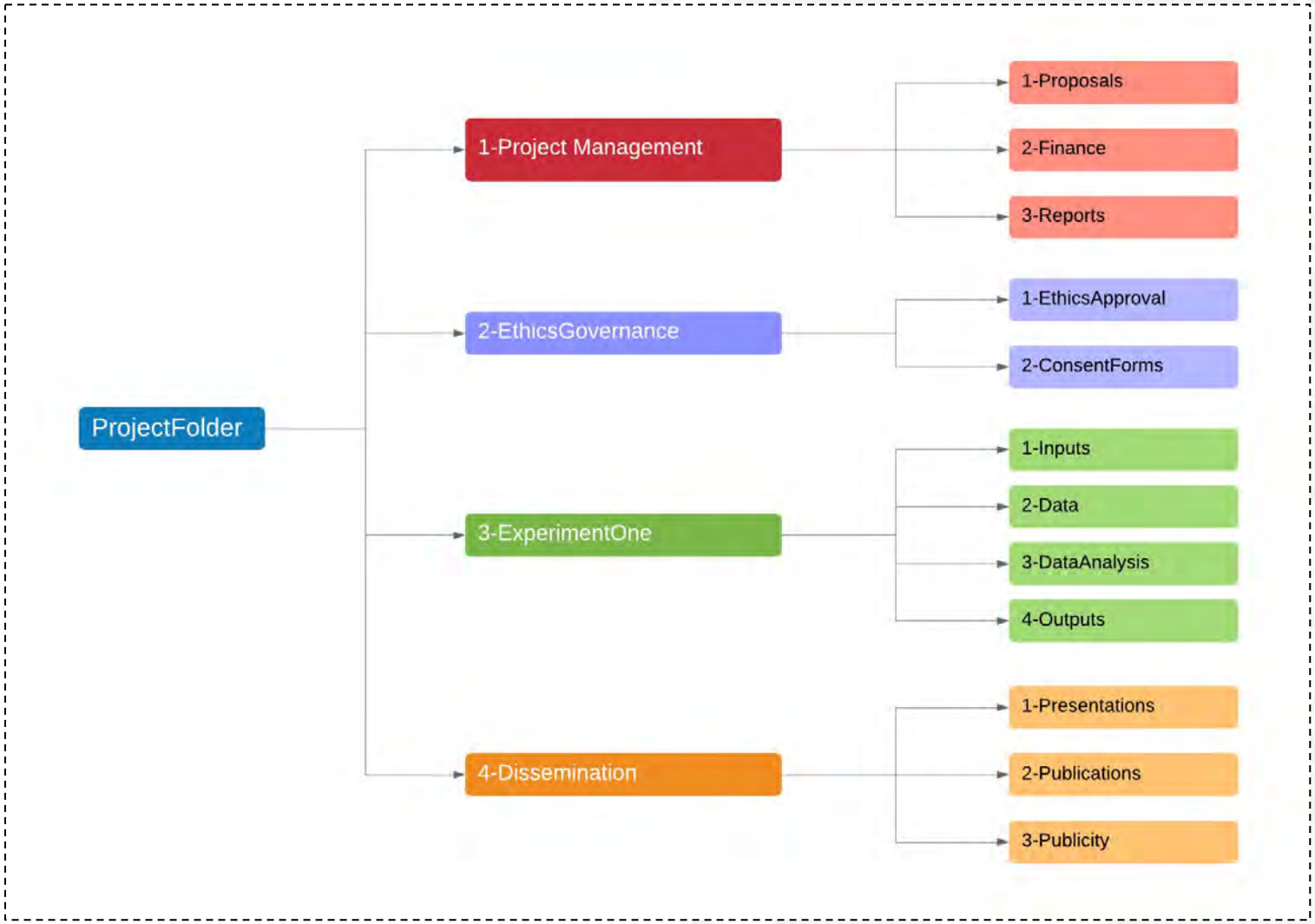
Where to store? - Look towards the clouds

1. Local Storage
2. Flash Drives
3. Cloud Tools
    a. Microsoft One Drive
        i. office.com
    b. Google drive
        i. go to drive.google.com and enter your UCT address
        ii. follow up with your UCT credentials
    c. Both
        i. Work from browser
        ii. Have an app that allows you to create a “local” folder that is synced (just don’t point them both to the same folder)
        iii. Size limit is flexible
        iv. Allow for collaboration
        v. Offer a suite of office tools (Word/Doc; Excel/Sheets; Powerpoint/Slides)

never call something “final”



Organizing your folders



Source: Nikola Vuković available at http://nikola.me/folder_structure.html

Using a file naming convention

1. The computer arranges **files by name, character by character**. Therefore, put the **most important information first**.
2. A good format for date designations is **YYYY-MM-DD** (see: The Problem with Dates: Applying ISO 8601 to Research Data Management).
3. When using a sequential numbering system, **use leading zeros** to make sure files sort in sequential order. Use "001, 002, ...010" instead of "1, 2, ...10"
4. Use **versioning** to indicate the most current version, e.g. **filename_v02.xxx**
5. Try **not** to make file names **too long**. Consider storing helpful metadata in a master spreadsheet that can be stored with your data for future reference. (see: Guide to writing "readme" style metadata)
6. **Avoid special characters**, such as: ~ ! @ # $ % ^ & * ( ) ` ; : < > ? . , [ ] { } ' " |
7. **Do not use spaces** as some software will not recognize file names with spaces.
8. **Punctuation** – decide on conventions for if and when to use punctuation symbols, capitals, hyphens and spaces.
9. Use **unique names** - do not give the same file name in different folders
10. Use "-" to separate connected items, and "_" for unrelated items

Source Harvard Biomedical Data Management: https://datamanagement.hms.harvard.edu/file-naming-conventions

File Naming Conventions

*Files **without** a naming convention:*	*Files **with** a naming convention:*
Test data 2016.xlsx	2016-01-04_ProjectA_Ex1Test1_SmithE_v1.xlsx
Meeting notes Jan 17.doc	2016-01-04_ProjectA_MeetingNotes_SmithE_v2.docx
Notes Eric.txt	ExperimentName_InstrumentName_CaptureTime_ImageID.tif
thesis-final.docx	2008_Scholtz_R.pdf

Source Harvard Biomedical Data Management: https://datamanagement.hms.harvard.edu/file-naming-conventions

Signs you might not be managing your data

1. Your only copy of your data is on a flash that you left behind in a Postnet
2. You know you saved it somewhere but the search function is not finding it and you can't remember the file name
3. Your collaborator can't make sense of the contents of the files: ie What does the column *love1* stand for?
4. The program that you used to collect the data doesn't work on your updated operating system and you can't open it anymore
5. You have four versions of the same file and can't tell which one is the right one, is it *final.docx*, or *final_final.docx*, or *copy of final_final.docx* or *thisoneistherightone_final - Copy.docx*

Storage & Backup ≠ **Preservation**

Yes, maintaining **backups** of your stored data is crucial! But this does **not mean** that they are **digitally preserved**. Digital preservation is an institutional endeavour to ensure that data remain accessible and usable **in the long term**, in view of:

- **technological change** (e.g. legacy media & formats)
- **bit-rot** (decay of digital files over time, e.g. on flash drives)
- **link-rot** (decay of identifiers over time, e.g. on websites)
- **media failure** (e.g. 'head crash' on hard drives, CD-Rs oxidising)

Digital preservation is generally handled by specialist staff, such as archivists and librarians, using dedicated hard- and software solutions. Researchers need to be aware that some of their data may legally require digital preservation, and ideally participate actively in the process of planning for it from the outset (see: DMP).

DIGITAL LIBRARY SERVICES

Digitisation for Digital Preservation & Access

http://www.digitalservices.lib.uct.ac.za/dls/what-we-digitise

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

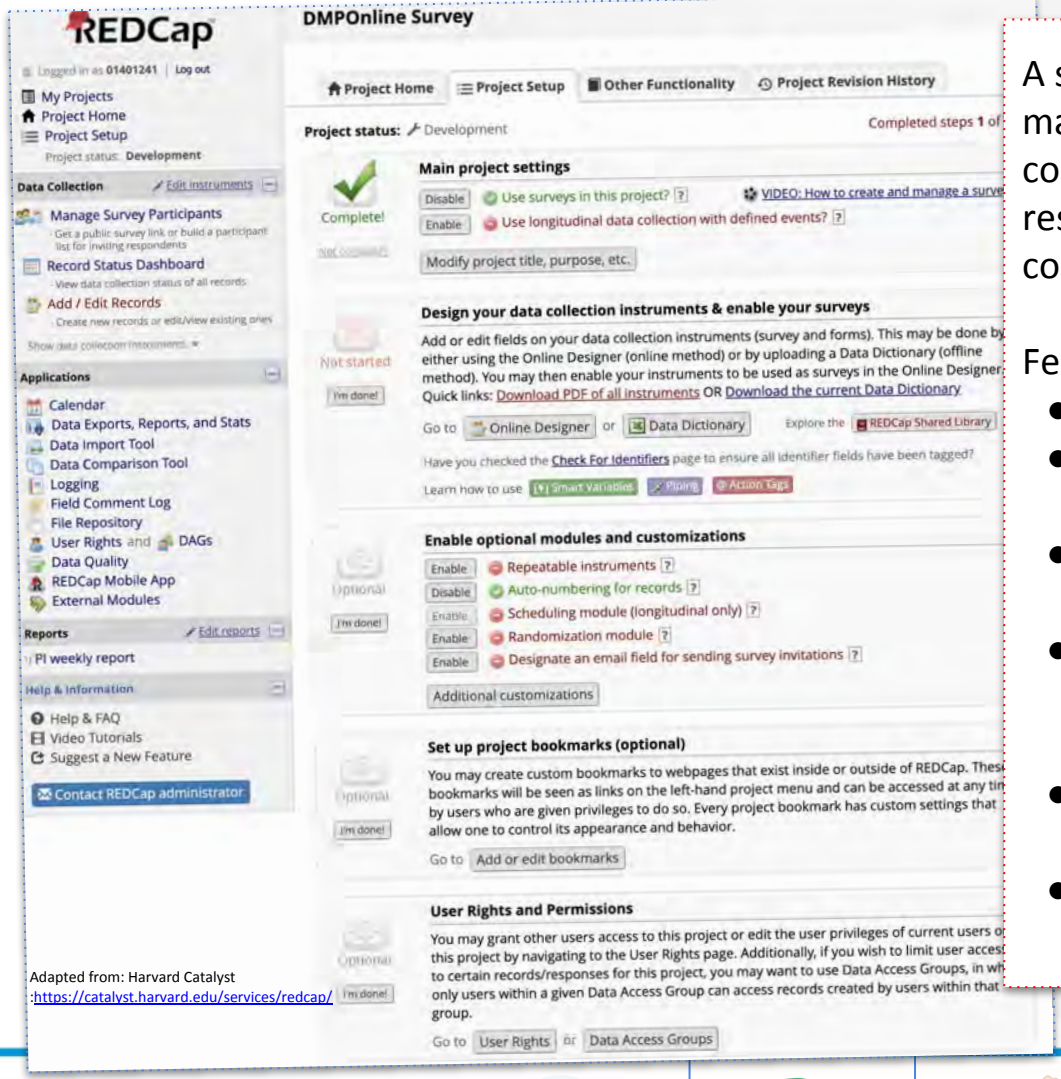
!

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

COLLECT & CAPTURE

RedCap

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



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

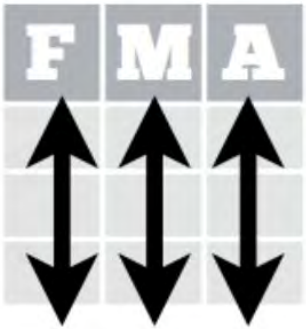
Make peace with spreadsheets

1. ORGANIZED DATA LIVES in TABLES
2. The idea is that in a spreadsheet
   a. each new **observation gets a new row**
   b. each **variable a new column.**
3. DO NOT USE
   a. merged cells,
   b. subheadings,
   c. color used to denote information,
   d. different data types within cells (numbers and letters),
   e. more than one piece of data in a cell (such as disaggregations).
4. If data is tidy, so many cool things can be done
   a. data visualization
   b. controlled vocab
   c. duplicate removal

Adapted from Merl Tech:http://merltech.org/making-some-sense-of-data-storage-and-presentation-in-excel/

Make peace with spreadsheets

In a tidy data set:



&

Each **variable** is saved in its own **column**

Each **observation** is saved in its own **row**

	country	continent	year	lifeExp	pop	gdpPercap
1	Afghanistan	Asia	1952	28.801	8425333	779.4453
2	Afghanistan	Asia	1957	30.332	9240934	820.8530
3	Afghanistan	Asia	1962	31.997	10267083	853.1007
4	Afghanistan	Asia	1967	34.020	11537966	836.1971
5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	Afghanistan	Asia	1977	38.438	14880372	786.1134
7	Afghanistan	Asia	1982	39.854	12881816	978.0114

Adapted from Julie Lowdee: https://jules32.github.io/2016-07-12-Oxford/dplyr_tidyr/#35_other_tidyr_awesomeness

Advice for the '*Collect & Capture*' phase

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

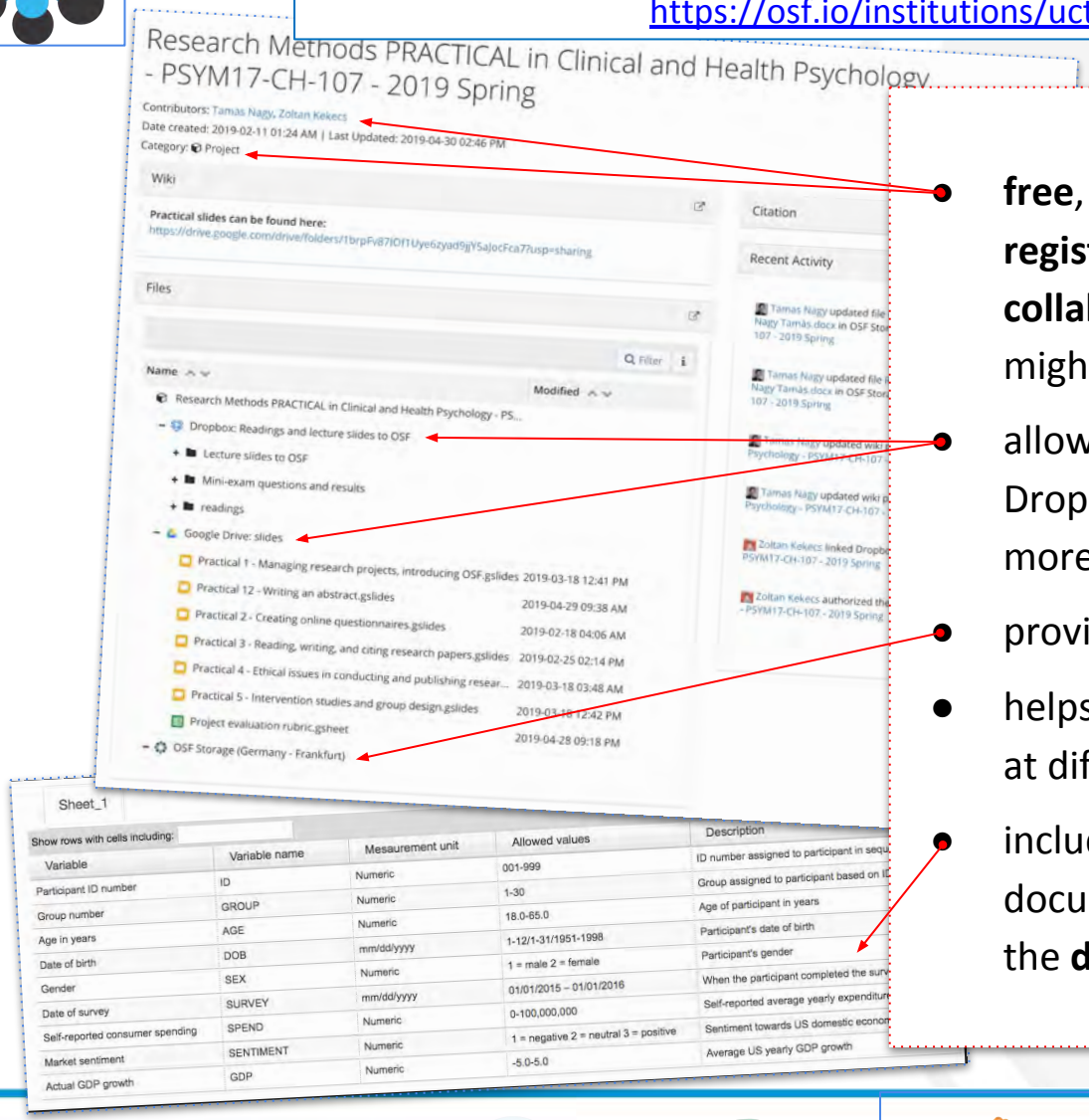
- Describe the type of data generated:
  - The **form** (*What kind of data does it hold?*)
  - The **stability** of each dataset (*How does it change over time?*)
  - Create **unique names** for each of your datasets
- Document the data you are capturing, and how you are identifying it within each data set by building a **data dictionary**.
- **Document your process** and store it together with your data (e.g. readme.txt).

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

COLLABORATE & ANALYSE

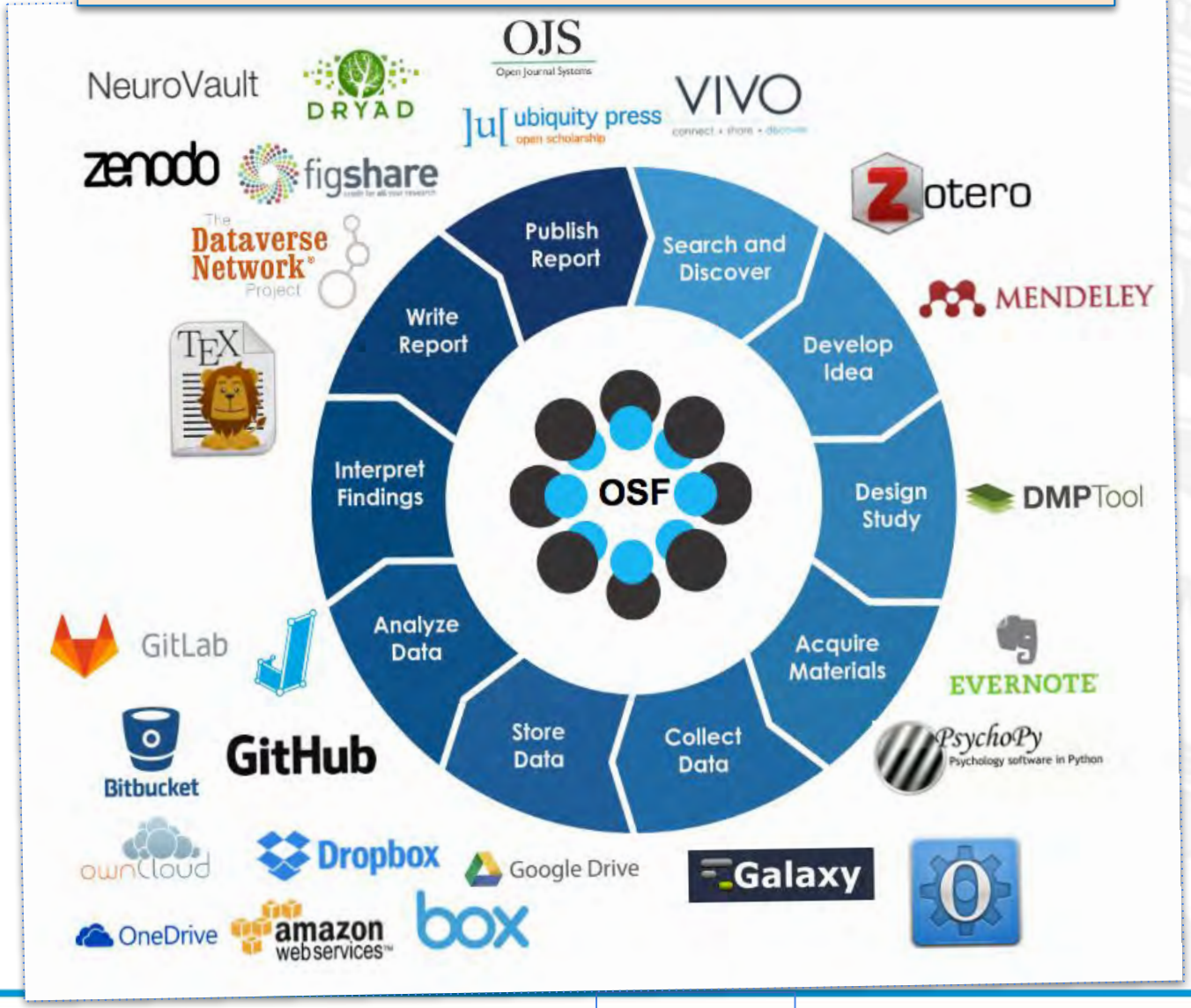
Open Science Framework (OSF)

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



- **free**, online platform that allows you to **register** your project, **manage collaborators,** and **centralise data** that might be stored at different locations
- 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**

OSF integrations along the research life-cycle



Open Science Framework (OSF)

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

- The UCT OSF service is an online platform that allows you to register your project, manage stakeholders, and centralise data that might be stored at different locations with different collaborators.

- Create an account in order by clicking on 'Sign in' on the home page, and then select ' Login through Your Institution.' Select *University of Cape Town*, and enter your SSO credentials after that.

- There are a lot of other features to assist you with managing the research project's workflows and procedures, as well as get DOIs for working data (NB: not uct-specific).

> further reading

Advanced digital scholarship

Data Analysis and Mining:

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

- Tools and processes for data de-identification, to safeguard privacy of patients.
- Tools and process for text analysis.

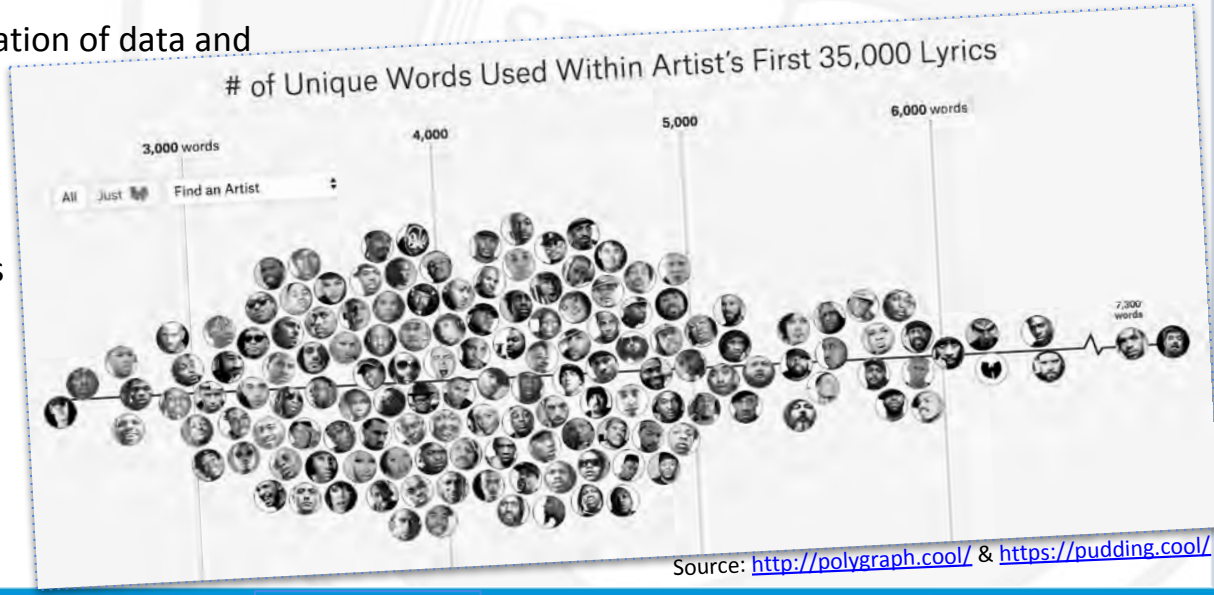
Source: **Tableau**, see: https://www.tableau.com/

Data Visualization:

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

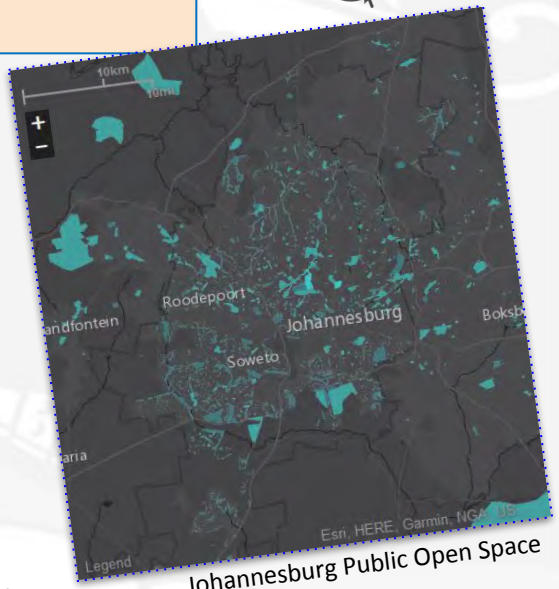
Digital Humanities:

Tools, processes and critical awareness found in the intersection between digital technologies and fields of study within the humanities.



Source: http://polygraph.cool/ & https://pudding.cool/

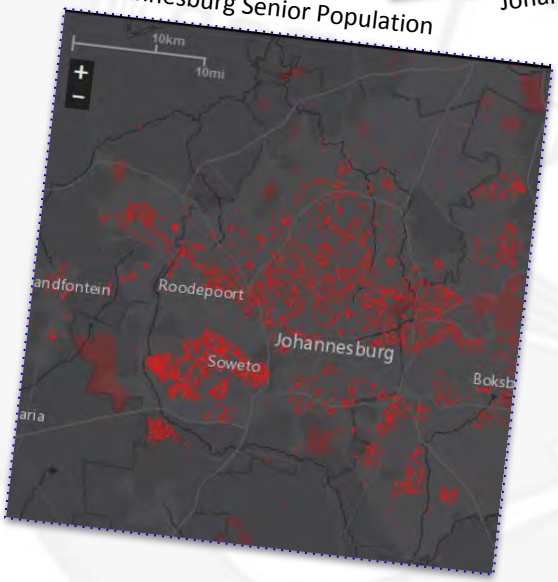
Geographic Information Systems



Johannesburg Public Open Space

Everything Happens Somewhere:

- Because everything happens somewhere everything can be associated with a spatial location.
- These locations can be mapped in space, either for simple visualisation or for complex analyses.



Johannesburg Senior Population

Data Visualisation (Maps):

- Maps are an incredibly powerful visualisation tool which allow us to view and display our data in interesting and informative ways. They allow us to see patterns in our data, not just find them.
- They also allow us to communicate our findings in a clear and succinct manner.

Images sourced from UrbanObservatory.org's App

GIS

Data Analysis (Making Information):

The full potential of GIS is realised when performing spatial analyses. Different types of analyses exist to satisfy various needs:

- **Overlay** Analysis allows us to compare different data types, e.g. Mean Annual Rainfall and Crop Type.
- **Geostatistical** Analysis allows us to perform statistical analyses of correlated spatial data, e.g. Hotspot Analysis.
- **Network** Analysis allows us to calculate travel times and service delivery areas, e.g. “Golden Hour” coverage or Clinic’s Service Area.
- **Dashboards** of real time sensor feeds for live monitoring, e.g. Resource Usage; Traffic Volumes; Fleet Management.

Difference in Malaria rates between 2000 and 2015.
From the urbanobservatory.org

DLS’ GIS services assist with GIS software acquisition, project planning, troubleshooting, analysis and cartographic design.

Find us @ www.gis.uct.ac.za

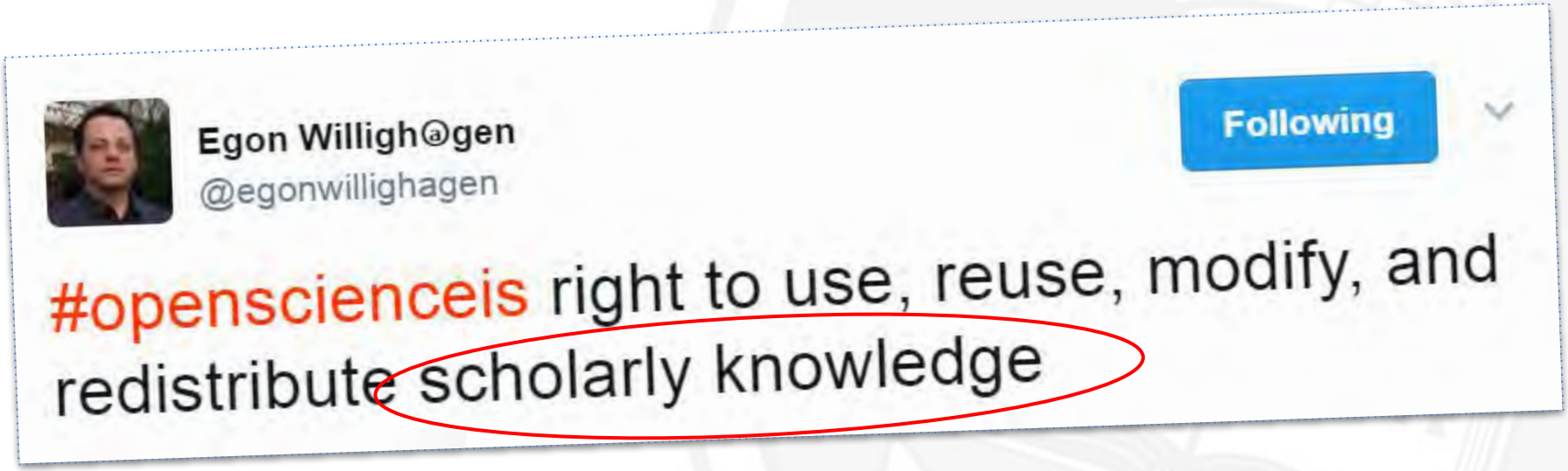
Selected analysis tools

tool	web address	what for?
Tableau	https://public.tableau.com/	interactive data visualization software
ArcGis	https://www.arcgis.com/home/index.html	tool for developing your own maps and analyzing spatial data
Amnesia	https://amnesia.openaire.eu/	data anonymization web-tool, that allows you to remove identifying information from data.
Rstudio	https://rstudio.com/	a set of integrated and productive tools for statistical computing and graphics programming language R
NVivo 12	https://www.qsrinternational.com/nvivo/home	qualitative data analysis software
SPSS	https://www.ibm.com/za-en/analytics/spss-statistics-software	software package used for interactive, or batched, statistical analysis.
ATLAS.ti	https://atlasti.com/	workbench for the qualitative analysis of large bodies of textual, graphical, audio and video data
RedCap	https://trn-redcap.uct.ac.za/	collaborative tool for data collection and capture (includes survey tools)
Otter.ai	https://otter.ai	Generate rich notes for meetings, interviews, lectures, and other important voice conversations

Full list here:

DISCOVER, REUSE & CITE

Open Science is ...



Where scholarly knowledge is ...

Open to participation	Open to (re)use	Open to the world

• No barriers based on race, gender, income, status
• Involvement of societal partners in research priority setting
• Evaluations that include societal relevance
• Citizen science
	• Open Access, for people and machines, to:
	  ○ Proposals and applications
	  ○ Data and code
	  ○ Preprints, working papers
	  ○ Papers and books
	  ○ Reviews and comments
	  ○ Posters and presentations
	• Open, non-proprietary standards
	• Open licences
	• Full documentation of process
		• Translations
		• Plain language explanations
		• Outreach beyond academia
		• Open to questions from outside academia
		• Curation and annotation of non-scholarly information
		• Participation in public debate

Source: Bianca Kramer & Jeroen Bosman, 2017. Defining open science definitions. Available at: https://im2punt0.wordpress.com/2017/03/27/defining-open-science-definitions/

Two research data repositories **at UCT**

DataFirst and **ZivaHub** are **registered**, **certified**, and **transparent**, through **independent review**, **standards**, and **policies**.

Adapted from: Zimmer, Niklas; King, Thomas (2018): **Data discovery and re-use**. figshare. Presentation. https://doi.org/10.25375/uct.7358423.v1

A *small* overview of data catalogues, registries and repositories

directly UCT-relevant

- BioLINCC – Clinical specimen database.
- Dataverse – Widely used open source repository system; Example: HARVARD Dataverse
- dataMED – prototype biomedical data search engine to discover data sets across data repositories or aggregators.
- 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.
- Deveo – Free, private Git, Mercurial, and SVN repository management platform.
- Dryad – Data repository system for any files associated with any published article in the sciences or medicine.
- Figshare(.com) – **Free cloud service for managing, sharing & publishing research data.**
- 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 – Information and tools to ensure your articles can be shared with your colleagues easily.
- Open Science Framework – **Open registration, version control & collaboration software system.**
- 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 data, code & computational methods to 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.**
- ZivaHub | Open Data UCT – UCT's digital repository.

Adapted from: Zimmer, Niklas; King, Thomas (2018): **Data discovery and re-use**. figshare. Presentation. https://doi.org/10.25375/uct.7358423.v1

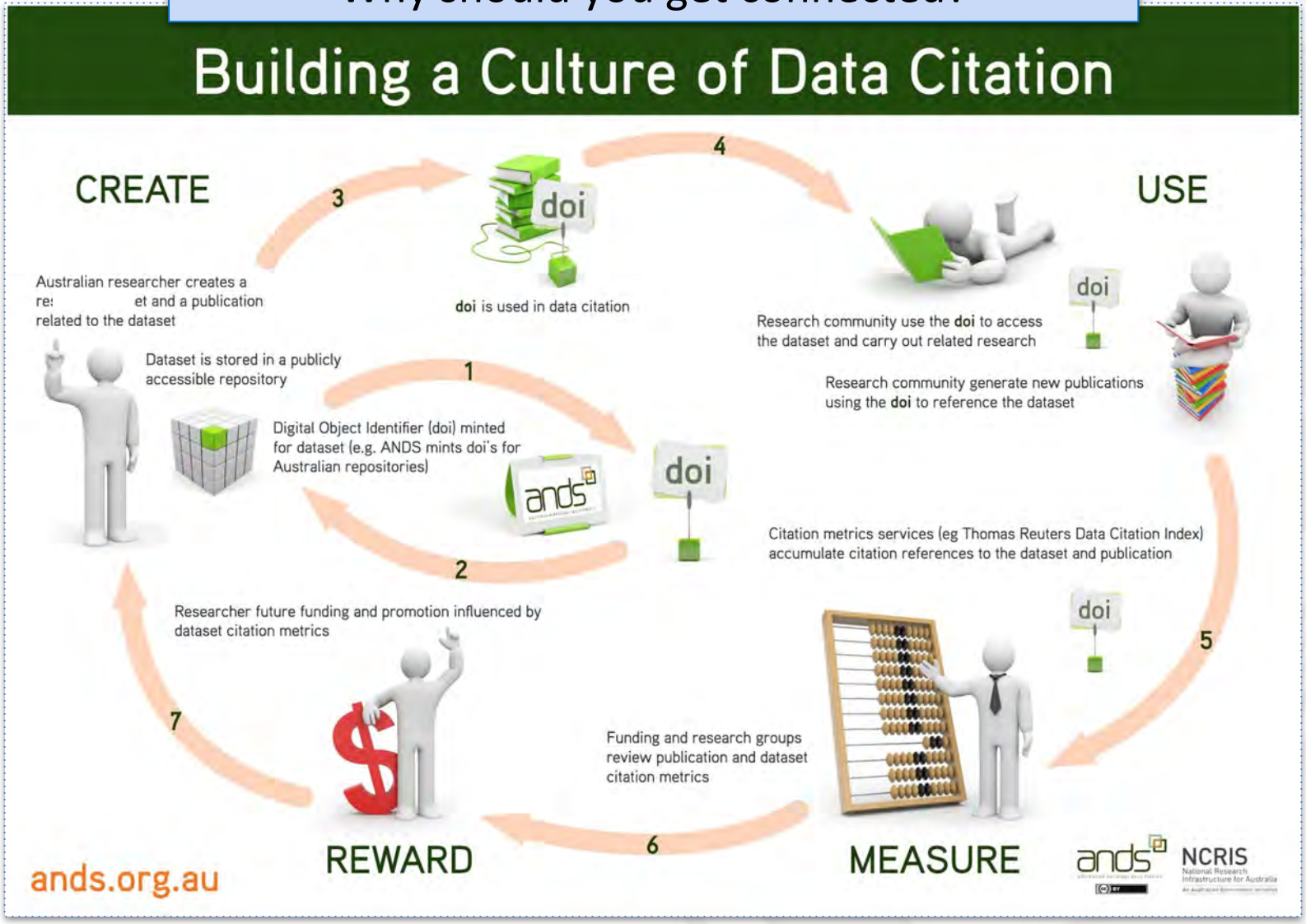
Paths to Open Science

OPEN

DATA	Open data is the process of sharing both the original, raw and the treated or processed data online. This helps others to redo your experiments, and re-use it for additional purposes, helping to verify and accelerate research discoveries.
ACCESS	Allows anyone to access and re-use research published in journal articles without payment or restriction.
PEER REVIEW	Includes publishing review reports, revealing the identity of reviewers, and making peer review a more continuous and collaborative process.
METHODS	Where the process of the research has been documented in a sufficient detail to allow others to *repeat*, *reproduce*, or *replicate* the work.
SOURCE	Much modern research relies on code and software, and Open Source is about providing free access and re-use rights to this to maximise its utility.

Source: J.Tennant; B. Caron; J. Havemann; S. Guay; J. Colomb; E. Lantsoght; E. Tóth-Czifra; K. Kriegel; J. Sègbédji Ahinon; C. Smout & G. O'Neill. (2019, March 16). OpenScienceMOOC. Module-1-Open-Principles 2.0.0 (Version 2.0.0). Zenodo. http://doi.org/10.5281/zenodo.2595951

Why should you get connected?



Source: https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation

SHARE & PUBLISH

Identify yourself :)

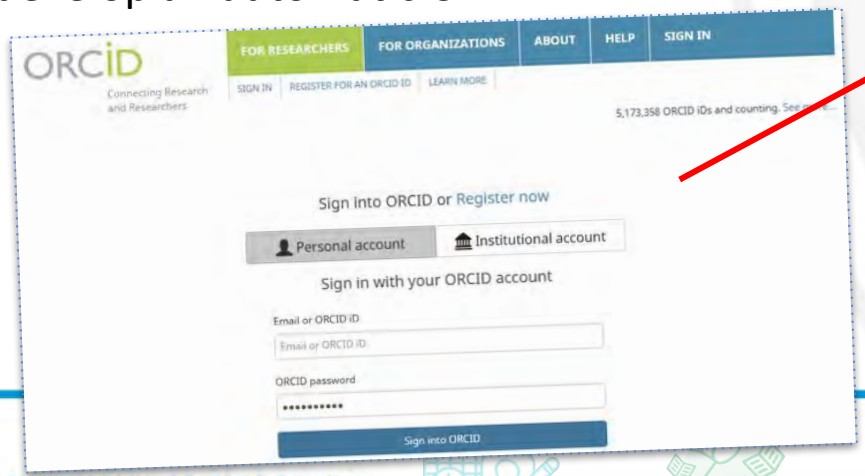
Get your own unique ID:

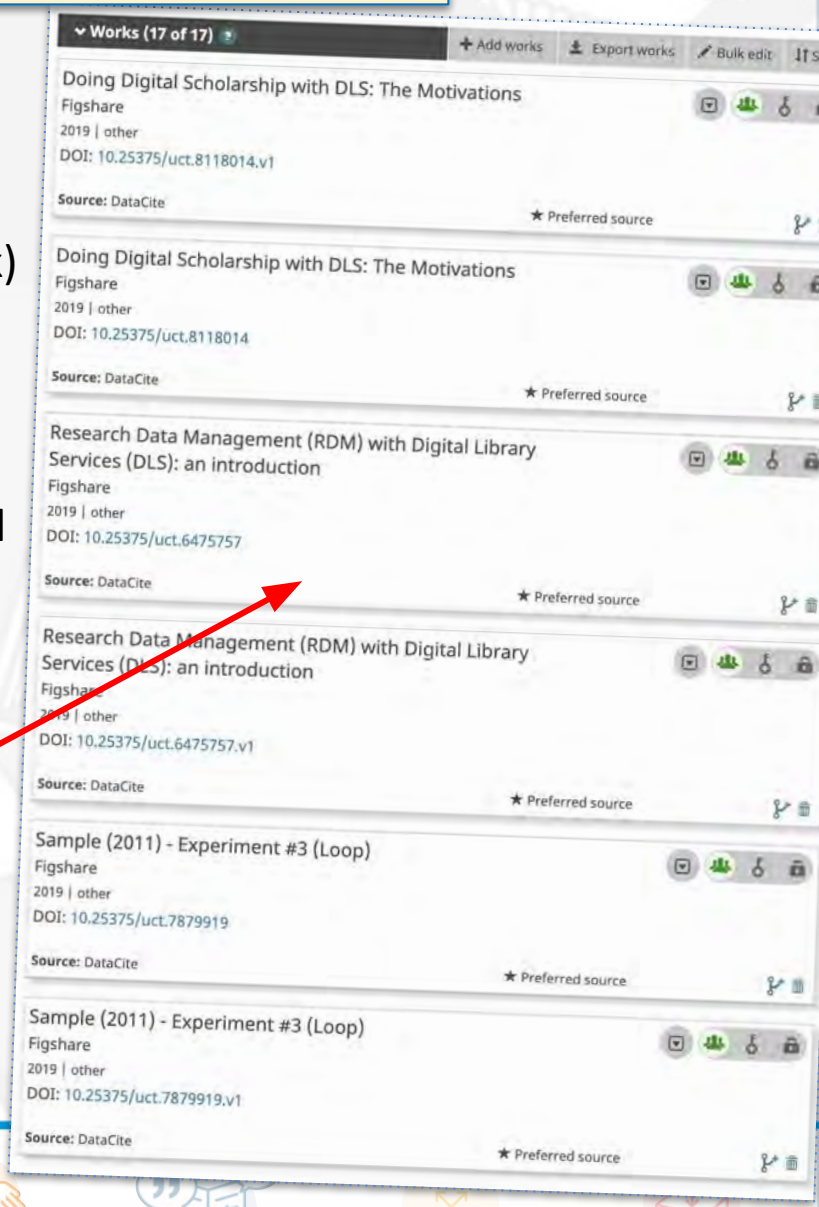
ORCiD is a persistent identifier that supports automated linkages between you and your professional activities (such as your published work) ensuring that your work is recognized

Go to https://orcid.org/ and click 'Sign in'

- Click 'Institutional login' and select UCT
- Add your normal UCT username and password

In future, when submitting journal articles, depositing data, etc. add your ORCiD number to develop an automatic CV





"What (are my) research data?"

… any information collected, stored, and processed to produce and validate original research results.

Compiled from: LibGuides@ Macalester University. Available at: https://libguides.macalester.edu/c.php?g=527786&p=3608583

Selected online academic platforms

tool	web address	what for?
ORCiD	https://orcid.org/	get your own unique identifier as a researcher
Publons (formerly ResearcherID)	https://publons.com/	collects information about peer reviews and builds public reviewer profiles for participating reviewers
Twitter	https://twitter.com/	follow academics and research organizations working in your field
LinkedIn (incl. ...*Learning*)	https://www.linkedin.com/	employment oriented service used for networking
ImpactStory	https://profiles.impactstory.org/	open-source website that helps researchers explore and share the the online impact of their research
Google Scholar	https://scholar.google.com/	freely accessible web search engine that indexes the full text or metadata of scholarly literature
ResearchGate	https://www.researchgate.net/	social networking site for scientists and researchers to share papers, ask and answer questions, and find collaborators.
Meta-Wiki (Wikimedia)	https://meta.wikimedia.org/	wiki which holds information for all Wikimedia projects.
Wikidata	https://www.wikidata.org/	collaboratively edited knowledge base hosted by the Wikimedia Foundation
Humanities Commons	https://hcommons.org/	network for people working in the humanities

Full list here:

Reasons **for** sharing research data

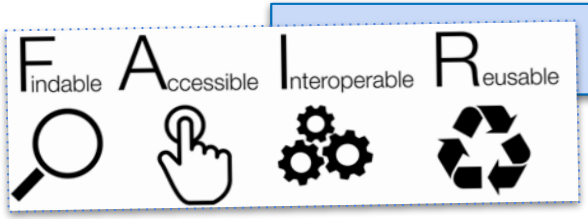
1. **To reproduce** / verify or falsify **research**.
2. To enable others to **ask new questions** of extant data.
3. **To advance the state of research and innovation** (Borgman, 2012).
4. **To confront** some of the **biases** in data collection and analysis (Atici et. al, 2013).
5. To **increase citation rate** (Piwowar, Day and Fridsma, 2007).
6. To **increase the visibility** of researchers and their work online (Peters et al., 2015).
7. **To comply with funding agencies' and institutions' mandates** for the results of scientific studies to be shared with the public.
8. **To comply with publishers** asking authors to deposit underlying datasets in publicly accessible platforms.

Digital Scholarship and DLS | **mission & vision**

...to transform the way research is conducted at UCT by accelerating discovery, increasing the value of research decision-making, and catalysing changes throughout the economy and society that are of value to all citizens.

The University seeks to ensure consistent research practice related to data management principles that support effective **data sharing**, including **open access**; and the need for **data to be discoverable**, **accessible**, **reusable** and **interoperable** to specific quality standards.

Source: DLS website: Policy_Research_Data_Management_2018.pdf



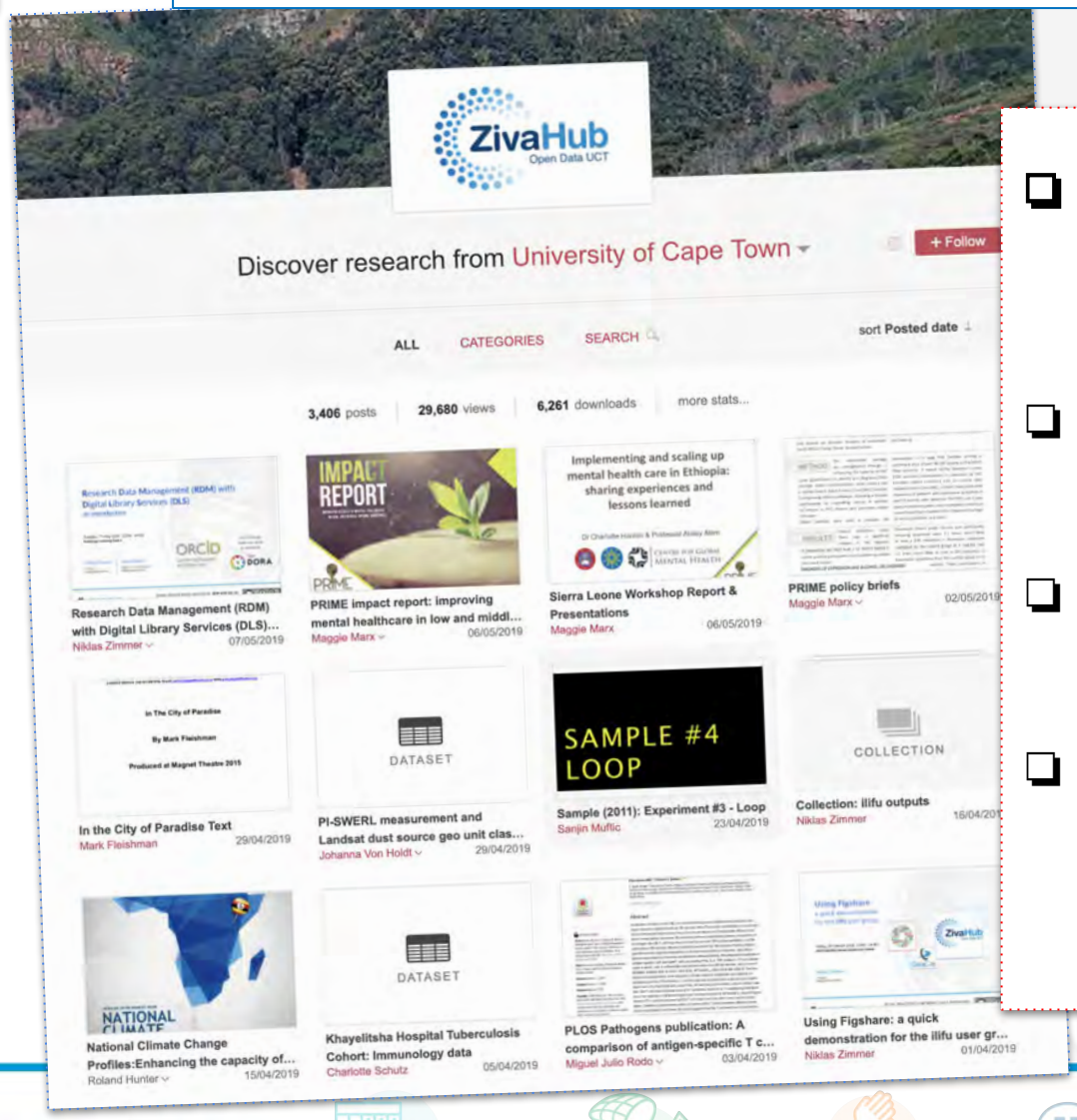
Preppin' for Visitors

Findable	...ensuring that your data can be found by both humans and machines.
Accessible	...once someone has found your data, they need to know how they can get access to them. This could include going through an authorisation and/or authentication process.
Interoperable	...ensuring that your data can be integrated with other data and that they can be utilised by applications or workflows for analysis, storage, and processing.
Reusable	...ensuring that your data - and their related metadata - are well-described and indicate how they can be reused with appropriate licensing.

Source: Jon Tennant; Bruce Caron; Jo Havemann; Samuel Guay; Julien Colomb; Eva Lantsoght; Erzsébet Tóth-Czifra; Katharina Kriegel; Justin Sègbédji Ahinon; Cooper Smout & Gareth O'Neill. (2019, March 16). OpenScienceMOOC/Module-1-Open-Principles 2.0.0 (Version 2.0.0). Zenodo. http://doi.org/10.5281/zenodo.2595951

ZivaHub | Open Data UCT

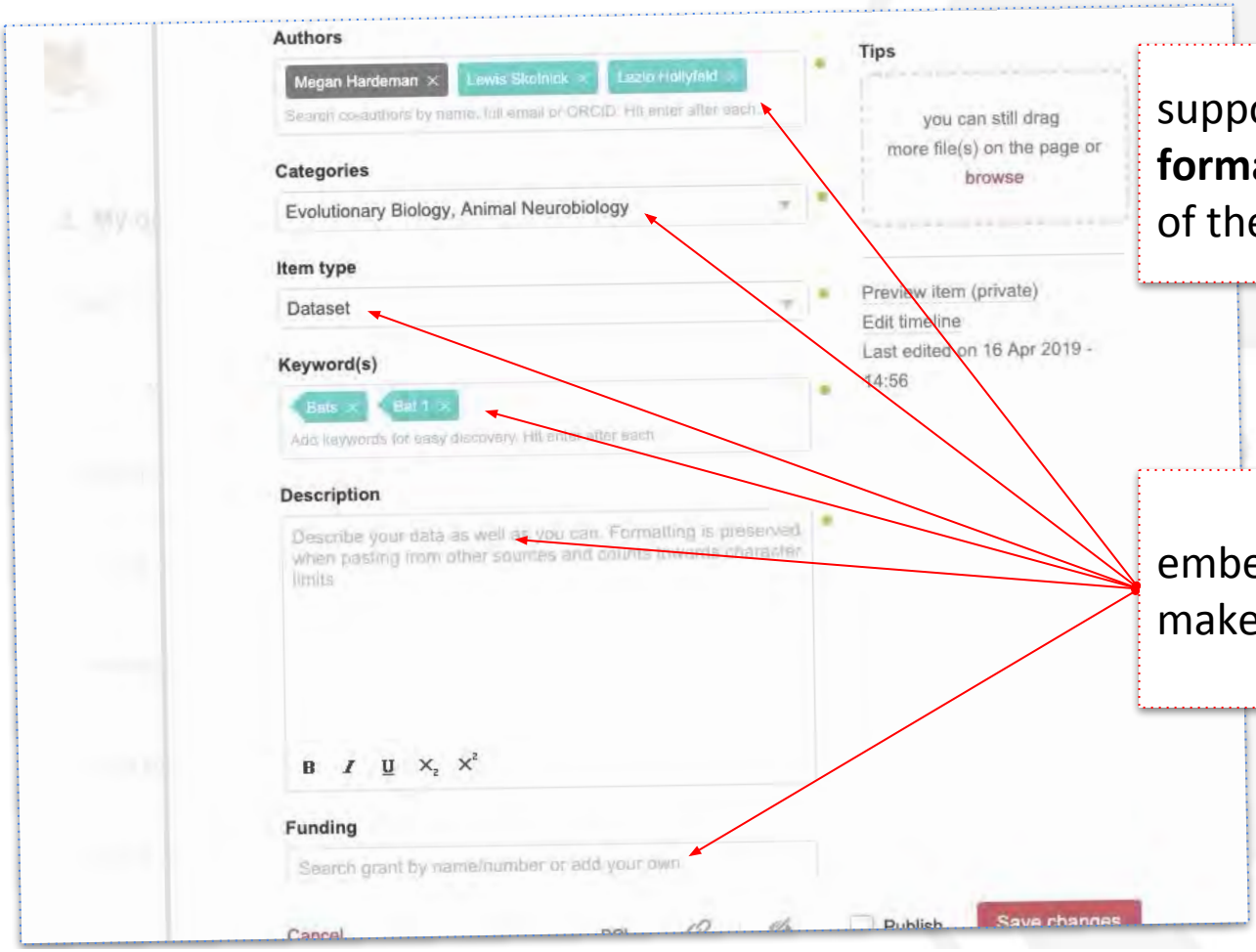
https://zivahub.uct.ac.za/



- **a repository to store and openly disseminate data**
- powered by ***Figshare*** *for institutions* (SaaS)
- keeps track of **views**, **downloads** and **citations**
- provides **universal search** & **linking** across *all* Figshare platforms in the world

ZivaHub | Open Data UCT

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



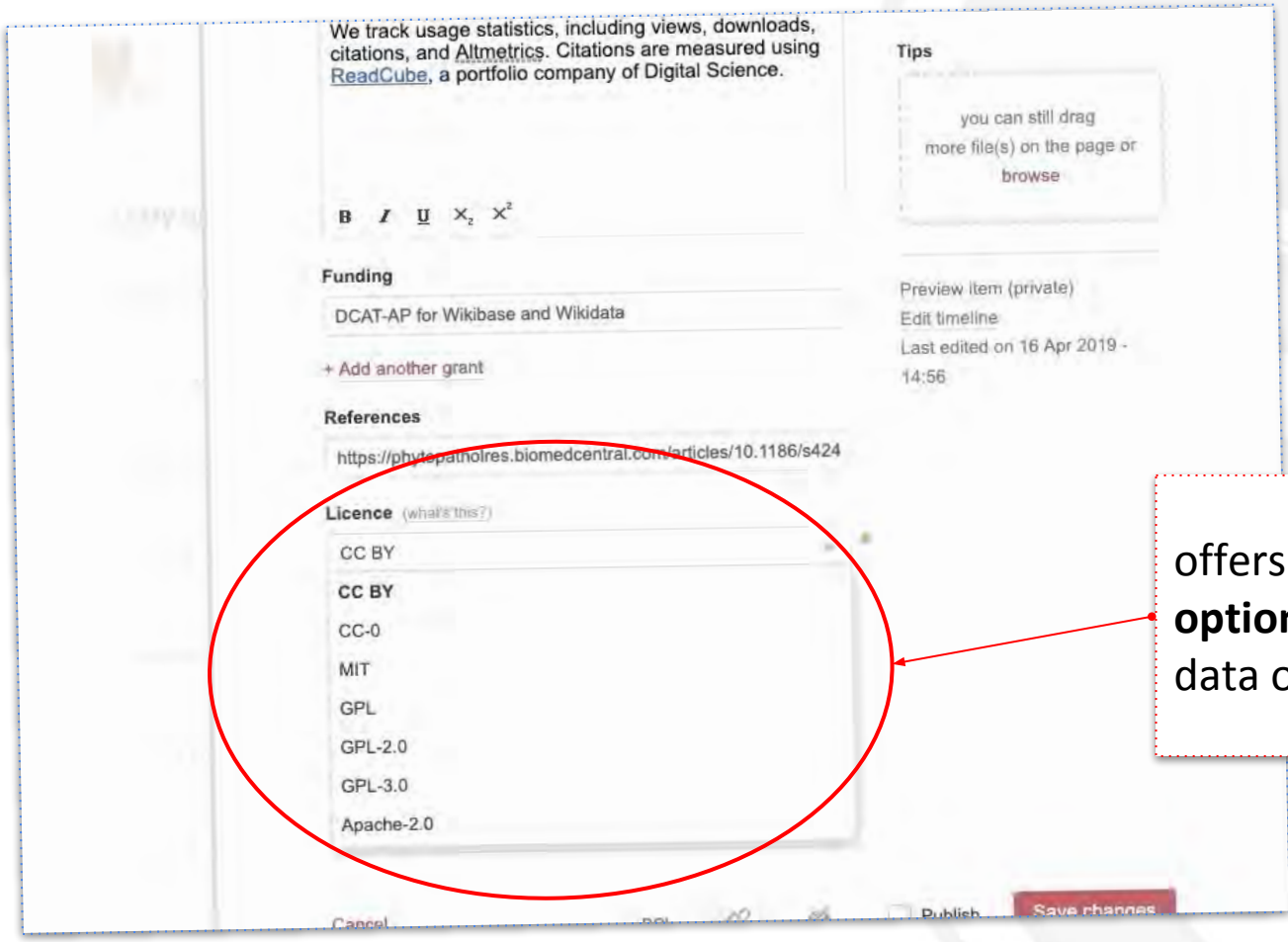
supports the upload of **any file format**, and aims to visualise all of them

embeds relevant **metadata**, to make data **FAIR** compliant

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

ZivaHub | Open Data UCT

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



offers a range of **licensing options** when publishing your data openly

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

Why do you need apply a license?

- Licensing is an important aspect of practising Open Science. By applying licenses to your outputs, you remove any ambiguity over what others can - and can't - do with your work.
- An open license, such as a Creative Commons license, consists of different elements that can be combined. Each element consists of a condition that needs to be followed by the re-user. The different combinations allow for great variation in the type of open license you apply: some being very open, others being very restrictive.

Authors: Foster. Available at:. https://datawizkb.leibniz-psychology.org/index.php/after-collection/what-should-i-know-about-licenses/

Sharing on ZivaHub

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

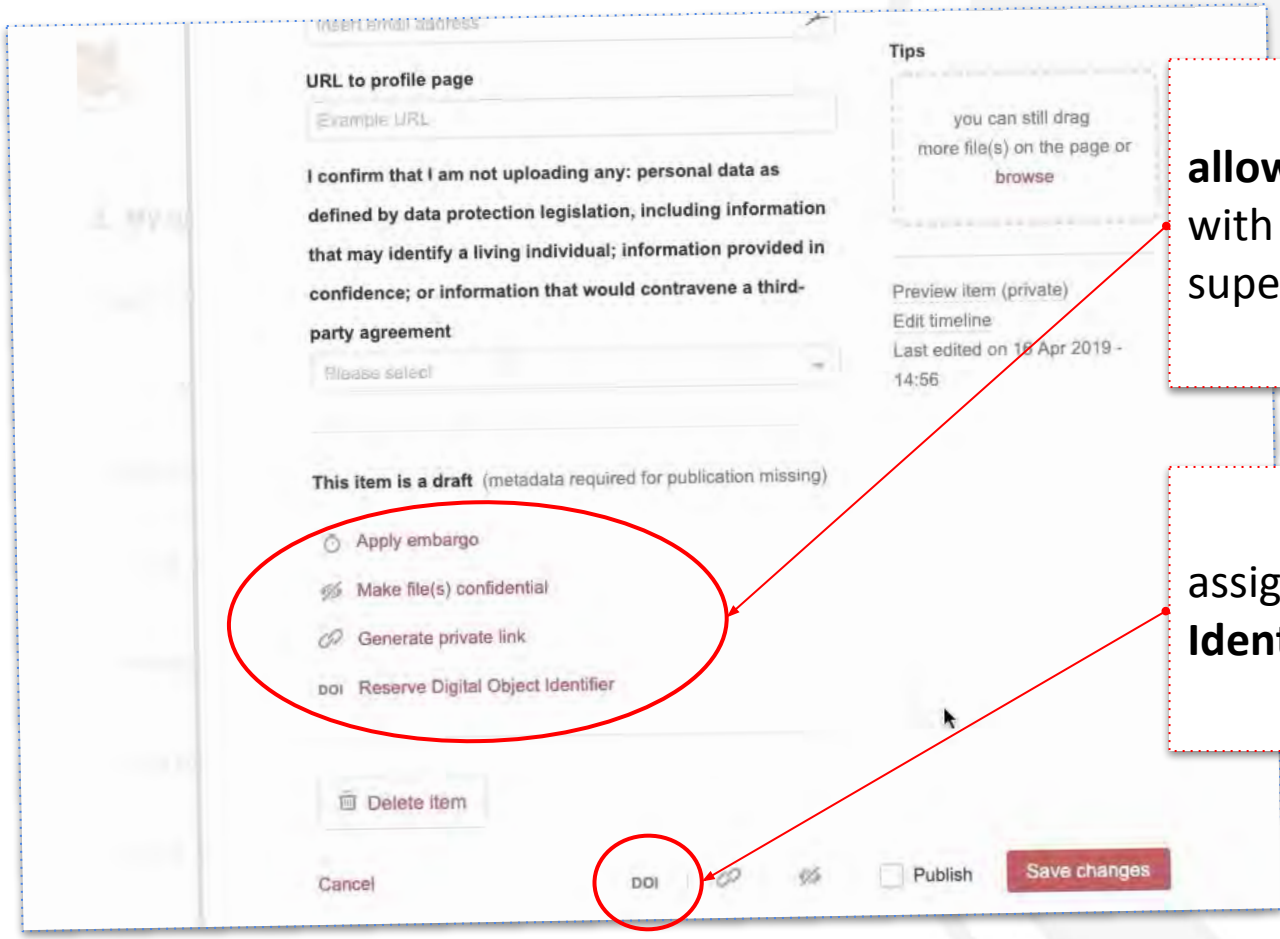
Private data	data uploaded into the repository, but nothing is shared or published (i.e. 'my data' storage space).
Metadata-only	metadata record links to where content is already stored, no data are uploaded.
Embargoed data	uploaded to the repository, metadata record is public, but the data are only available after specified date.
Private link	uploaded to the repository and shared via a private link only (useful for peer review of data, incl. 'blind').
Confidential	uploaded to the repository, metadata is publicly accessible, but data are inaccessible.

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

ZivaHub | Open Data UCT

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

DIGITAL LIBRARY SERVICES

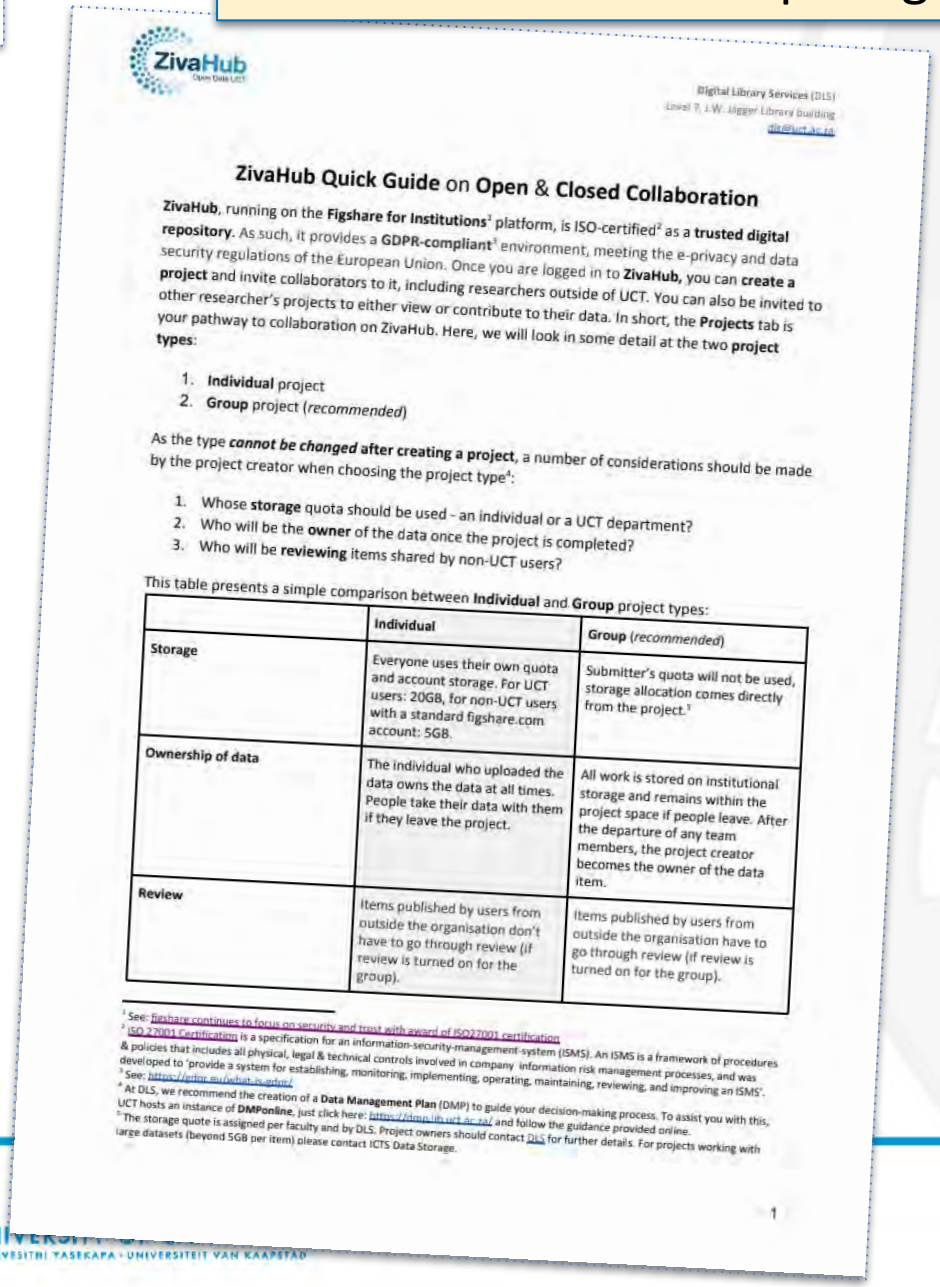


allows private data sharing, i.e. with funders, reviewers, or supervisors prior to publication

assigns a **Digital Object Identifier** (DOI) to all items

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

ZivaHub quick guides



Digital Library Services (DLS)
Level 7, J.W. Jagger Library Building
dls@uct.ac.za

ZivaHub Quick Guide on Open & Closed Collaboration

ZivaHub, running on the **Figshare for Institutions**[1] platform, is ISO-certified[2] as a **trusted digital repository**. As such, it provides a **GDPR-compliant**[3] environment, meeting the e-privacy and data security regulations of the European Union. Once you are logged in to **ZivaHub,** you can **create a project** and invite collaborators to it, including researchers outside of UCT. You can also be invited to other researcher's projects to either view or contribute to their data. In short, the **Projects** tab is your pathway to collaboration on ZivaHub. Here, we will look in some detail at the two **project types**:

1. **Individual** project
2. **Group** project (*recommended*)

As the type ***cannot be changed*** **after creating a project**, a number of considerations should be made by the project creator when choosing the project type[4]:

1. Whose **storage** quota should be used - an individual or a UCT department?
2. Who will be the **owner** of the data once the project is completed?
3. Who will be **reviewing** items shared by non-UCT users?

This table presents a simple comparison between **Individual** and **Group** project types:

	Individual	**Group** (*recommended*)
Storage	Everyone uses their own quota and account storage. For UCT users: 20GB, for non-UCT users with a standard figshare.com account: 5GB.	Submitter's quota will not be used, storage allocation comes directly from the project.[5]
Ownership of data	The individual who uploaded the data owns the data at all times. People take their data with them if they leave the project.	All work is stored on institutional storage and remains within the project space if people leave. After the departure of any team members, the project creator becomes the owner of the data item.
Review	Items published by users from outside the organisation don't have to go through review (if review is turned on for the group).	Items published by users from outside the organisation have to go through review (if review is turned on for the group).

[1] See: figshare continues to focus on security and trust with award of ISO27001 certification
[2] ISO 27001 Certification is a specification for an information-security-management-system (ISMS). An ISMS is a framework of procedures & policies that includes all physical, legal & technical controls involved in company information risk management processes, and was developed to 'provide a system for establishing, monitoring, implementing, operating, maintaining, reviewing, and improving an ISMS'.
[3] See: https://gdpr.eu/what-is-gdpr/
[4] At DLS, we recommend the creation of a **Data Management Plan** (DMP) to guide your decision-making process. To assist you with this, UCT hosts an instance of **DMPonline**, just click here: https://dmp.lib.uct.ac.za/ and follow the guidance provided online.
[5] The storage quote is assigned per faculty and by DLS. Project owners should contact DLS for further details. For projects working with large datasets (beyond 5GB per item) please contact ICTS Data Storage.

1

Guidance for publishing data on **ZivaHub:**

- Open and Closed collaboration
- Creating Your User Profile on ZivaHub
- Making your data FAIR with ZivaHub
- Your data availability statement
- Getting published on ZivaHub
- Publishing videos on ZivaHub
- ZivaHub - GitHub integration
- The ZivaHub Start-up guide

Further info:

- http://www.digitalservices.lib.uct.ac.za/dls/data-sharing-guidelines

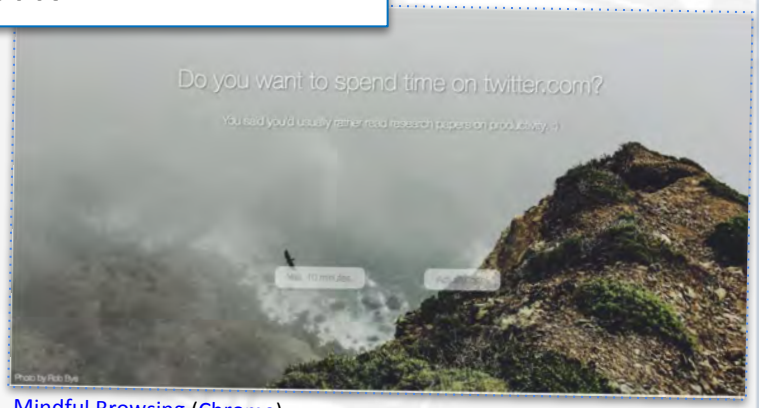
DIGITAL LIBRARY SERVICES

Digital Scholarship ‘hacks’

Focus and Deep Work

Dedicate your time and eliminate distractions

- Self-driven:
  - *'Pomodoro'* technique
    - Decide on the task to be done.
    - Set the pomodoro timer (traditionally to 25 minutes).
    - Work on the task.
    - End work when the timer rings and put a checkmark on a piece of paper.
    - If you have fewer than four checkmarks, take a short break (3–5 minutes), then go to step 2.
    - After four pomodoros, take a longer break (15–30 minutes), reset your checkmark count to zero, then go to step 1.
  - Web-only Version: https://tomato-timer.com
  - List of similar tools: https://techwiser.com/pomodoro-timer-for-windows/
- Computer-assisted:
  - Windows - Focus Assist
  - Mac - Do not Disturb in Notification Centre
  - Apps that can block access to sites/windows and keep track of your 'productivity': https://zapier.com/blog/stay-focused-avoid-distractions/



Mindful Browsing (Chrome)

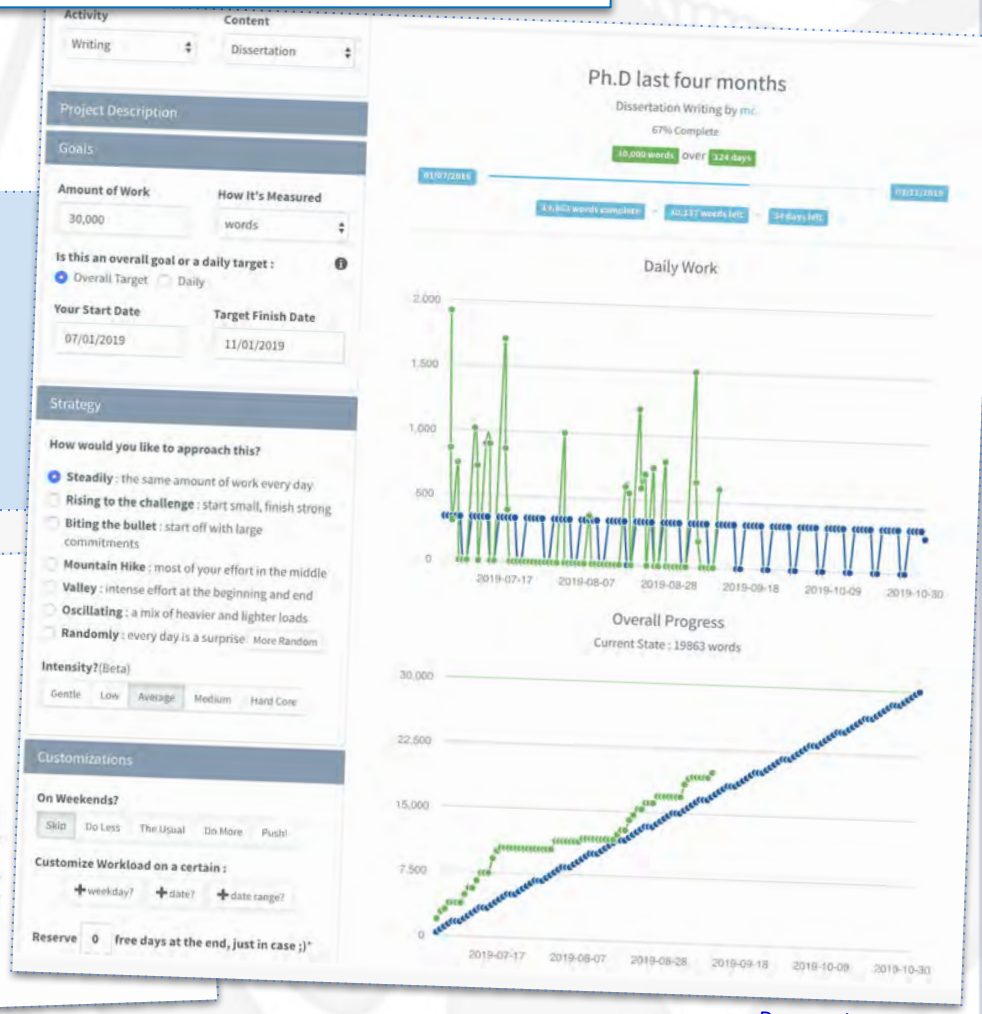
Tomato One

Adding to the word count

Pacemaker.Press

- Useful tool for setting word count milestones
- Can set flexible goalposts for your projects
- Can adjust based on your progress

Pacemaker.Press



Pacemaker.Press

DIGITAL LIBRARY SERVICES

Closing remarks, Upcoming workshops & Staying in touch with DLS

DIGITAL LIBRARY SERVICES

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 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

Digital Scholarship | RDM | Open Science

Digital Scholarship is the application and integration of digital tools and methods in learning, teaching and research. When you integrate digital technologies, work within networked environments and subscribe to Open Science practises, your Digital Scholarship has the power to transform the research landscape, and to serve the public good.

Research Data Management (**RDM**) is the organization and documentation of research data (ideally towards making it **FAIR**: **F**indable, **A**ccessible, **I**nteroperable and **R**eusable).

Open Science is a set of *practices* that drives all aspects of research to be more efficient, accountable, collaborative, and of good quality.

Digital Scholarship at DLS | **mission & vision**

We provide Digital Scholarship services to the University of Cape Town, including the following:

- Data Curation activities supporting best practices in **Research Data Management (RDM)**;
- specialist **Digitisation** services towards **Digital Preservation**;
- expertise in **Geographic Information Systems (GIS)**.

We advocate for **Open Science**, to make research done at UCT more *efficient, collaborative, accessible, findable and reusable*. We spearhead these practices as contributions to a more equitable and sustainable social order in the higher education landscape.

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

Examples of sustainably planned, strategic support

Proposal/Planning

- Data Management Planning (DMPonline, examples of existing DMPs)
- Advice on best practices for research data workflows

Process

- Advice on ways to better manage your data
- Where to find tools for collection, capture and analysis
- Advice on doing geospatial analysis

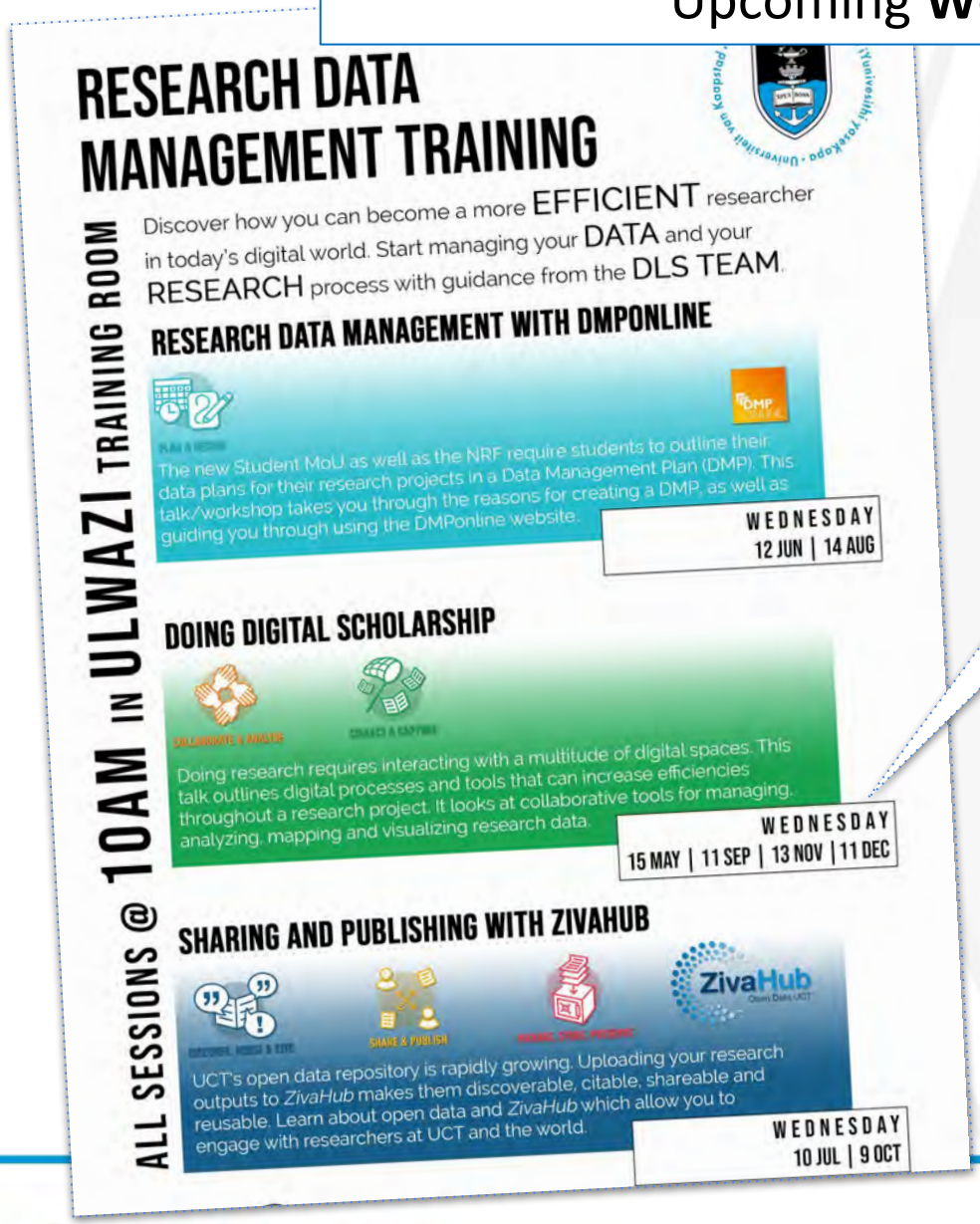
Publication

- Advice on what can be published (ethically; in terms of dataset size; ...)
- Advice on where to publish (i.e. a subject repository, ZivaHub; ...)
- Support on curating and developing an online showcases of data

Preservation

- Transfer analogue objects to digital files
- Ensure that your files will be accessed in perpetuity

Upcoming **Workshops**



Digital Scholarship: Tools for Analysis
13 Nov @ 10 AM

‘RDM at UCT’ Slack workspace

Slack = **S***earchable* **L***og of* **A***ll* **C***onversation and* **K***nowledge*

[RDM at UCT (Slack)]

[UCT DMP]

[OneDrive / Google Drive etc.]

[UCT Open Science Framework (OSF)]

[Digital preservation]

[ZivaHub | Open Data UCT]

Let's change what we value in research.

Sign DORA

DIGITAL LIBRARY SERVICES

Thank you!

dls@uct.ac.za

@DigitalUct

rdm-at-uct.slack.com

DIGITAL LIBRARY SERVICES http://www.digitalservices.lib.uct.ac.za/

Want to leave some feedback?