

Building Research Data Services at SA University Libraries

An offering of resources towards our communities of practice

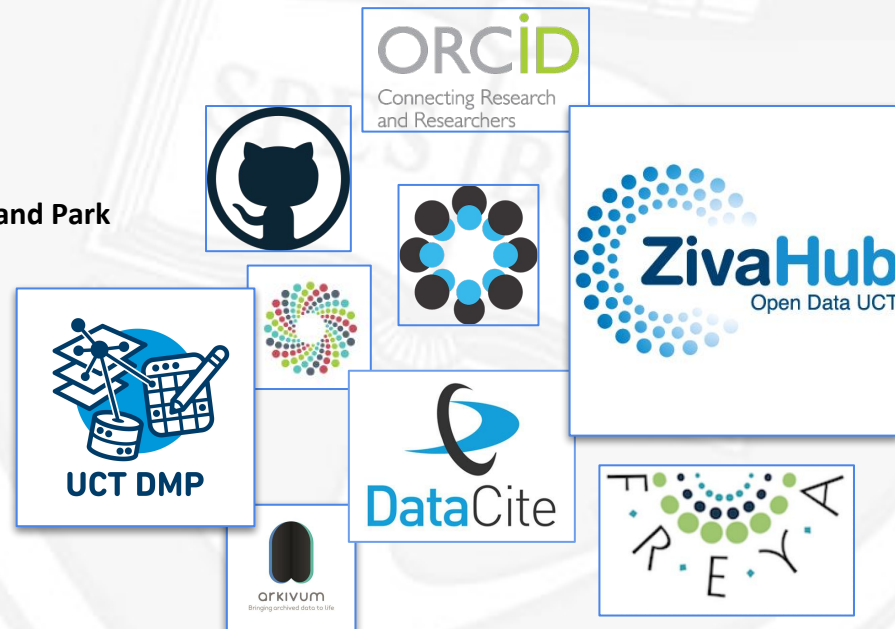
Tuesday 15th October 2019, 12:00 - 14:00
Wednesday 16th October 2019, 09:00 - 12:00

Training Room, Kingsway Campus Library, UJ Main Campus Auckland Park

UCT Libraries

Digital Library Services

- [Niklas Zimmer](#) (manager)
- [Sanjiin Muftić](#) (digital scholarship specialist))
- [Patricia Chikuni](#) (data curation officer)
- [Ya'qub Ebrahim](#) (data curation officer)
- [Thomas Slingsby](#) (GIS officer)



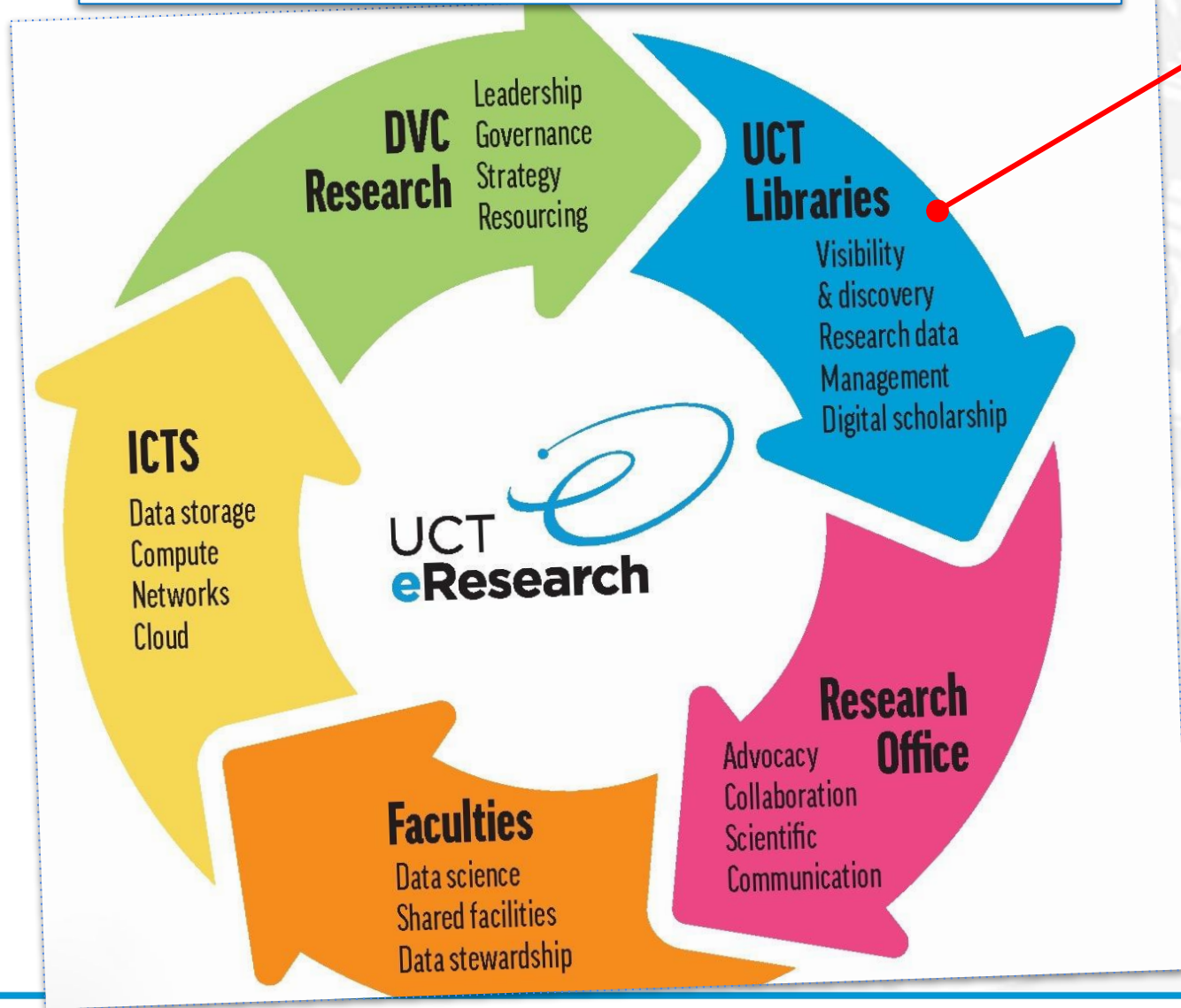


1. Digital Scholarship, Open Science and Research Data Management
2. The RDM Lifecycle: Aims & Principles
 - Support at each stage (Plan, Process, Publish)
3. The RDM Lifecycle: Stages, Methods & Tools
 - Plan & Design: UCT DMP platform
 - Collect & Capture: RedCap
 - Collaborate & Analyse: UCT Open Science Framework (OSF)
 - Discover, Reuse & Cite: selected repositories/systems (re3data)
 - Share & Publish: ZivaHub: Open Data UCT | Omeka S & iiif server
 - Manage, Store, Preserve: Arkivum Perpetua (AtoM / Archivematica)
4. Drivers for Digital Preservation (*Business Owners; Value Propositions; Policies & Frameworks*)
5. Practicalities of Digital Preservation (*Best Practices; Tools & Systems; Vendors; Roles; Activities*)
6. Digital Preservation: an all-inclusive conversation
7. Closing remarks, Upcoming workshops & Staying in touch with DLS

Who is talking?



DIGITAL LIBRARY
SERVICES



UCTL Digital Library Services staff



DIGITAL LIBRARY
SERVICES

Staff



Manager
Niklas Zimmer
niklas.zimmer@uct.ac.za
+27 21 650 2957



Data Curation Officer
Dr Patricia Rudo Chikuni
patricia.chikuni@uct.ac.za
+27 21 650 7073



Data Curation Officer
Ya'qub Ebrahim
yaqub.ebrahim@uct.ac.za
+27 21 650 2074



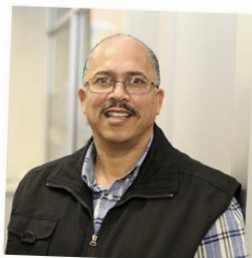
Digital Scholarship Specialist
Sanjin Muftić
sanjin.muftic@uct.ac.za
+27 21 650 5537



GIS Officer
Thomas Slingsby
thomas.slingsby@uct.ac.za
+27 21 650 3392



SLA: Digital Preservation
Mark Hanmer
mark.hanmer@uct.ac.za
+27 21 650 1253



SLA: Digital Preservation
Gary November
gary.november@uct.ac.za
+27 21 650 1275



GIS Specialist
Nicholas Lindenberg
nicholas.lindenberg@uct.ac.za
+27 21 650 3917

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



Contact

Search this website

HOME

ABOUT

ENGAGEMENT

AMBASSADORS

RESOURCES

NEWS & BLOG

FREYA is a 3-year project funded by the European Commission under the Horizon 2020 programme. The project aims to build the infrastructure for persistent identifiers as a core component of open science, in the EU and globally. FREYA will improve discovery, navigation, retrieval, and access of research resources. **New provenance services will enable researchers to better evaluate data and make the scientific record more complete, reliable, and traceable.** By engaging with the global community through the Research Data Alliance and other research infrastructures, **we work together to realise the vision of fully and effectively accessible data.**



Dr. Swijghuisen Reigersberg (<http://orcid.org/0000-0003-2337-7962>, @MurielSR) is an applied ethnomusicologist and practice-researcher affiliated with PARADISEC, Conservatorium of Music, and a researcher development manager (strategy) in the DVC Research Portfolio at The University of Sydney. Her research centres on Australian Indigenous Christian choral singing and the relationship between music, health and wellbeing. In her own time she works as a researcher and consultant.

Niklas Zimmer



Niklas Zimmer is the Manager of Digital Library Services at University of Cape Town (UCT) Libraries. He holds a MA(FA) and a BA(Hons) from UCT, as well as a BA in education from University of Cologne. Before his employment at UCT Libraries in 2015, Niklas worked as digitisation manager at the Centre for Popular Memory at UCT, as an art teacher heading the Visual Arts Department at the German International School in Cape Town, and as an archivist at the Art and Exhibition Hall of the Federal Republic of Germany.

ORCID: 0000-0001-8078-0403 | Academia.edu: [NiklasZimmer](#) | Twitter: [NiklasZimmer](#)

Guo Xiaofeng



Ambassador Programme



FREYA | **Connected Open Identifiers for Discovery, Access and Use of Research Resources** | www.project-freya.eu | twitter: @freya_eu



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.

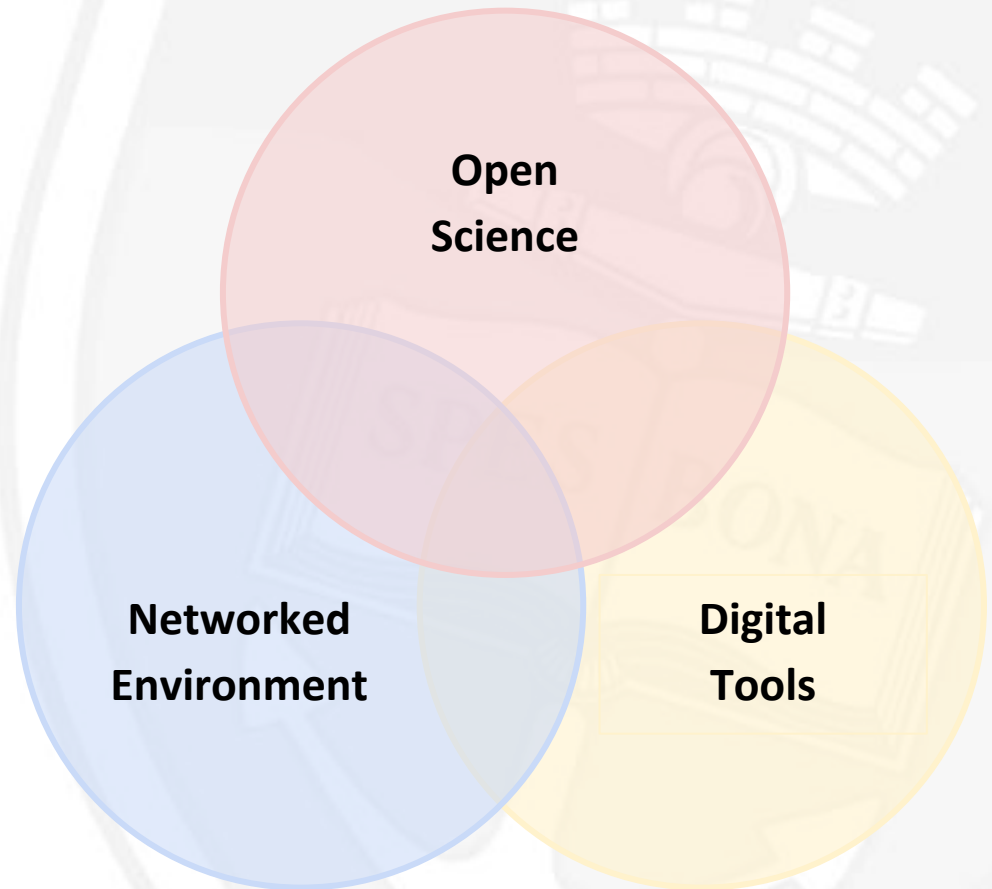
Source: DLS website: <http://www.digitalservices.lib.uct.ac.za/>



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

What is **Digital Scholarship**?

Digital Scholarship *is the application and integration of digital tools and methods in 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§ion=2>

“What (are my) research data?”

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

QUALITATIVE | QUANTITATIVE

Micro
Unit record
Raw
Field
Experimental
Spatial
Cleaned
Processed
Primary
Secondary
De-identified

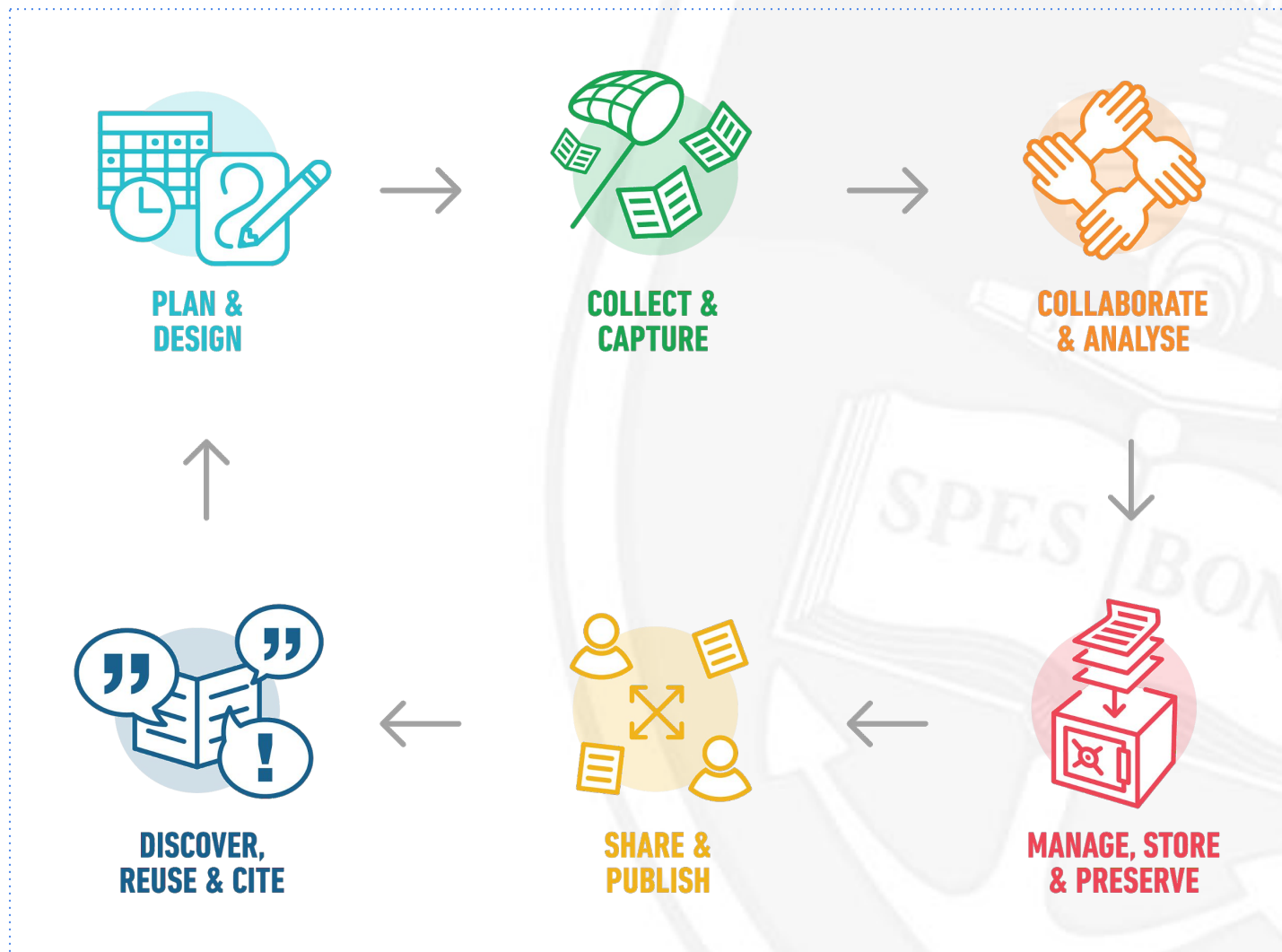
**RESEARCH
DATA**

Documents (text, spreadsheets)
Lab notebooks, field notebooks, diaries
Questionnaires, transcripts, surveys
Codebooks
Films, audio or video tapes/files
Photographs, image files
Sensor readings
Test responses
Artifacts, specimens, physical samples
Models, algorithms, scripts
Content analysis
Focus group recordings; interview notes

OBSERVATIONAL | EXPERIMENTAL | SIMULATION | DERIVED

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

The research project (& its data) lifecycle





Digital Scholarship and Research Data Management *towards Open Science*



MANAGEMENT - Data about Data

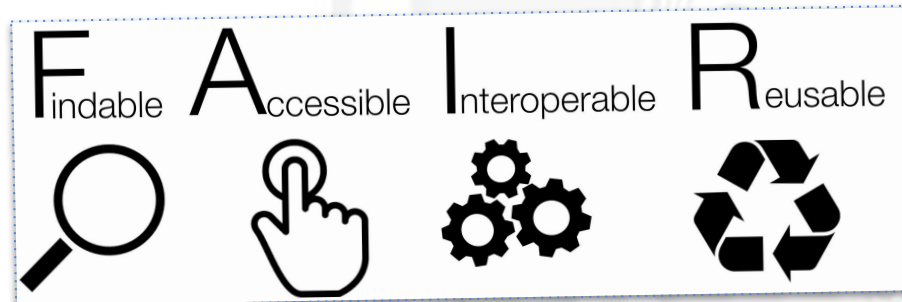
- **Digital Scholarship** tools exist to help us work with our data - keep track, store, secure, cleanup, analyze, collect, collaborate, visualize and organize
- Part of management is having sufficient and clear data describing the data and what has happened to it - or ***metadata***
- In the digital 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.



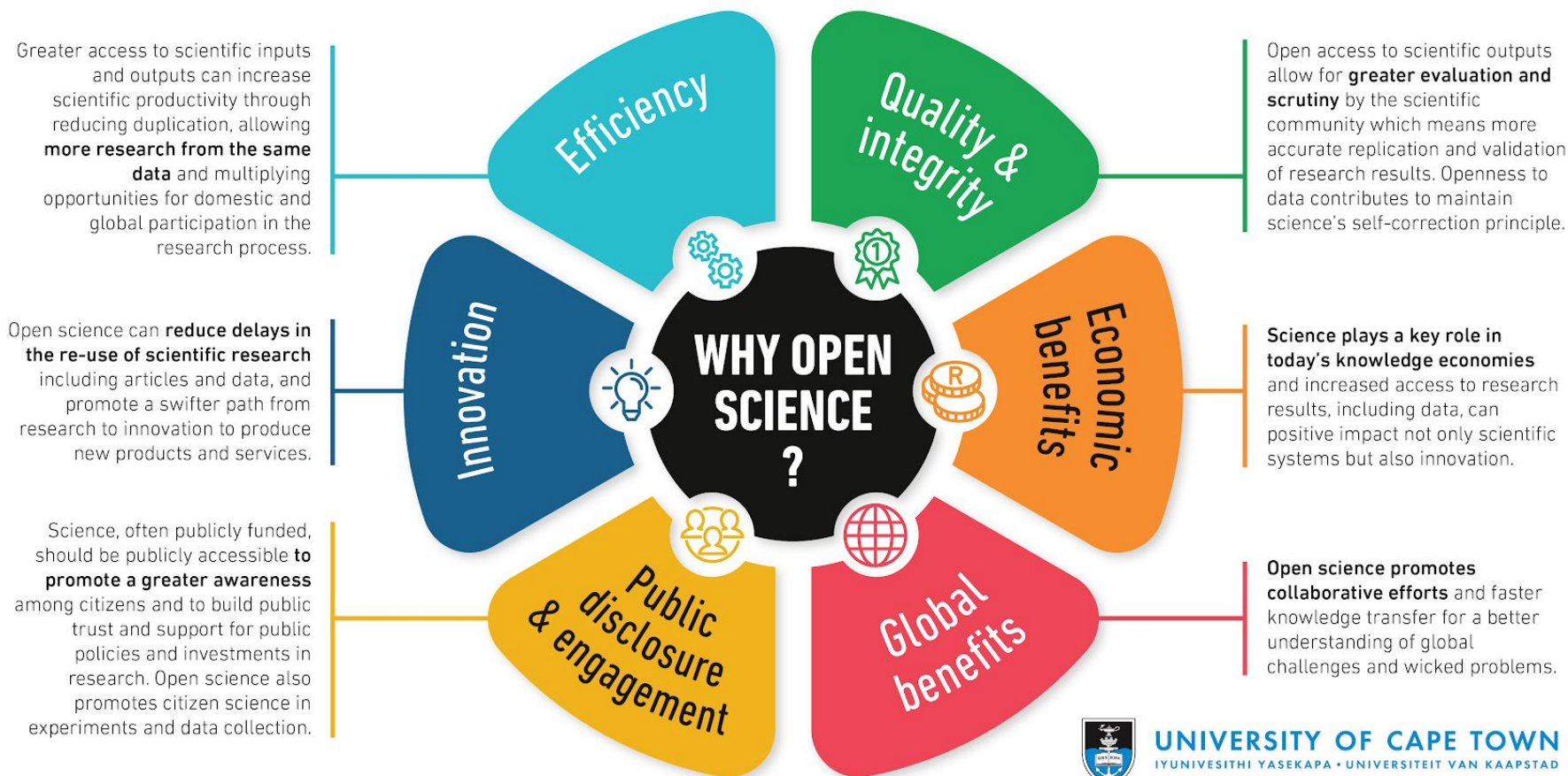
- Good **Digital Scholarship** practices along every step of the research lifecycle help data management and enable Open Science.

What is **Open Science**?



Source: 'Prometheus' <https://www.xkcd.com/1228/>

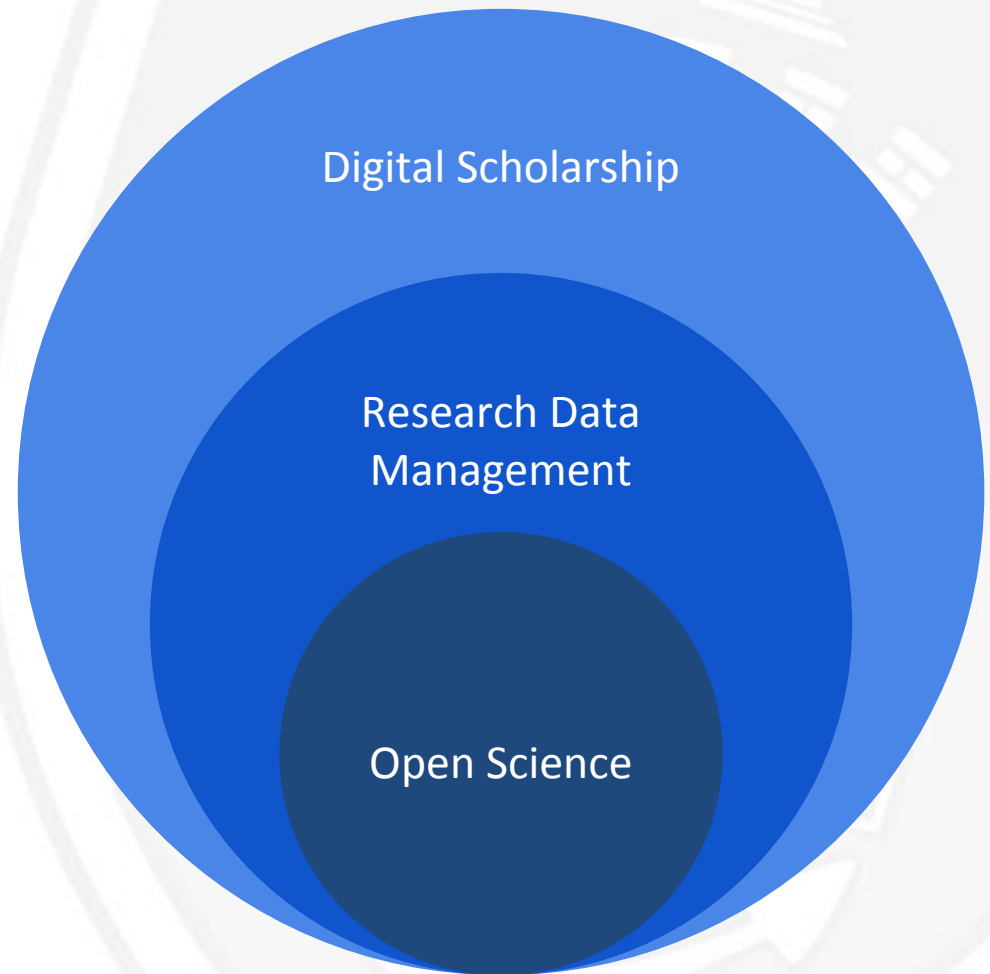
Open Science at UCT



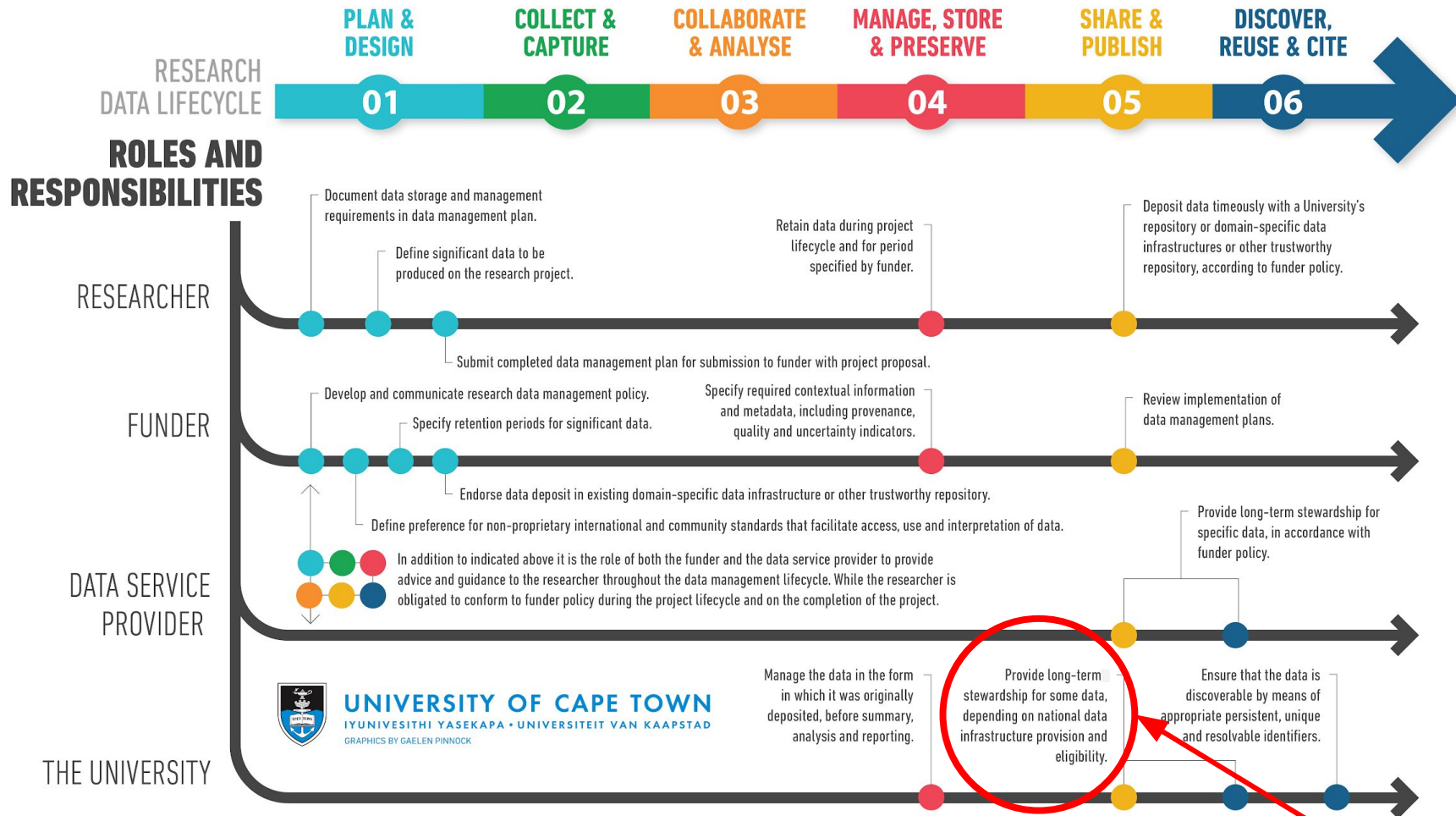
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD
GRAPHICS BY GAELIN PINNOCK

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

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



RDM roles & responsibilities at UCT



Source: UCT RDM Roles and Responsibilities: https://upload.wikimedia.org/wikipedia/commons/0/04/UCT_RDM_Roles-and-Responsibilities.png

'Good RDM makes data reusable'



DIGITAL LIBRARY
SERVICES



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



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



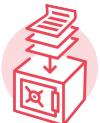
COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Open science workflows & tools

You can make your workflow more open by ...



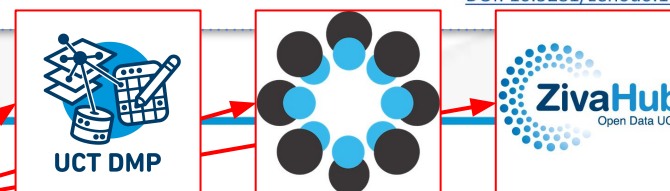
- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing 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](https://doi.org/10.5281/zenodo.1147025)

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





Get Connected

Identifying yourself to the machines of the internet



Identify yourself :)

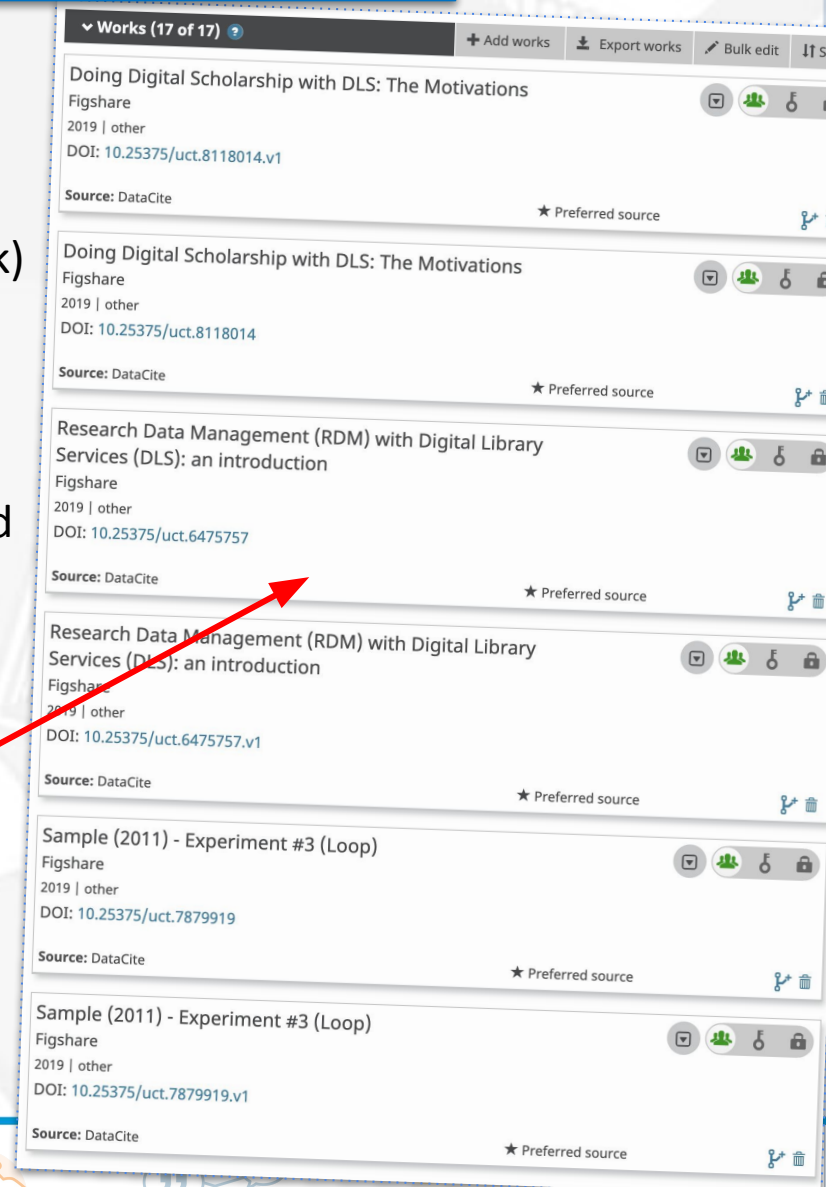
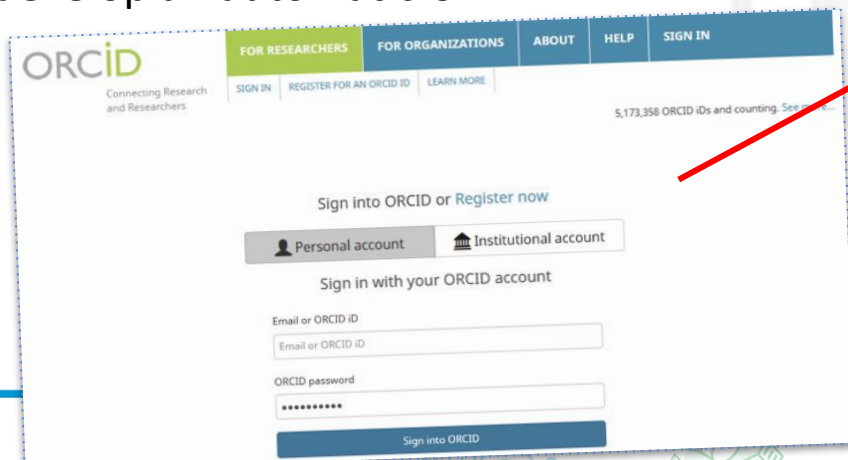
Get your own unique ID:

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

The ORCID login page shows the following elements:

- ORCID logo and tagline: 'Connecting Research and Researchers'
- Navigation tabs: FOR RESEARCHERS, FOR ORGANIZATIONS, ABOUT, HELP, SIGN IN
- Links: SIGN IN, REGISTER FOR AN ORCID ID, LEARN MORE
- Text: 5,173,358 ORCID iDs and counting. See more...
- Sign in options: Personal account, Institutional account
- Text: Sign in with your ORCID account
- Input fields: Email or ORCID ID, ORCID password
- Button: Sign into ORCID

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:](#)



Why should you get connected?

Building a Culture of Data Citation



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



The Research Data Lifecycle

*Digital Scholarship tools and methods to assist
with Research Data*





PLAN & DESIGN



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Plan & Design

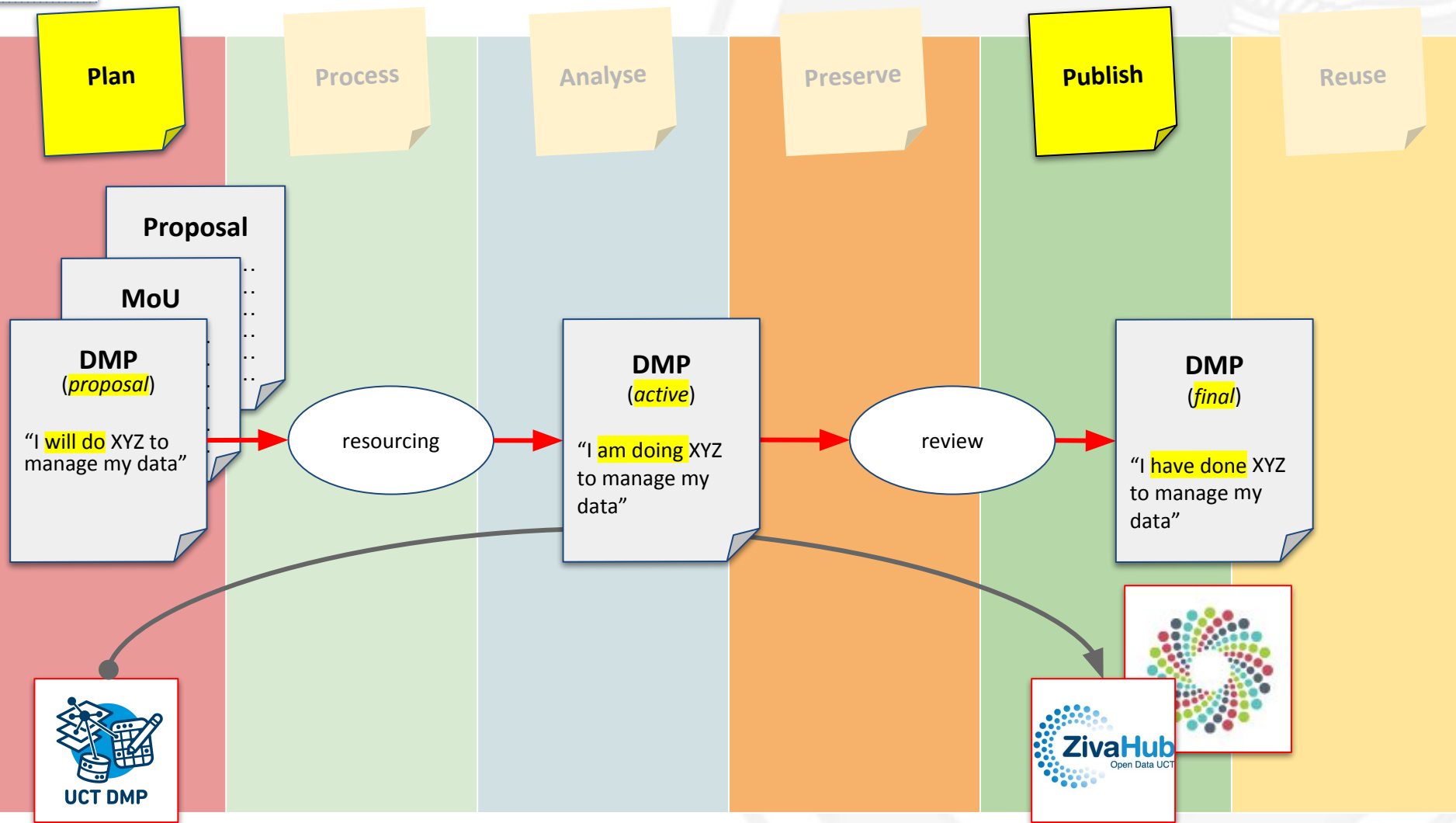
Planning for your various needs upfront, you are able to resource them appropriately. Collaborate with all your stakeholders at the budget application stages in order to strategically approach all the phases of digital preservation.

- ★ design digitisation & preservation projects
- ★ plan consent for publishing (IP / copyright, release forms, etc.)
- ★ locate & (re-)process existing data
- ★ plan data management (formats, storage, etc.)
- ★ digitisation to intl. standards and best practice
- ★ capturing of technical metadata
- ★ creation of descriptive metadata

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 DMP is already a **requirement** by many **funders** (NIH, Wellcome Trust, NRF).
- Even when 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.
- A DMP 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	<input type="checkbox"/>
Student	Yes	<input type="checkbox"/>

10 January 2019
Page 6
ACA47a

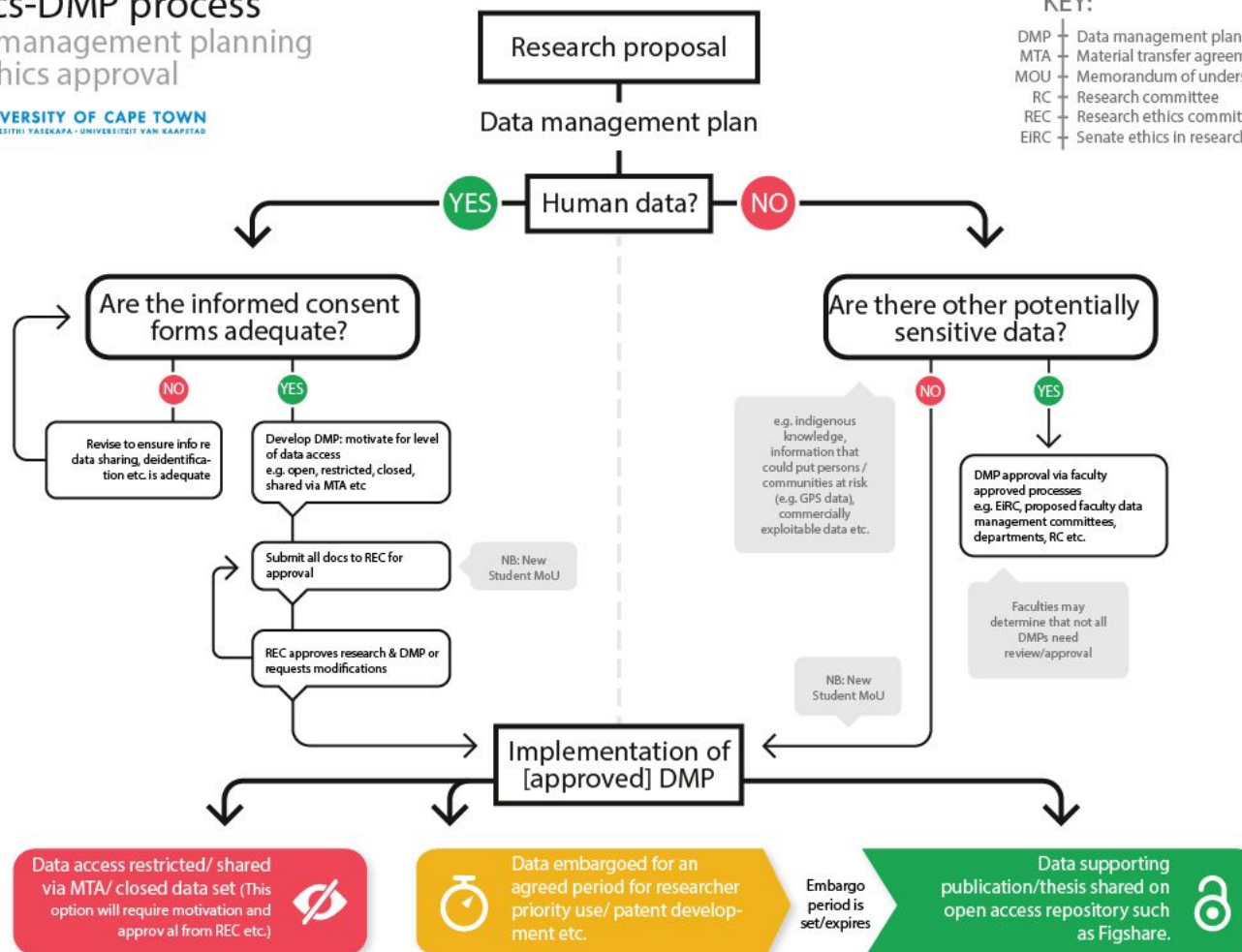
For more information, see: OSF Guides > Best Practices > Handling Data > **Creating a data management plan (DMP)**. Available: <http://help.osf.io/m/bestpractices/l/618674-creating-a-data-management-plan-dmp>

Data Management Planning & Ethics at UCT

Ethics-DMP process
Data management planning
for ethics approval



KEY:
DMP – Data management plan
MTA – Material transfer agreement
MOU – Memorandum of understanding
RC – Research committee
REC – Research ethics committee
EIRC – Senate ethics in research committee



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



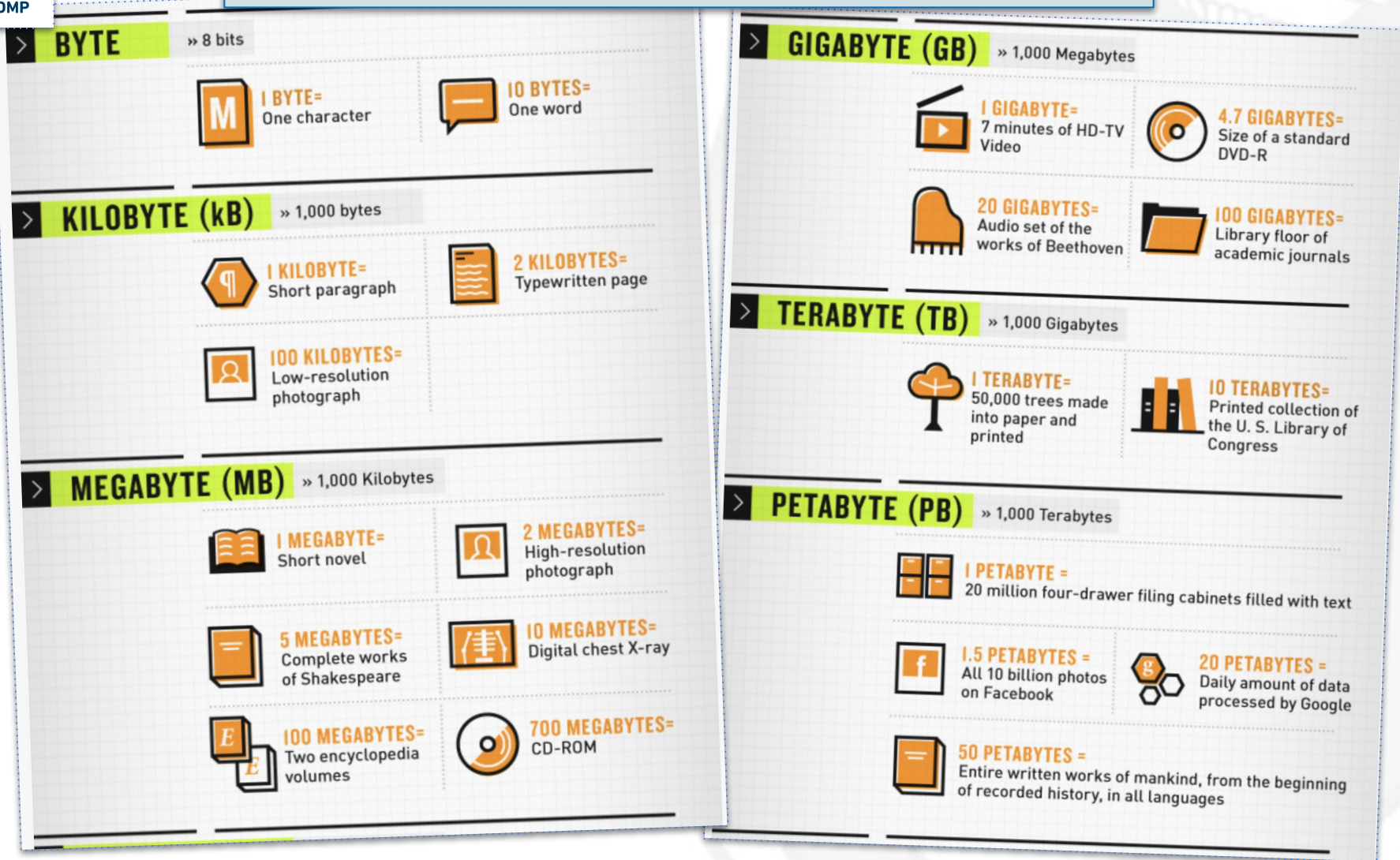
Typical DMP questions

- **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>



Example for tips / guidance: Data sizes



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

The UCT DMP platform

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

My plan (Gender, Health and Justice Research Unit)

Plan details

GHJRU 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-Generis)

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.

UCT DMP

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

Create your own account

My plan (Gender, Health and Justice Research Unit)

0/18 questions answered

approx. 12% of 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 icon] [table icon] [link icon] [grid icon]

Save

Not answered yet

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

B I [list icon] [table icon] [link icon] [grid icon]

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.

Useful information is provided at every step.

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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



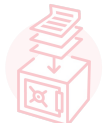
COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE

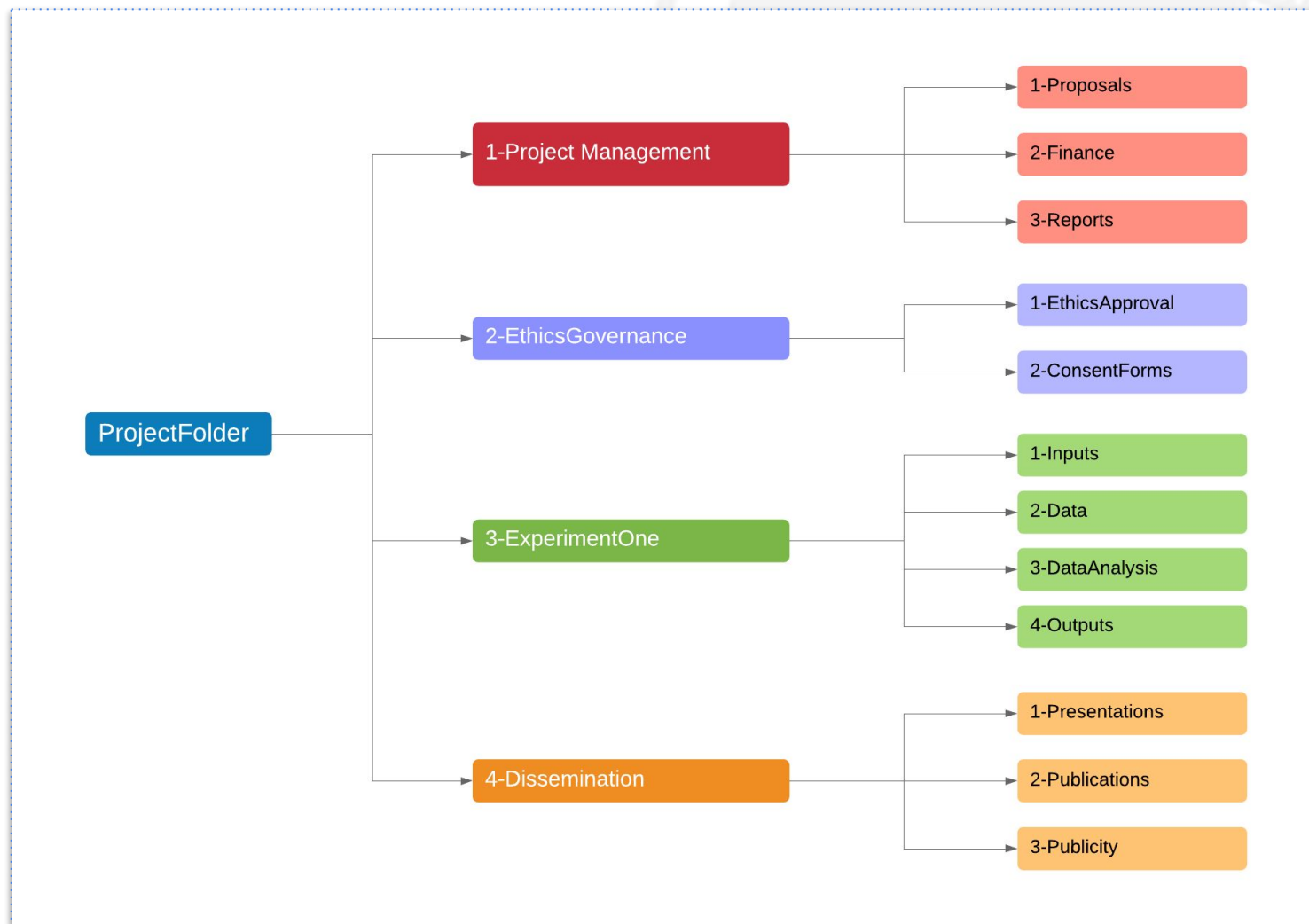


SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Organizing your folders



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



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE

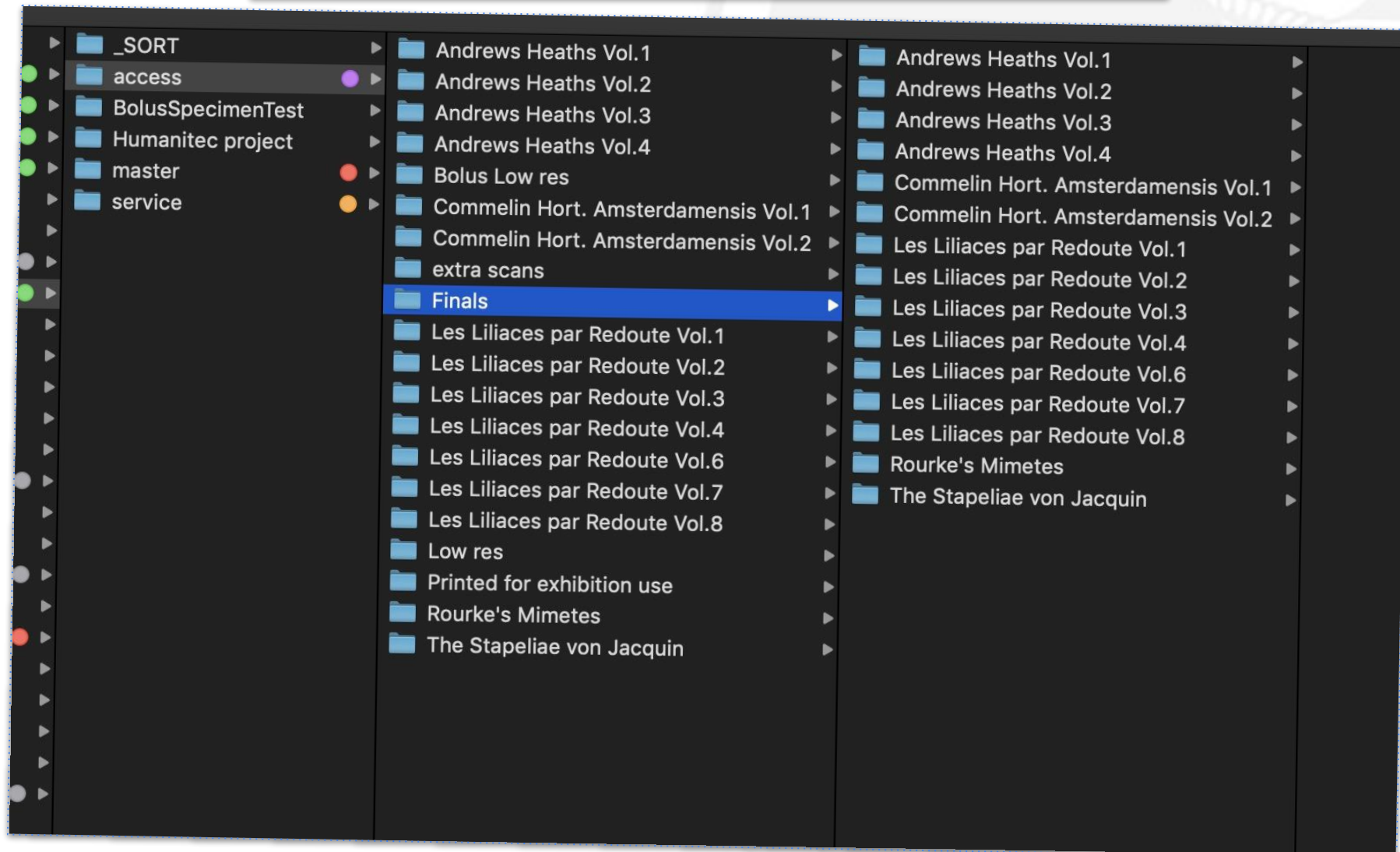


SHARE & PUBLISH



MANAGE, STORE, PRESERVE

'final' is just plain rude



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

<i>Files without a naming convention:</i>	<i>Files with a naming convention:</i>
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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE



COLLECT & CAPTURE



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Collect & Capture

Obtaining and/or creating data often requires the use of complex tools and software, which collect raw data that then need to be stored and managed. Ensuring that data are captured and stored effectively involves setting up relevant soft- and hardware, as well as NB: well-documented, cross-departmental workflows.

- ★ creating / logging data
- ★ checking, validating, and cleaning data
- ★ managing active data storage
- ★ enhancing metadata
- ★ transcription, translation
- ★ anonymise / de-identify data (where necessary)



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)



RedCap

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



REDCap
Logged in as 01401241 | Log out

My Projects
Project Home
Project Setup
Project status: Development

Data Collection [Edit instruments](#)

Manage Survey Participants
Get a public survey link or build a participant list for inviting respondents

Record Status Dashboard
View data collection status of all records

Add / Edit Records
Create new records or edit/view existing ones

Show data collection instruments ▼

Applications

- Calendar
- Data Exports, Reports, and Stats
- Data Import Tool
- Data Comparison Tool
- Logging
- Field Comment Log
- File Repository
- User Rights and DAGs
- Data Quality
- REDCap Mobile App
- External Modules

Reports [Edit reports](#)

1) PI weekly report

Help & Information

- Help & FAQ
- Video Tutorials
- Suggest a New Feature

[Contact REDCap administrator](#)

DMPOnline Survey

Project Home | Project Setup | Other Functionality | Project Revision History

Project status: Development Completed steps 1 of 5

Main project settings

Complete!

Disable ☒ Use surveys in this project? [?](#) [VIDEO: How to create and manage a survey](#)

Enable ☐ Use longitudinal data collection with defined events? [?](#)

Not complete? [Modify project title, purpose, etc.](#)

Design your data collection instruments & enable your surveys

Not started

[I'm done!](#)

Add or edit fields on your data collection instruments (survey and forms). This may be done by either using the Online Designer (online method) or by uploading a Data Dictionary (offline method). You may then enable your instruments to be used as surveys in the Online Designer. Quick links: [Download PDF of all instruments](#) OR [Download the current Data Dictionary](#)

Go to [Online Designer](#) or [Data Dictionary](#) Explore the [REDCap Shared Library](#)

Have you checked the [Check For Identifiers](#) page to ensure all identifier fields have been tagged?

Learn how to use [Smart Variables](#) [Piping](#) [Action Tags](#)

Enable optional modules and customizations

Optional

[I'm done!](#)

Enable ☐ Repeatable instruments [?](#)

Disable ☒ Auto-numbering for records [?](#)

Enable ☐ Scheduling module (longitudinal only) [?](#)

Enable ☐ Randomization module [?](#)

Enable ☐ Designate an email field for sending survey invitations [?](#)

[Additional customizations](#)

Set up project bookmarks (optional)

Optional

[I'm done!](#)

You may create custom bookmarks to webpages that exist inside or outside of REDCap. These bookmarks will be seen as links on the left-hand project menu and can be accessed at any time by users who are given privileges to do so. Every project bookmark has custom settings that allow one to control its appearance and behavior.

Go to [Add or edit bookmarks](#)

User Rights and Permissions

Optional

[I'm done!](#)

You may grant other users access to this project or edit the user privileges of current users of this project by navigating to the User Rights page. Additionally, if you wish to limit user access to certain records/responses for this project, you may want to use Data Access Groups, in which only users within a given Data Access Group can access records created by users within that group.

Go to [User Rights](#) or [Data Access Groups](#)

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

Adapted from: Harvard Catalyst

<https://catalyst.harvard.edu/services/redcap/>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH

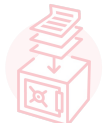


MANAGE, STORE, PRESERVE

Make peace with spreadsheets

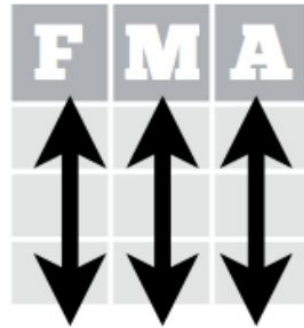
1. ORGANIZED DATA LIVES in TABLES
2. The idea is that in a spreadsheet
 - a. each new **observation gets a new row**
 - b. each **variable a new column.**
3. DO NOT USE
 - a. merged cells,
 - b. subheadings,
 - c. color used to denote information,
 - d. different data types within cells (numbers and 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/>



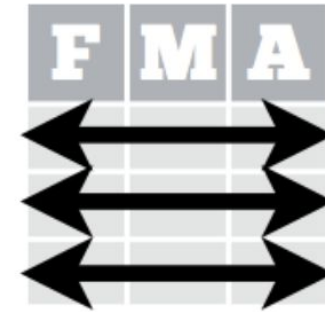
Make peace with spreadsheets

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_tidy/#35_other_tidy_awesome



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/1/618674-creating-a-data-management-plan-dmp>



Signs you might not be managing your data

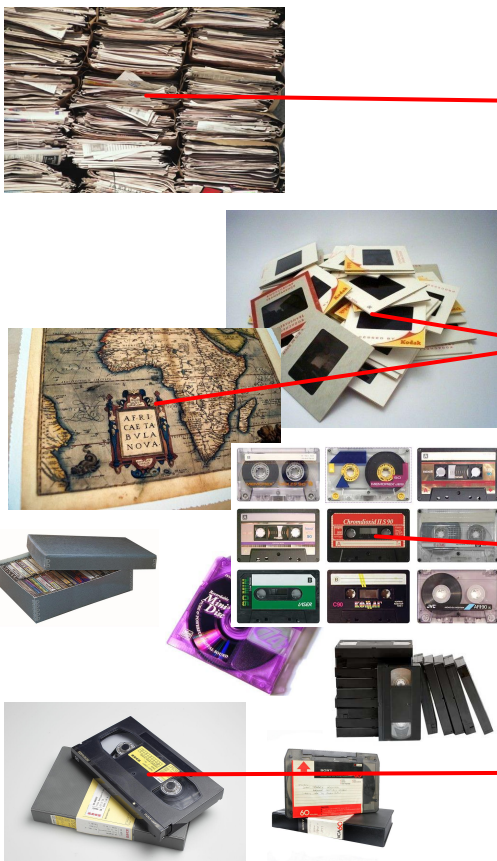
1. Your only copy of your data is on a flash that you left behind in a Postnet
2. You know you saved it somewhere but the search function is not finding it and you can't remember the file name
3. Your collaborator can't make sense of the contents of the files: ie What does the column *love1* stand for?
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*



Digitisation for Digital Preservation & Access

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

legacy formats



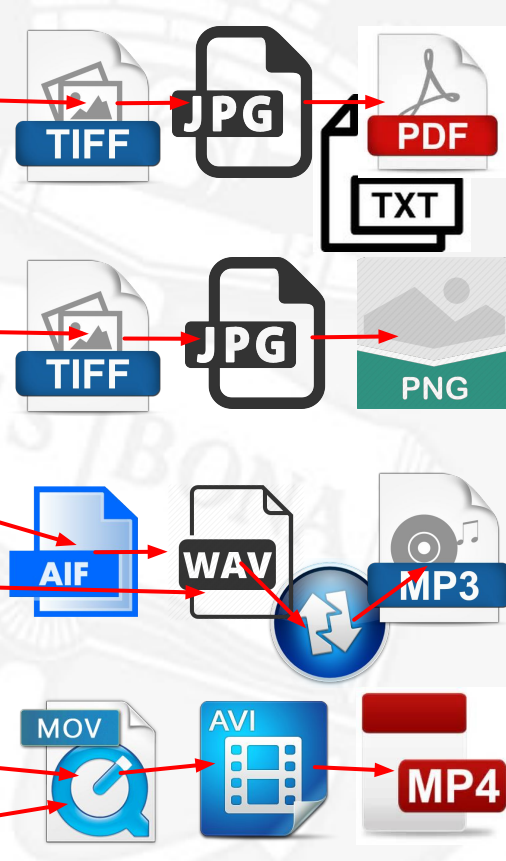
hardware



software



digital files



Some requirements, and an example ...

Requirements (overall):

- specialised equipment - (legacy and current)
- special skills - (scarce)

Example: video tapes:

- 1) **Legacy media** (analogue or digital)
e.g.: film collections (NB: 20+ formats at SC!)
- 2) **Documentation** regarding mandates, rights, description, QA, delivery
e.g.: Digital Preservation Policy; AV Digitisation guideline; service manuals
- 3) **Physical storage**, cleaning and preparation facilities, equipment, materials and skills
e.g.: cold room, tape cleaning/conditioning machines (per format), training
- 4) **Legacy playback equipment**; related knowledge and skills to operate & troubleshoot
e.g.: playback machines (per format), training, servicing/repair
- 5) Up to date **digitisation equipment** (DACs), capture software (DAW) and skills
e.g.: video capture cards, fast computer and HDD, video software, training
- 6) Large, fast **repository solutions** for storage, management, preservation and access
e.g.: Media Asset Management system, high-availability server; systems and processes for preservation (NB: size!) and online exhibition (NB: streaming server).



COLLABORATE & ANALYSE



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH

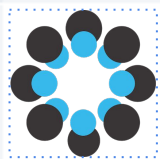


MANAGE, STORE, PRESERVE

Collaborate & Analyse

Archival projects today often include teams scattered geographically, who need access to the same data at the same time. Consider creating shared data stores, dataset transfers, file sharing and other facilities or software required for effective collaboration, including data processing (e.g. virtual machines, cloud-based software etc.).

- ★ checking, validating, and cleaning data
- ★ creating / enhancing metadata; transcription; translation
- ★ anonymising / de-identifying data
- ★ describing, managing and storing data
- ★ prepare data products for preservation and publication



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 Kekacs

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/drive/folders/1brpFv87IOftUye6zyad9jYSajocFca77usp=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

Practical 12 - Writing an abstract.gslides

Practical 2 - Creating online questionnaires.gslides

Practical 3 - Reading, writing, and citing research papers.gslides

Practical 4 - Ethical issues in conducting and publishing resear...

Practical 5 - Intervention studies and group design.gslides

Project evaluation rubric.gsheet

OSF Storage (Germany - Frankfurt)

Citation

Recent Activity

Tamas Nagy updated file

Nagy Tamas.docx in OSF Stor

107 - 2019 Spring

Tamas Nagy updated file

Nagy Tamas.docx in OSF Stor

107 - 2019 Spring

Tamas Nagy updated wiki

Psychology - PSYM17-CH-107

Tamas Nagy updated wiki

Psychology - PSYM17-CH-107

Zoltan Kekacs linked Dropbo

PSYM17-CH-107 - 2019 Spring

Zoltan Kekacs authorized the

- PSYM17-CH-107 - 2019 Spring

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

Sheet_1

Show rows with cells including:

Variable	Variable name	Mesaurement unit	Allowed values	Description
Participant ID number	ID	Numeric	001-999	ID number assigned to participant in sequ
Group number	GROUP	Numeric	1-30	Group assigned to participant based on IC
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 surv
Self-reported consumer spending	SPEND	Numeric	0-100,000,000	Self-reported average yearly expendit
Market sentiment	SENTIMENT	Numeric	1 = negative 2 = neutral 3 = positive	Sentiment towards US domestic econo
Actual GDP growth	GDP	Numeric	-5.0-5.0	Average US yearly GDP growth



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE



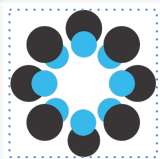
Open Science Framework (OSF)

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

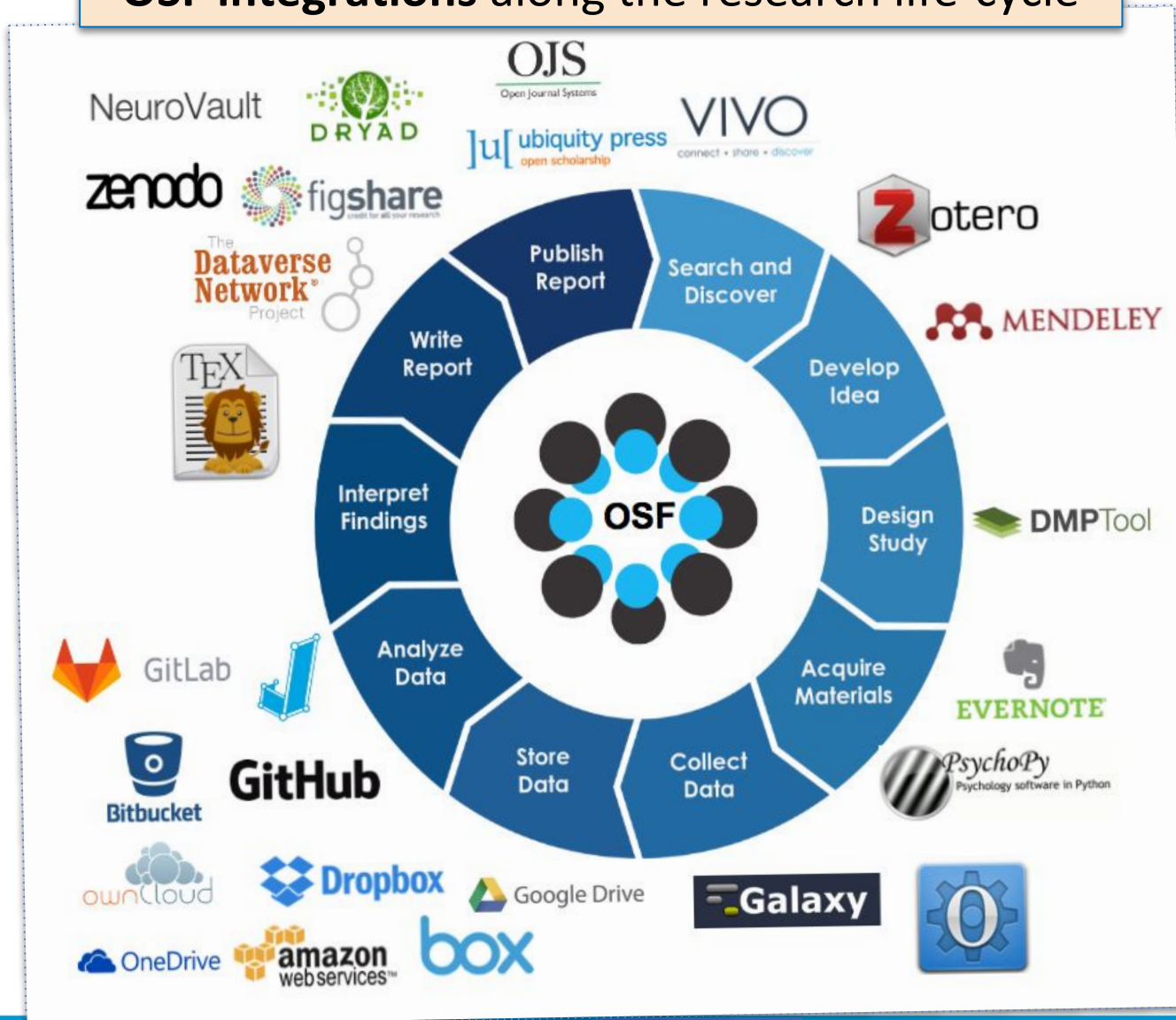
- The [UCT OSF](#) service is an online platform that allows you to register your project, manage stakeholders, and centralise data that might be stored at different locations with different collaborators.
- Create an account in order by clicking on 'Sign in' on the home page, and then select '[Login 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](#)





OSF integrations along the research life-cycle

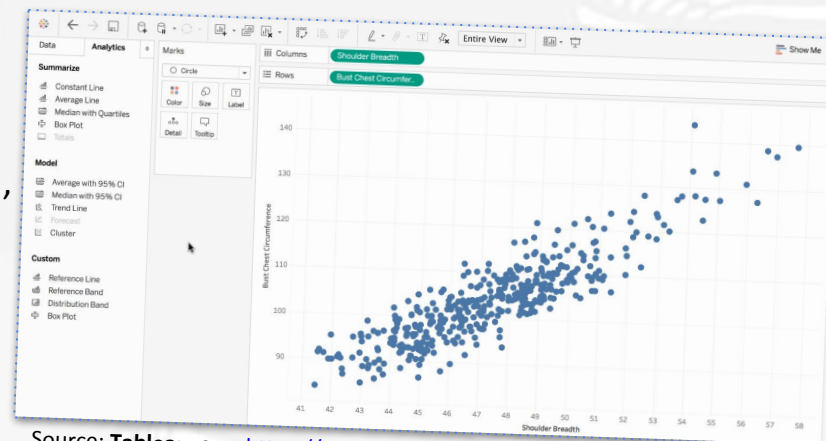


Advanced digital scholarship

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.



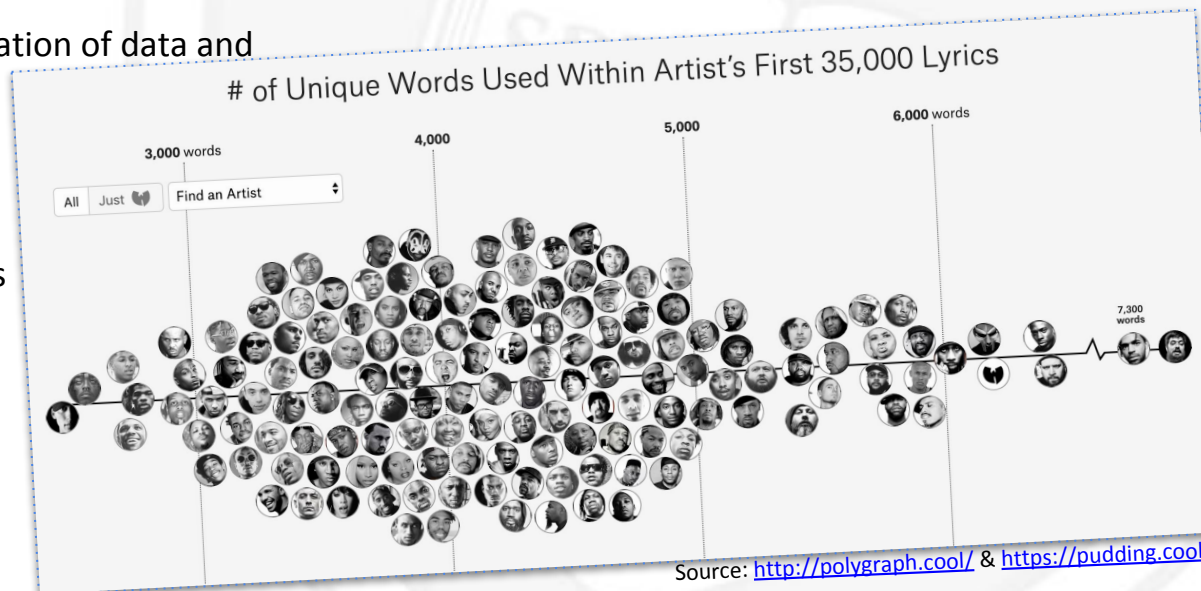
Source: Tableau, see: <https://www.tableau.com/>

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/>

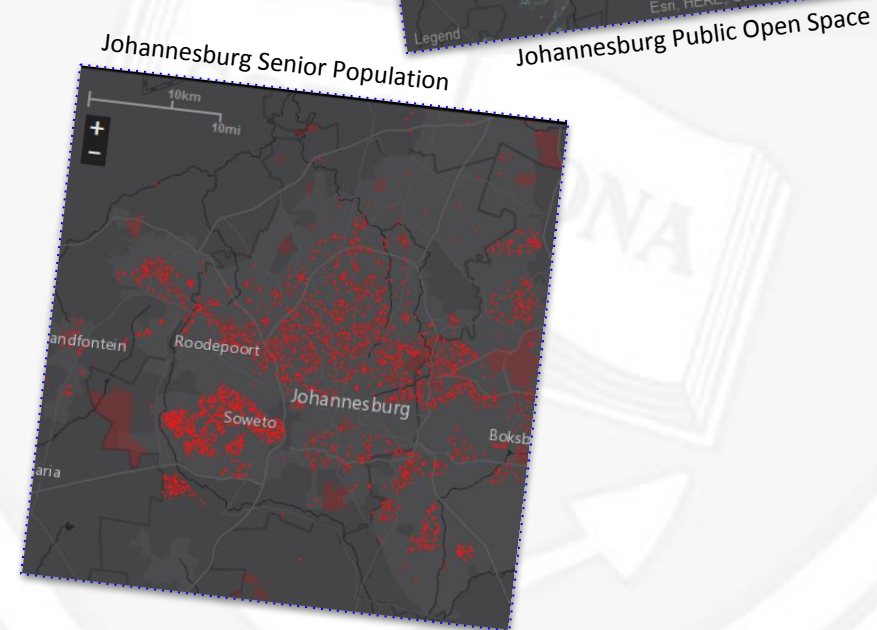
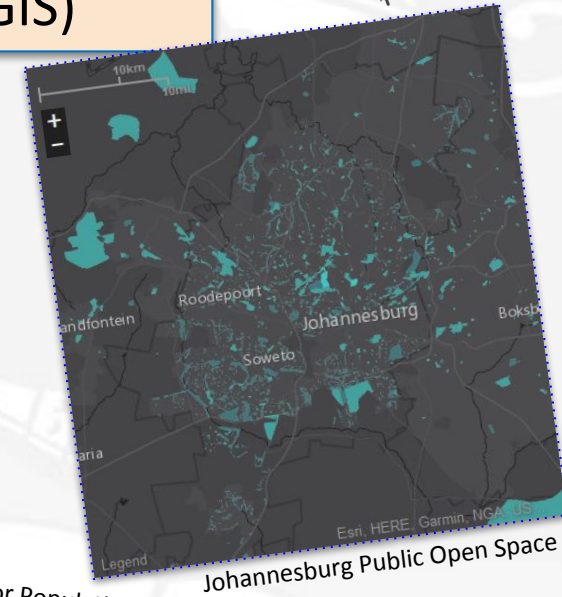
Geographic Information Systems (GIS)

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.

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



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



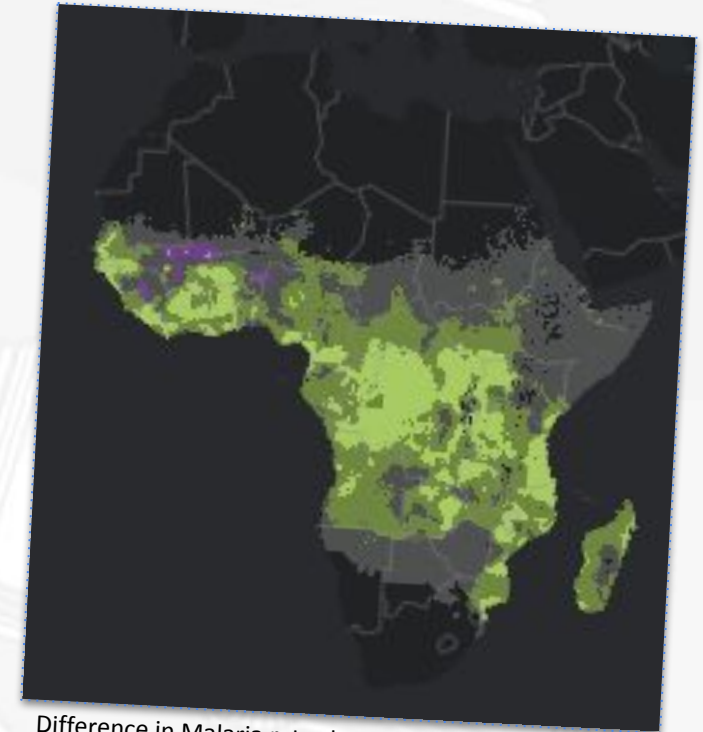
MANAGE, STORE, PRESERVE

Geographic Information Systems (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

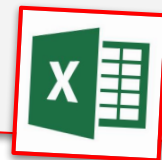
DLS’ GIS services assist with GIS software acquisition, project planning, troubleshooting, analysis and cartographic design.

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



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



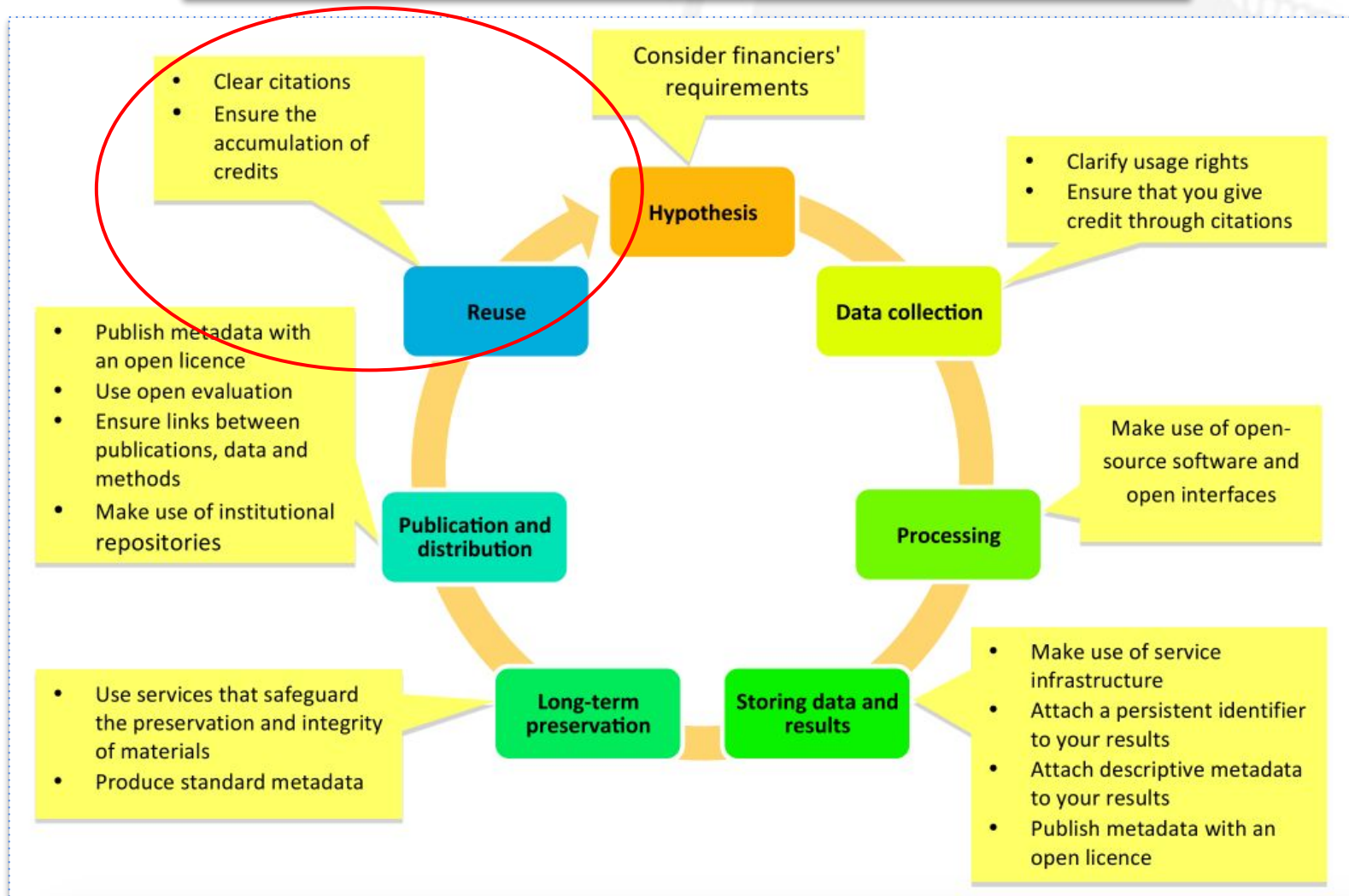
MANAGE, STORE, PRESERVE

Discover, Reuse & Cite

The outputs of digitisation / digital preservation projects should be publicly available as citable, scholarly resources. Open-access publishing of data should go hand-in-hand with other OA publishing efforts.

- ★ follow-up research
- ★ new research
- ★ research interviews
- ★ crowdsourced annotations / public feedback
- ★ teaching and learning

Open discovery, reuse and citation



Source: Foster Open Science: **What is Open Science?** Figure 1. Promoting openness at different stages of the research process. <https://www.fosteropenscience.eu/content/what-open-science-introduction>

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.

Open Science is ...



Egon Willighagen
@egonwillighagen

Following

#openscience is right to use, reuse, modify, and redistribute scholarly knowledge



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%2Fncchem.1149>



Where scholarly knowledge is ...

Open to participation	Open to (re)use	Open to the world
<ul style="list-style-type: none"> • No barriers based on race, gender, income, status • Involvement of societal partners in research priority setting • Evaluations that include societal relevance • Citizen science 	<ul style="list-style-type: none"> • Open Access, for people and machines, to: <ul style="list-style-type: none"> ○ 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 	<ul style="list-style-type: none"> • 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/>

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 <i>repeat</i> , <i>reproduce</i> , or <i>replicate</i> 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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Working with the **FAIR** guiding principles

- 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

Interoperable

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

Reusable

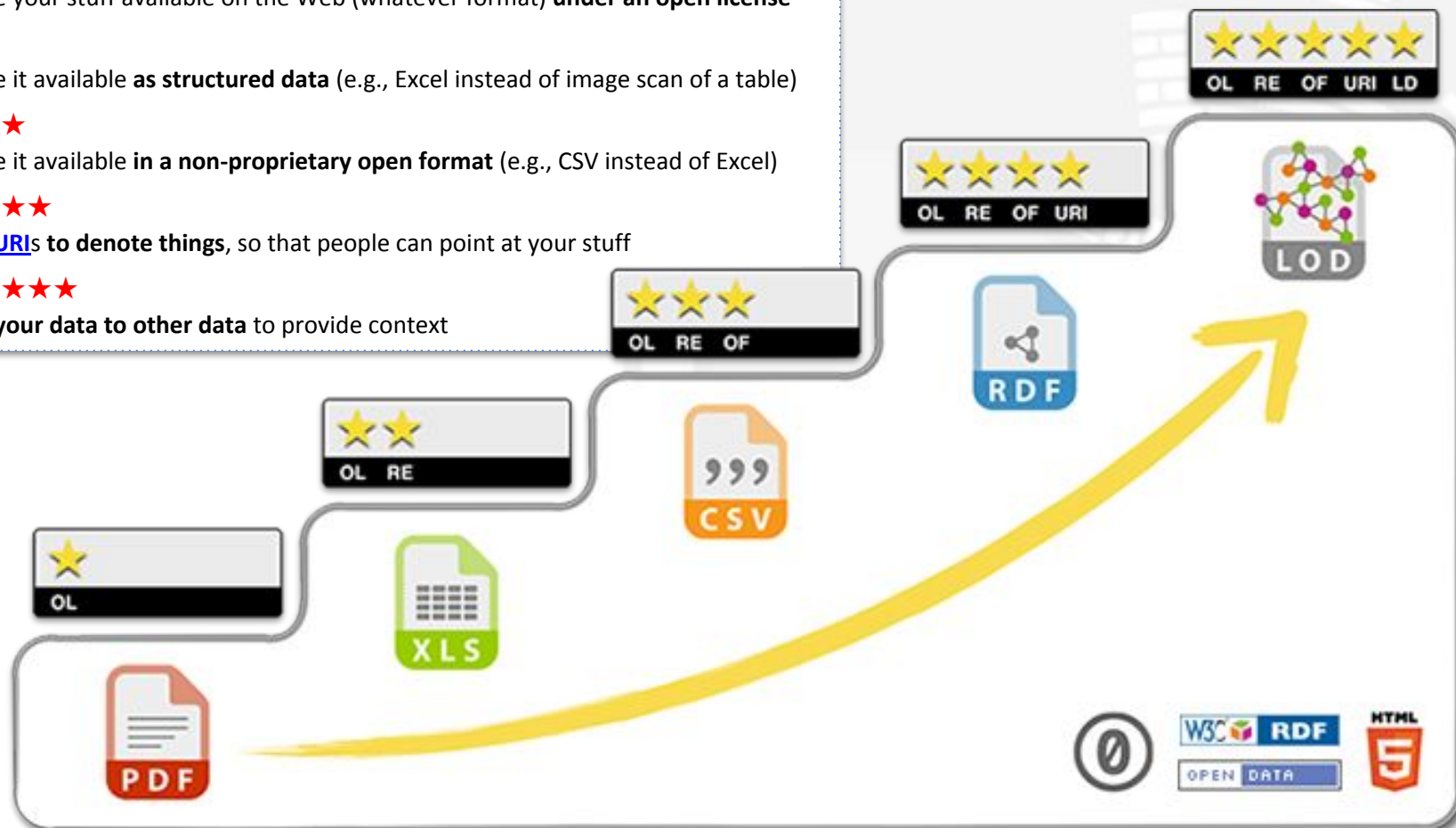
- Consider permitted use
- Apply machine-readable open licenses (e.g. CC-BY etc.)

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



5 ★ Open Data [Tim Berners-Lee]

- ★
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/>



SHARE & PUBLISH



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Share & Publish

Scholarly resources should be '*as open as possible, as closed as necessary*,' i.e. publicly available on an open-access platform.

- ★ manage copyright
 - ★ distribute data
 - ★ control access
 - ★ promote data
- Guidance on publishing and sharing sensitive data:
 - <http://www.ands.org.au/guides/sensitivedata>
 - Tools for anonymisation, de-identification, and disclosure control:
 - [Amnesia](#) (OpenAIRE)
 - [Statistical Disclosure Control \(sdcMicro\)](#) (IHSN)

'Reasons' for **not** sharing research data

1. **Misinterpretation** of the data.
2. **Misappropriation** of the data.
3. Damage to the researcher's **reputation** (CODATA-ICSTI, 2013).
4. The **myth** that scientific findings using shared data cannot be published in high-impact journals (Milham, et.al 2018).

... however, these fears immediately disappear the moment the data are properly managed and documented (CODATA-ICSTI, 2013).

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](#)



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



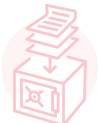
COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Why share your research data?

1. To **comply with Funding agencies and institutions** requiring results of scientific studies to be shared with the public as a condition for providing grants or awards.
2. To satisfy **publishers'** requests to deposit some datasets in public platforms.
3. To **confront some of the biases in data collection and analysis** (Atici et. al, 2013).
4. To **reproduce** or to **verify** research, and to **ask new questions** of extant data.
6. To **advance the state of research** and innovation (Borgman, 2012).
7. To **increase citation rate** (Piwowar, Day and Fridsma, 2007).
8. To **increase the visibility of individual researchers and their work online** (Peters et al., 2015).

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

🏠 > Using figshare

Select a category

Getting Started

How-To Guides

Information & Tips

Our Policies

Top 10 FAQs

Looking for another answer?

[Visit Community Forum](#)

[File Support Ticket](#)

How-To Guides

How to delete my account?

How to choose the most appropriate licence

A guide to licensing your data, for institutional users only.

How to edit or delete my data

Find out how to make changes to your data. Please note that only private data can be deleted.

How to upload confidential files, linked files, embargoed files, and metadata records only

There are many ways to upload your data and/or metadata.

How to sync ORCID and DataCite for figshare

How to connect figshare with your GitHub account

How to use Collections

How to use projects and collaborate on Figshare

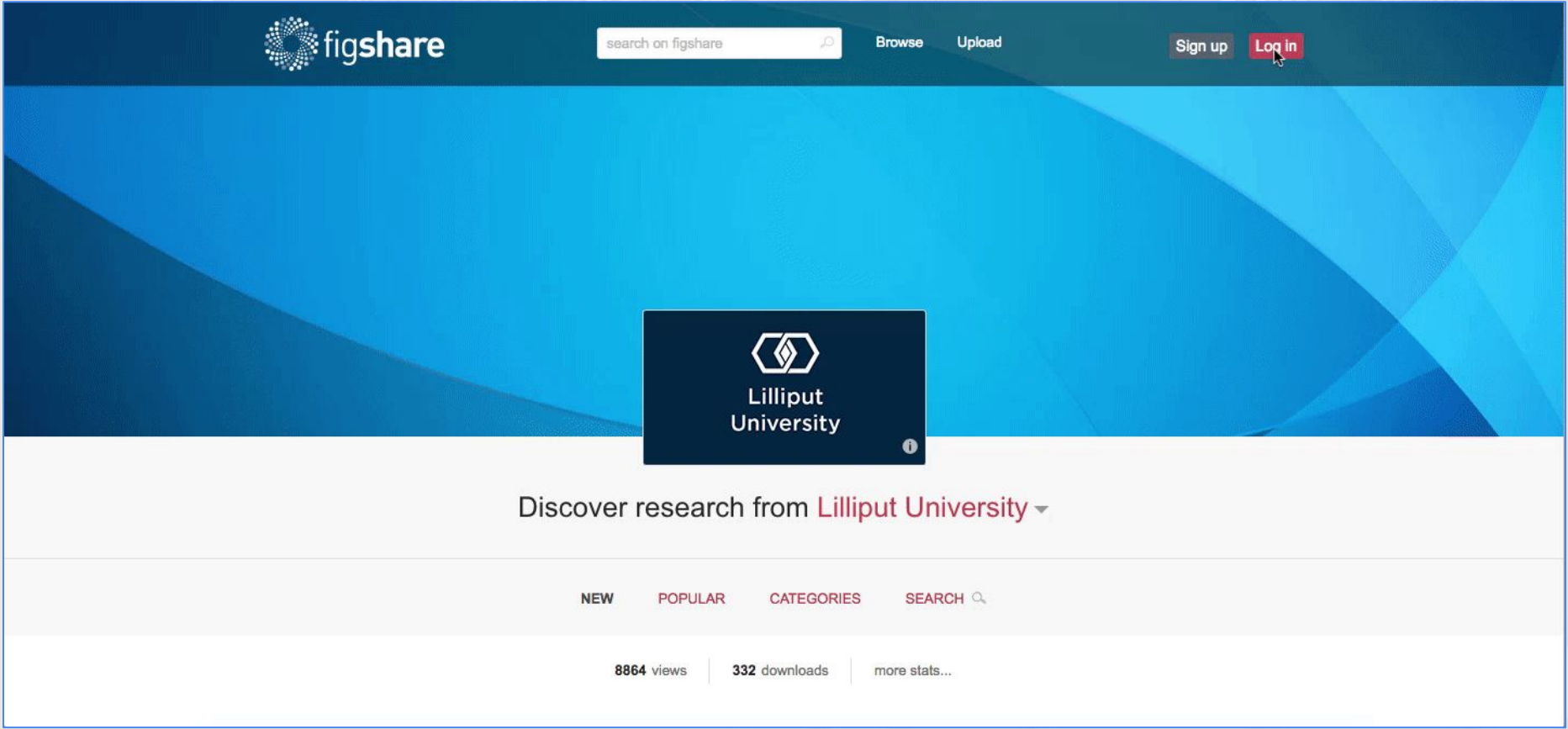
How to get a Personal Token

How to use the OAI-PMH

Source: <https://knowledge.figshare.com/>

Log in

Navigate to your institutions' Figshare portal and select 'Log in' at top right and enter your SSO (Single Sign-On) details. The process may vary depending on the respective institutional setup, but most Figshare instances in South Africa will likely be using [SAFIRE](#) for authentication.



Source: <https://knowledge.figshare.com/>

+Create a new item



Select the **+Create a new item** button from the top left (or just drag and drop your file into the browser). If you are uploading a big file size then you can use our desktop uploader, see: [how to use the desktop uploader](#).

	STATUS	TYPE	CREATED	SIZE
New draft item OU cool stuff		DATASET	31.8.2016 16:22	0 kB
Title		FILESET	31.8.2016 14:25	617.62 kB
New draft item OU cool stuff		DATASET	31.8.2016 10:42	0 kB
Example title one		METADATA ONLY	31.8.2016 10:23	0 kB
New draft item		METADATA ONLY	31.8.2016 10:13	0 kB

Source: <https://knowledge.figshare.com/>

Insert a Title

For information on how best to complete the fields for maximum exposure of your data, see: [Figshare how-to guide](#). Tips on how to fill out each of the fields can also be found on the right side of the screen.



Manage

needed to publish & get DOI

Title

figshare logo.jpg

Authors

Christopher George

Add co-authors by name or full email

Categories

Select categories

Group

Aeronautical, Automotive, Chemical and Materials Engineering

File type (what's this?)

Fileset

Keyword(s)

Add keywords for easy discovery

Description

Describe your data as well as you can

Tips

you can still drag more file(s) on the page or browse

Preview item (private)

or Esc

New draft item

studies

Promo material examples

figshare logo.jpg

peace_example.ZIP

Source: <https://knowledge.figshare.com/>

Add Authors

You can rearrange the order in which the authors appear and remove yourself as an author if you are uploading on behalf of someone. You can also search by entering the email address of the author you want to add.

figshare logo.jpg 34.57 kB

peace_example.ZIP 625.57 kB

Manage

needed to publish & get DOI

Title

How to upload your data

Authors

Christopher George

Add authors by name or full email

Categories

Select categories

Group

Aeronautical, Automotive, Chemical and Materials Engineering

File type (what's this?)

Fileset

Keyword(s)

Add keywords for easy discovery

Description

Describe your data as well as you can

Tips

Add the names of your co-authors. You can invite authors to join figshare or you can simply add them. After they were added, you can drag and drop the names to arrange them in the order that you wish.

Preview item (private)

or Esc

New draft item

studies

Promo material examples

figshare logo.jpg

peace_example.ZIP

Source: <https://knowledge.figshare.com/>

Select a Category

These are taken from the Australian Fields of Research classification system. You can choose more than one and either select from the drop-down menu or search for your subject area. It is suggested to find the closest fit, and then get more specific in the Keywords section, as required.

My drafts

+ Create a new draft

figshare logo

New draft item

OU code

Title

New draft item

OU code

Example title

needed to publish & get DOI

Title

How to upload your data

Authors

Alan Hyndman

Mark Hahnel

Add co-authors by name or full email

Categories

Select categories

Group

Aeronautical, Automotive, Chemical and Materials Engineering

File type (what's this?)

Fileset

Keyword(s)

Add keywords for easy discovery

Description

Describe your data as well as you can

manage

or Esc

you can still drag more file(s) on the page or browse

Preview item (private)

SIZE

USING

DATE

TYPE

NAME

FILE

New draft item

studies

Promo material examples

figshare logo.jpg

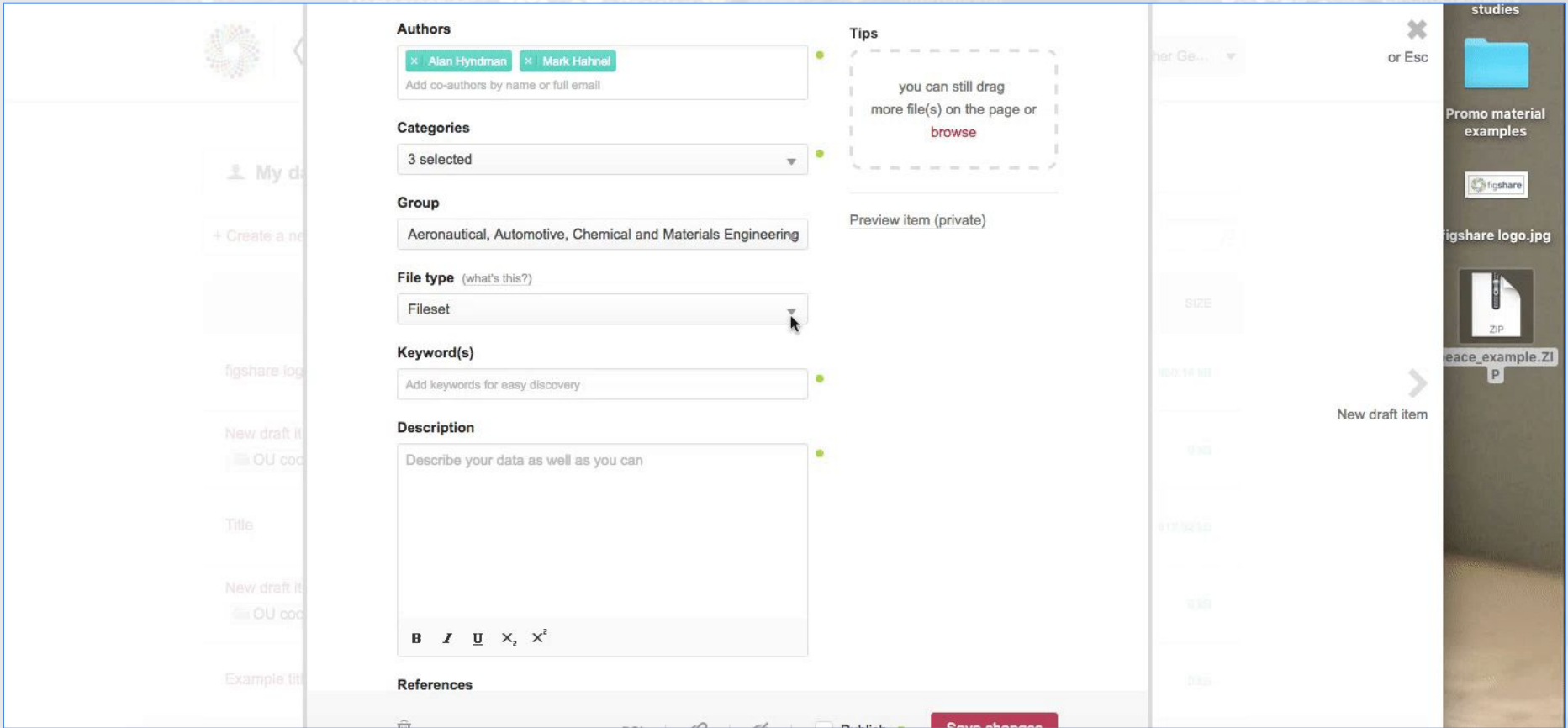
ZIP

peace_example.ZIP

Source: <https://knowledge.figshare.com/>

Select the File Type

Select the File type from the drop-down menu. Figshare accepts over 1,000 file extensions, and many are parsed right in the browser. If there's something that isn't supported yet, they will likely build it for you.



The screenshot shows the Figshare upload interface. On the left is a sidebar with navigation options like 'My drafts', 'Create a new draft', and 'New draft item'. The main area contains several sections: 'Authors' with a list of 'Alan Hyndman' and 'Mark Hahnel'; 'Categories' with a dropdown showing '3 selected'; 'Group' with the text 'Aeronautical, Automotive, Chemical and Materials Engineering'; 'File type' with a dropdown menu currently set to 'Fileset'; 'Keyword(s)' with a text input field; 'Description' with a large text area and a rich text toolbar; and 'References' at the bottom. A 'Tips' box on the right side of the main area says 'you can still drag more file(s) on the page or browse'. On the far right, there is a vertical sidebar with a 'studies' folder icon, 'Promo material examples', a 'figshare' logo, 'figshare logo.jpg', a 'ZIP' file icon, and 'peace_example.ZI'. At the bottom right of this sidebar is a 'New draft item' button.

Source: <https://knowledge.figshare.com/>

Add Keyword(s)

These should be more specific than the category and help others find your research. Add as many keywords as you want - just hit enter after each keyword. figshare remembers ones you have entered previously, and will also suggest ones.

The screenshot shows the figshare upload form. A red box highlights the 'Keyword(s)' field. A callout box shows a keyword suggestion 'ilifu' with a close button 'x'. The main form includes fields for Group, File type, Keyword(s), Description, References, and Funding.

Source: <https://knowledge.figshare.com/>

Write a **Description**

'The abstract' - include any relevant information that pertains to your research, e.g. information about the methodology, approval for data collection, or legal or ethical requirements. Be as descriptive as you can!

My drafts

+ Create a new draft item

figshare logo

New draft item

OU code

Title

New draft item

OU code

Example title

Presentation

Keyword(s)

x chemistry x Biological Datasets
x data acquisition software Add keywords for easy discovery

Description

B I U \times_2 \times^2

References

Link to references or related content

Funding

Add grant number or funding authority

Licence (what's this?)

CC-BY

GeoLocation

Tips

Add as much context as possible so that others can interpret your research and reproduce it. Make sure you include methodology, techniques used and if relevant information about approval for data collection to confirm adherence to legal or ethical requirements. The description should have at least four characters.

Preview item (private)

Source: <https://knowledge.figshare.com/>

Add **Funding** info

When typing in your funder information, Figshare will search the [Dimensions](#) database for your funder and hyperlink to the Dimensions page bridging the connection between your research outputs and the funder.

land use mapping ×

Project XYZ123 ×

XYZ123-Maps-Folder ×

XYZ123-England-Subfolder ×

Add keywords for easy discovery. Hit enter after each

Description

This set of maps accompanies my related publication, entitled 'A Land Cover Atlas of the United Kingdom', <https://doi.org/10.15131/shef.data.5266495>, which was published at the same time.

Some of the maps in this set feature in the Atlas, but at a lower resolution. I have deposited them here as high resolution images (300dpi PNG files) so that interested users can access and download them.

B *I* U x_2 x^2

Funding

Search grant by name/number or add your own

+ Add another grant info

References

Link to references or related content

Licence (what's this?)

CC BY

Preservation status

Tips

Search by grant number, name or funder body to add the funding information about your work. Multiple grants are supported.

Preview item (private)

Last edited on 06 Nov 2018 - 22:38

Cancel

Discard


Publish

Save changes

Source: <https://knowledge.figshare.com/>

Select a License

You can choose from a number of licenses based on your reuse requirements. For more information see: [How to choose the most appropriate licence.](#)



My drafts

+ Create a new draft

figshare logo

New draft item

OU content

Title

New draft item

OU content

Example title

Presentation

Keyword(s)

chemistry

Biological Datasets

data acquisition software

Add keywords for easy discovery

Description

Describe the data as best as you can.

Include information about approval for data collection.

B

I

U

x₂

x²

References

Link to references or related content

Funding

Add grant number or funding authority

Licence (what's this?)

CC-BY

GeoLocation

Tips

Add as much context as possible so that others can interpret your research and reproduce it. Make sure you include methodology, techniques used and if relevant information about approval for data collection to confirm adherence to legal or ethical requirements. The description should have at least four characters.

Preview item (private)

Other Ge...

or Esc

SIZE

1024x768

1024x768

1024x768

1024x768

New draft item

studies

Promo material examples

figshare logo.jpg


ZIP

peace_example.ZI

Source: <https://knowledge.figshare.com/>

References | Linking Files

1. You can copy & paste URLs into the **References** box (add more fields by hitting Enter).
2. You can link to data that are hosted elsewhere by selecting '**Link file.**'



My drafts

Create a new draft

New draft item

New draft item

test

New draft item

Role of E54

to upload, drag file(s) on the page or [browse](#)

☐ Metadata record only

Link file

needed to publish & get DOI

Title

New draft item

Authors

Christopher George

Add co-authors by name or full email

Categories

Select categories

Group

Aeronautical, Automotive, Chemical and Materials Engineering

File type (what's this?)

Figure

Keyword(s)

DOI

Publish


Save changes


Tips


Your file(s) will not be visible to the public. The confidentiality is set as soon as the item becomes public.


Preview item (private)


Source: <https://knowledge.figshare.com/>


















Tick Publish (or not)

Make sure everything you have submitted is accurate - once it's published, it's **permanently available**. You can go back and edit items after you've made them publicly available. Some changes may trigger a new version. See [here](#) to find out which amends will generate a new version of your item.



My drafts

+ Create a new item

figshare logo

New draft item

OU code

Title

New draft item

OU code

Example title

Funding

Add grant number or funding authority

Licence (what's this?)

CC-BY

GeoLocation

Access Level Commitment

Location

Strain

Publish this item

Apply embargo

Make file(s) confidential

Generate private link

DOI Reserve Digital Object Identifier

Tips

you can still drag more file(s) on the page or browse

Preview item (private)

Other Ge...

or Esc

SIZE

DESCRIPTION

NAME

STATUS

DATE

studies

Promo material examples

figshare logo.jpg

peace_example.ZIP

New draft item

Save changes

Source: <https://knowledge.figshare.com/>

Review / Curation

All of your items will first be reviewed by curator before they are approved (i.e. published). Someone may contact you, generally with suggested edits towards improving the metadata. In some cases, they may also take down files that have been published in error.

My drafts

+ Create a new draft item

figshare logo

New draft item

OU code

Title

New draft item

OU code

Example title

Add grant number or funding authority

Licence (what's this?)
CC-BY

GeoLocation

Access Level Commitment

Location

Strain

Publish this item

Apply embargo

Make file(s) confidential

Generate private link

DOI Reserve Digital Object Identifier

Tips
you can still drag more file(s) on the page or browse

Preview item (private)

Other Ge... ▾

X or Esc

SIZE

500 x 140 px

40.00

517.55x140px

42.00

512px

New draft item

studies

Promo material examples

figshare logo.jpg

peace_example.ZIP

DOI

Publish

Save changes

Source: <https://knowledge.figshare.com/>



Those are the basics, and now
for a few more features ...

Applying an Embargo

Embargoing data serves to allow the author(s) a period of sole use of the data. After you click 'save' & 'publish' - and approval by the curator -, users are presented with a countdown timer.

My drafts

+ Create a new draft

New draft item

New draft item

test

New draft item

Role of E54

Licence (what's this?)
CC-BY

GeoLocation

Access Level Commitment

Location

Strain

☒ Publish this item

☐ Apply embargo

☒ Make file(s) confidential

☐ Generate private link

☐ DOI Reserve Digital Object Identifier

Tips


The embargo period begins as soon as the item is made public.
If you set a period instead of a specific date, that period will start on the publication day.

Preview item (private)

Source: <https://knowledge.figshare.com/>

Publishing a Metadata-only record

In this case no files are uploaded, but a description of the study is available, including information about how to contact the authors. This is a good option for getting a DOI for - and linking to - **data that are stored / hosted somewhere else.**



My drafts

Create a new draft

New draft item

New draft item

test

New draft item

Role of E54

to upload, drag file(s) on the page or [browse](#)

☐ Metadata record only

Link file

needed to publish & get DOI

Title

New draft item

Authors

Christopher George

Add co-authors by name or full email

Categories

Select categories

Group

Aeronautical, Automotive, Chemical and Materials Engineering

File type (what's this?)

Dataset

Keyword(s)

DOI

☐ Publish

Save changes

Tips

The embargo period begins as soon as the item is made public. If you set a period instead of a specific date, that period will start on the publication day.

Preview item (private)

Source: <https://knowledge.figshare.com/>

Creating a Project

Projects are **collaborative spaces used for ongoing work**. You can upload data that is in progress and have users make comments. **Projects are secure spaces that can be used for sensitive data**. You can also **collaborate with people outside your institution** by inviting them to your project.

	STATUS	TYPE	CREATED	SIZE
New draft item		FILESET	29.9.2016 14:32	0 kB
New draft item		DATASET	29.9.2016 14:32	0 kB
test		DATASET	29.9.2016 14:29	0 kB
New draft item		DATASET	28.9.2016 11:07	0 kB
utations of PIK3CA in breast cancer: a comparative ...		DATASET	26.9.2016 17:12	1.93 MB

Source: <https://knowledge.figshare.com/>

Individual vs. Group Projects

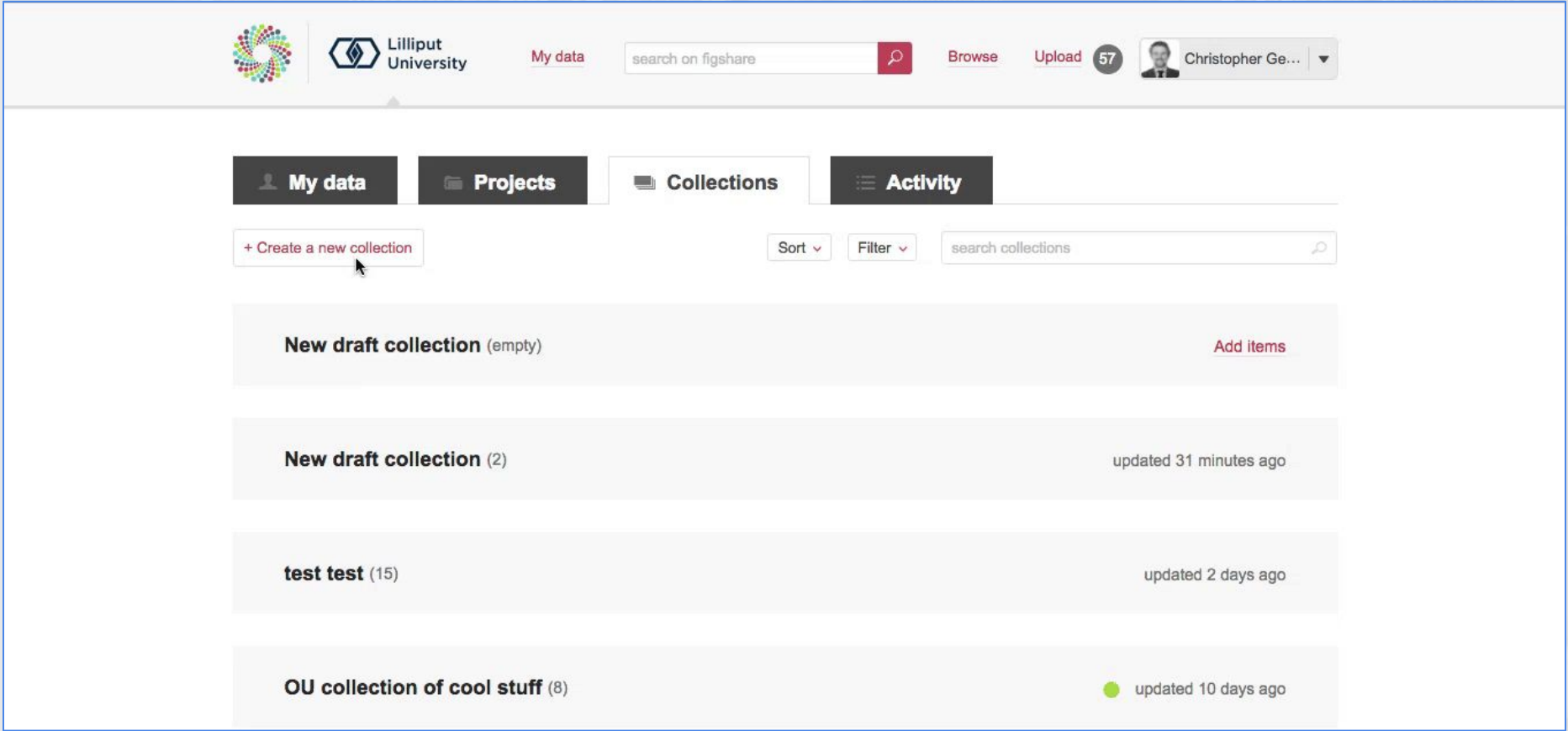
If a project is for individual use, select the **individual** option under Allocate storage. To use group storage, select the **group** option and find the relevant group. Add users to the project by searching on the right hand side (collaborator or viewer). Collaborators can comment on the project and the data within it and upload items, while viewers can only view the data. NB: Items must be downloaded, edited, and re-uploaded to the Project.

Individual Projects	Group Projects
Everyone uses their own quota and account storage.	Submitters' quota will not be used, storage allocation comes directly from the project.
People take their work with them if they leave the project.	All work is stored on institutional storage and remains within the project space if people leave.
Items are created using the metadata schema of the submitter.	Contributors must adopt the metadata schema of the project owner.
Items appear in the subgroup of the uploader.	Items appear in the subgroup of the project owner.

Source: <https://knowledge.figshare.com/>

Creating a Collection

Collections are ways of collating data that bring it together under a theme. They can be either private or public and can be assigned a DOI. One key difference the Projects feature is that this offers a way of gathering and re-using data published on Figshare which you did **not** create yourself.



The screenshot shows the Figshare interface for a user named Christopher Ge... (57 uploads). The 'Collections' tab is selected, showing a list of collections. A button '+ Create a new collection' is highlighted with a mouse cursor. The list includes:

- New draft collection** (empty) - Add items
- New draft collection** (2) - updated 31 minutes ago
- test test** (15) - updated 2 days ago
- OU collection of cool stuff** (8) - updated 10 days ago

Source: <https://knowledge.figshare.com/>

+Add items to a Collection

Once you've created a Collection, select **Add public items** or **My data**. If you select from public items, you can search and select the items you wish to add to your collection. The same process applies if you choose from your data. If you find a public item you wish to add to a Collection, simply select **+Collect** and choose the Collection to add it to.

My data

Projects

Collections

Activity

← Cool stuff

MANAGE ⚙

⌵ Show collection details

+

Add public items

Add to your collection one or more public items from figshare

🔗

Add from My data

Add items that are already in your My data section

Source: <https://knowledge.figshare.com/>

Publish a Collection

Once you've collected your data, you can make your Collection public by selecting the **Manage** gear wheel on the right side of the screen and selecting **Publish collection**. Once you've published a collection, it's permanent - the same as publishing an item.

Source: <https://knowledge.figshare.com/>



- Data availability statement
- Institution-specific DOI

Data availability statement

The [ilifu Research Data Management and Open Science Guidelines](#) (V.0.5 June 2018) require users to share their data openly by default (with various exceptions provided for). This also requires users to provide a link to where the data underpinning their traditional, written research outputs can be found. Provided are a few examples of wording data availability statements based on whether:

1. you collected or generated your own data
2. you re-used publicly-available data
3. your data requires an embargo period, or
4. (if applicable), why your data is not publicly available.

Data availability statements should be included in the abstract or executive summary, as well as in the methods/methodology section. When data is uploaded to Figshare, a doi is automatically generated on saving the new item. This doi can be copied & pasted into the data availability statement.

Data availability statement



DIGITAL LIBRARY
SERVICES

Generic version:

The data supporting this paper can be accessed on the University of Cape Town's Institutional Data Repository (ZivaHub) under the following doi:



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Data availability statement



DIGITAL LIBRARY
SERVICES

(1) You produced, collected or generated **your own data**:

The data that support the findings of this study are openly available in {repository name, e.g. 'ZivaHub'} under the following {identifier(s), e.g. DOI(s)}: '...'

Example from the Department of Biological Sciences:

The data that support the findings of this study are openly available in ZivaHub with the identifier [10.25375/uct.6561776](https://doi.org/10.25375/uct.6561776)



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Data availability statement

(2) You re-used public data:

The data that support the findings of this study are available in [repository name] at [URL/DOI], reference number [reference number]. These data were derived from the following resources available in the public domain: [list resources and URLs]

Example from the Faculty of Commerce:

The data that support the findings of this study are openly available in DataFirst with the identifier

<https://datafirst.uct.ac.za/dataportal/index.php/catalog/570>

Data availability statement

(3) You collected your own data, but require an **embargo period**:

The data that support the findings **will be available** in [repository name] at [URL / DOI link] following a [6 month] embargo from the date of publication to allow for commercialisation of research findings.

Example from Faculty of Science:

The data that support the findings will be available in ZivaHub at 10.23456/uct.examplelink following a 12 month embargo from the date of publication to allow for commercialisation of research findings.

Data availability statement

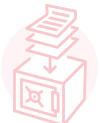
(4) Ethical concerns **prevent** data sharing:

Note: even if your data cannot be ethically shared, the metadata describing the study should still be made available as it builds your academic profile and may lead to future collaboration. UCT's data repository [ZivaHub](https://uct.figshare.com/) (<https://uct.figshare.com/>) allows the creation of **metadata-only records** that can be used to describe a study without openly sharing data objects.

Example:

Due to ethical concerns, supporting data cannot be made openly available. Further information about the data and conditions for access are available at the University of Bath data archive: <https://doi.org/10.15125/12345>

If none of the above adequately describe your data sharing requirements, the University of Bath has comprehensive guidelines (<http://www.bath.ac.uk/research/data/archiving-data/writing-a-data-access-statement/>) on how to write different kinds of data accessibility statements.



The (institution-specific) doi on Figshare

The dois created and managed via Figshare are minted by [DataCite](#). A doi is an alphanumeric string beginning with ‘10’, i.e 10.25375/**uct**.7143581.v1. To make this resolve to a digital object, it needs to be prefixed with ‘https://doi.org/’.

Example: <https://doi.org/10.25375/uct.7143581.v1>, further explained:

https://	doi.org/	10.25375/	uct.	7143581.	v1
A secure extension of HTTP.	International DOI foundation, who ensure that DOIs stay true.	‘10’ is the standard doi directory code; ‘25375’ is the registrant code for ZivaHub.	Indicates that the DOI originates from a UCT researcher or research project.	Administrative subdivision code.	Version number.



Groups in Figshare



Discover research from **Monash University** ▲

1.



+ Follow

ARC Centre of Excellence for Climate
System Science

COLLOQUY
text theory critique

Faculty of Information Technology

Learning Spaces



Monash University Theses

2.



Schooling Food



NEW

POPULAR

CATEGORIES

SEARCH 🔍

18,711 posts

2,155,510 views

346,062 downloads

more stats...

Source: <https://monash.figshare.com/sciencetheses>

level 1 group

Discover research from **Monash University Theses** ▲ ⓘ



+ Follow

Faculty of Art, Design and Architecture
Theses

Faculty of Arts Theses

Faculty of Business and Economics
Theses

Faculty of Education Theses

Faculty of Engineering Theses

Faculty of Information Technology
Theses

Faculty of Law Theses

Faculty of Medicine, Nursing and
Health Sciences Theses

Faculty of Pharmacy and
Pharmaceutical Sciences Theses

Faculty of Science Theses

Monash University Accident Research
Centre (MUARC)

NEW

POPULAR

CATEGORIES

SEARCH 🔍

7,287 posts

1,206,789 views

149,391 downloads

more stats...

Source: <https://monash.figshare.com/sciencetheses>

level 2 group

Discover research from **Faculty of Science Theses**

3.



+ Follow

NEW

POPULAR

CATEGORIES

SEARCH 🔍

908 posts

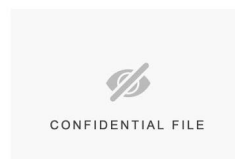
139,530 views

18,682 downloads

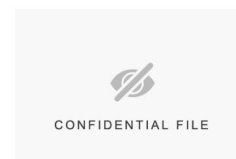
more stats...



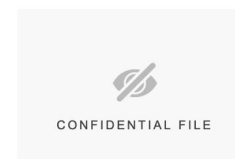
Working the food and beverage waste puzzle: extractants and ame...
TEMMA CARRUTHER... today



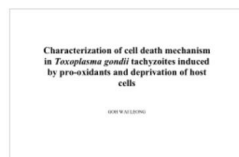
Alkaline earth and lanthanoid complexes of sterically encumberi...
CASPAR NATANIEL D... today



A Study of Service Triads
Sourav Sengupta today



The Specialised Metabolite Dhurrin Plays Important Roles in the Growt...
MUHAMMAD NOUMAN... 14/03/2019



Characterization of cell death mechanism in *T. gondii* induced by...



Designing and Optimising Electrocatalysts for CO₂ Reduction



Extremal Graph Theory for Minors, Improper Colourings and Gonality



Rapid and repeatable local adaptation in the global invader A...

Source: <https://monash.figshare.com/sciencetheses>

level 3 group



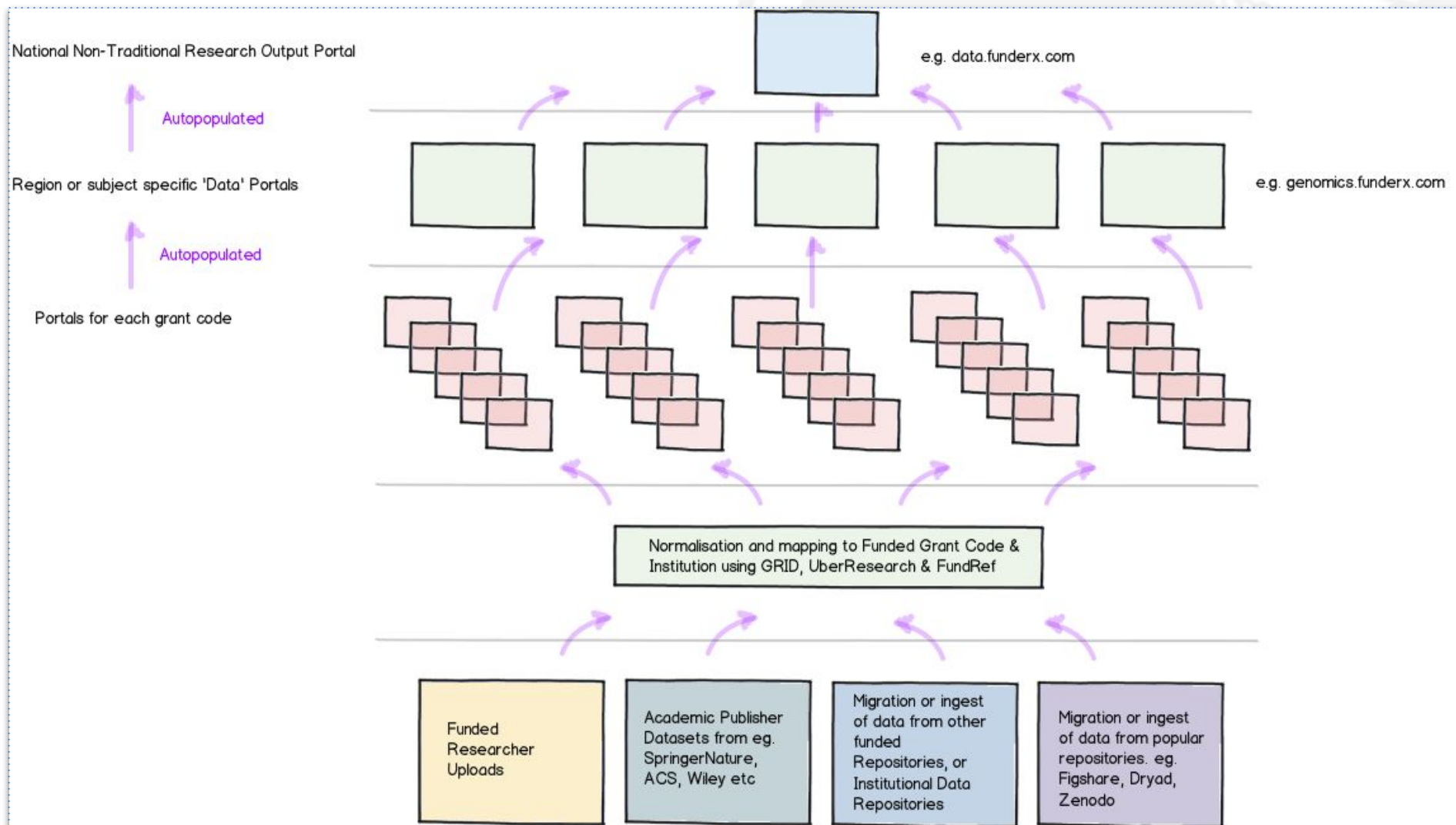
Tiered **instances** of Figshare

*Tier 1: **national*** - DIRISA storage? curation?

*Tier 2: **regional*** - (WC: ilifu) instance? group?

*Tier 3: **institutional*** - (WC: ilifu) sub-groups? projects?

*Tier 4: **individual*** - (WC: ilifu) affiliates? tags?



Source: Hahnel, Mark: *Introduction to Figshare*. DIRISA - Figshare pilot in Pretoria and Durban. Presentation. Available: <https://www.dirisa.ac.za/wp-content/uploads/2017/07/figsharefest-UCT-July-17.pdf>

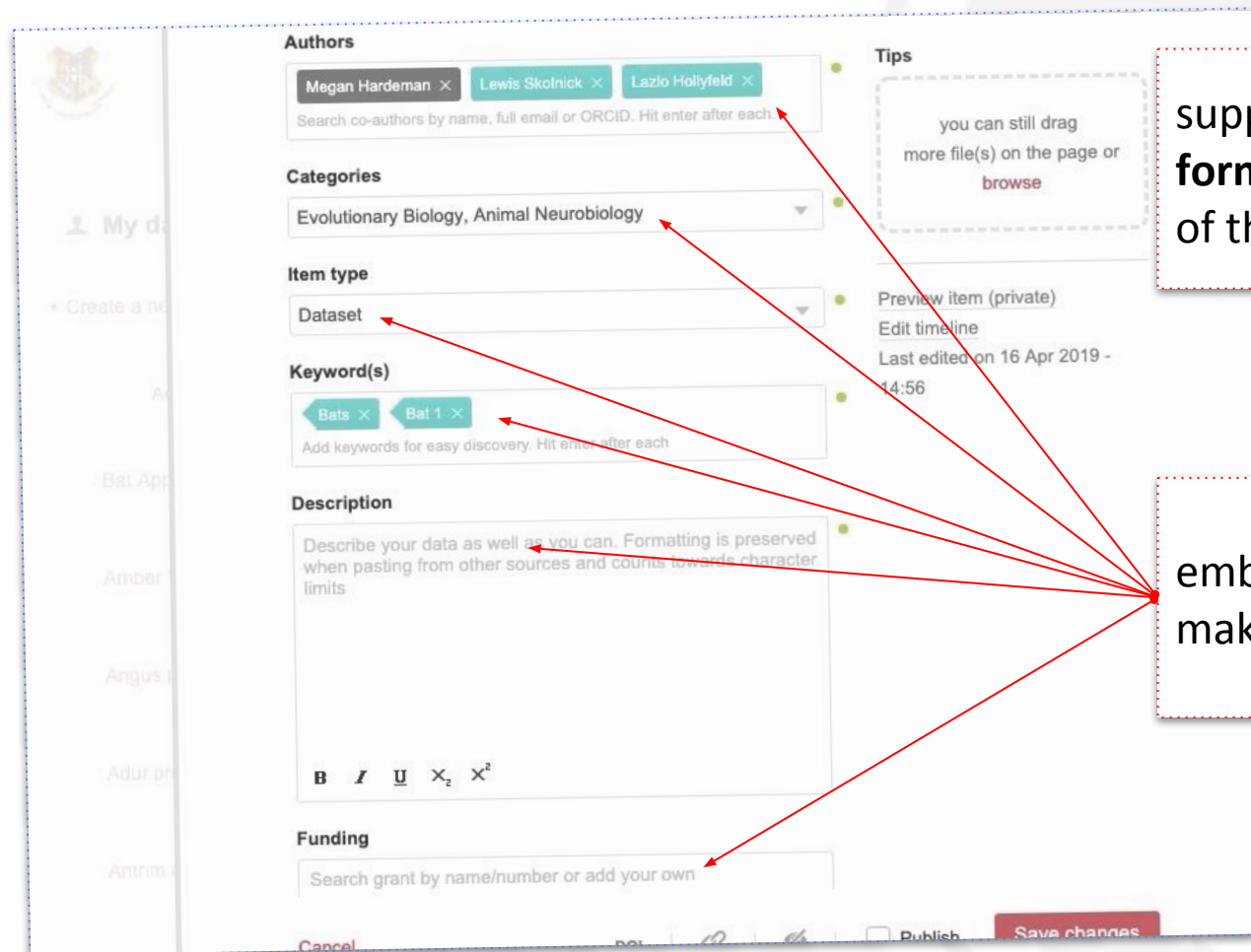




recap (Figshare for Institutions)

ZivaHub | Open Data UCT

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



The screenshot shows the ZivaHub data upload form. Red arrows point from the following fields to a central point on the right:

- Authors:** Megan Hardeman, Lewis Skolnick, Lazlo Hollyfeld
- Categories:** Evolutionary Biology, Animal Neurobiology
- Item type:** Dataset
- Keyword(s):** Bats, Bat 1
- Description:** Describe your data as well as you can. Formatting is preserved when pasting from other sources and counts towards character limits
- Funding:** Search grant by name/number or add your own

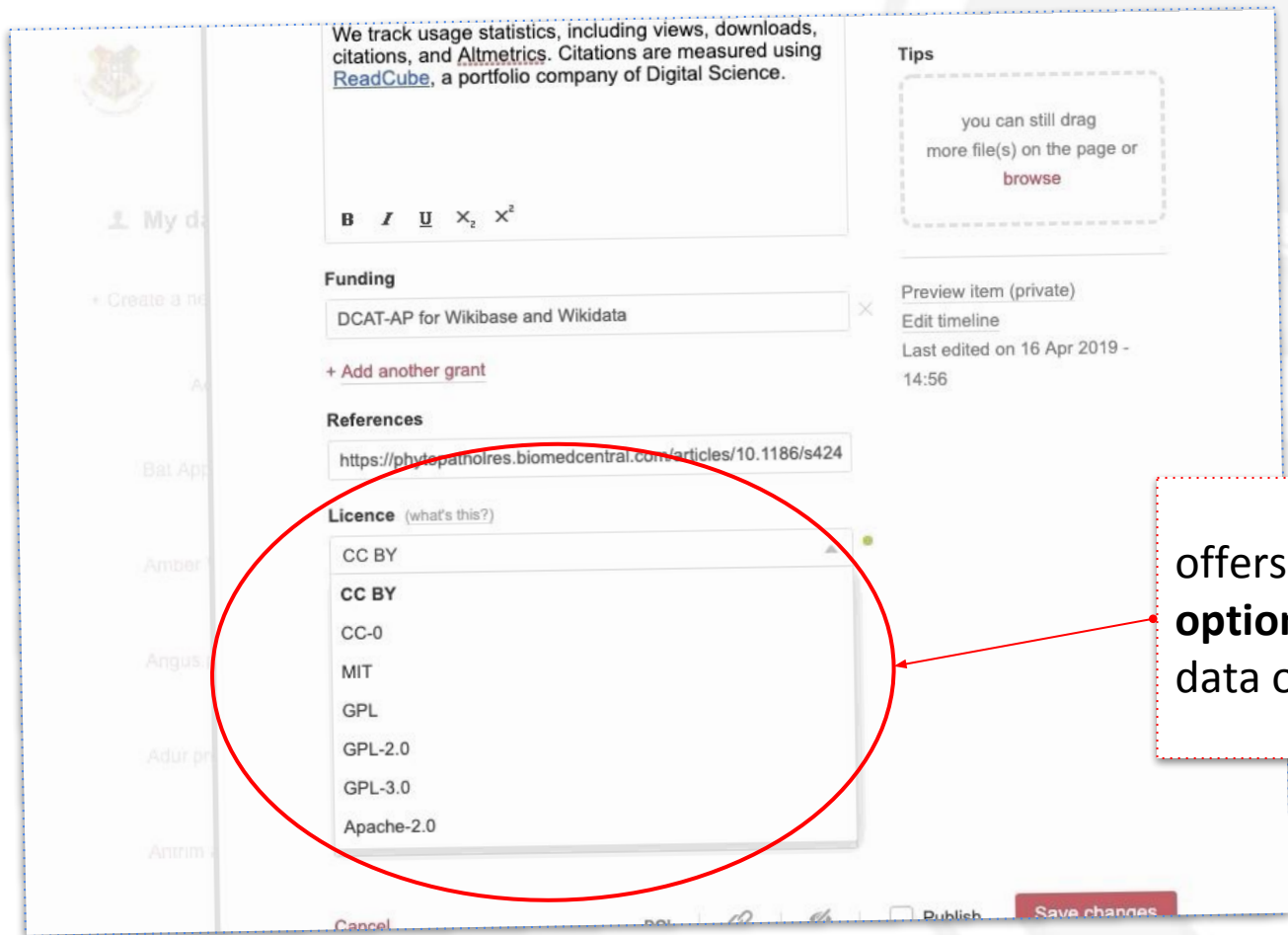
supports the upload of **any file format**, and aims to visualise all of them

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/>



We track usage statistics, including views, downloads, citations, and Altmetrics. Citations are measured using [ReadCube](#), a portfolio company of Digital Science.

Tips
you can still drag more file(s) on the page or [browse](#)

Funding
DCAT-AP for Wikibase and Wikidata
[+ Add another grant](#)

References
<https://phytopatholres.biomedcentral.com/articles/10.1186/s424>

Licence (what's this?)

- CC BY
- CC BY
- CC-0
- MIT
- GPL
- GPL-2.0
- GPL-3.0
- Apache-2.0

Buttons: Cancel, Publish, Save changes

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/>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Guidance for open licencing

- Check if you have the right to assign a license to the data (no third parties' rights involved)
- Check if the chosen license is in conformance with conditions prescribed by your informed consent and other contractual agreements.
- If you are depositing your data with a publisher (e.g. as supplement to a print publication), you should ensure that only non-exclusive rights of use are granted to the publisher

Authors: DataViz Knowledge Base. Available at: <https://datawizkb.leibniz-psychology.org/index.php/after-collection/what-should-i-know-about-licenses/>



Which open license to choose?

Choose a License

Ans Click to select the license e the search to find the license you want

Start again

What do you want to deposit?

Software Data

Search for a license...

Public Domain Mark (PD)

See full text

The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.

Publicly Available PD

Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available CC 0 OPEN DATA

Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Publicly Available CC BY OPEN DATA

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.

Publicly Available CC BY SA OPEN DATA

Creative Commons Attribution-NoDerivs (CC-BY-ND)

The no derivatives creative commons license is straightforward; you can take a work released under this license and re-distribute it but you cannot change it.

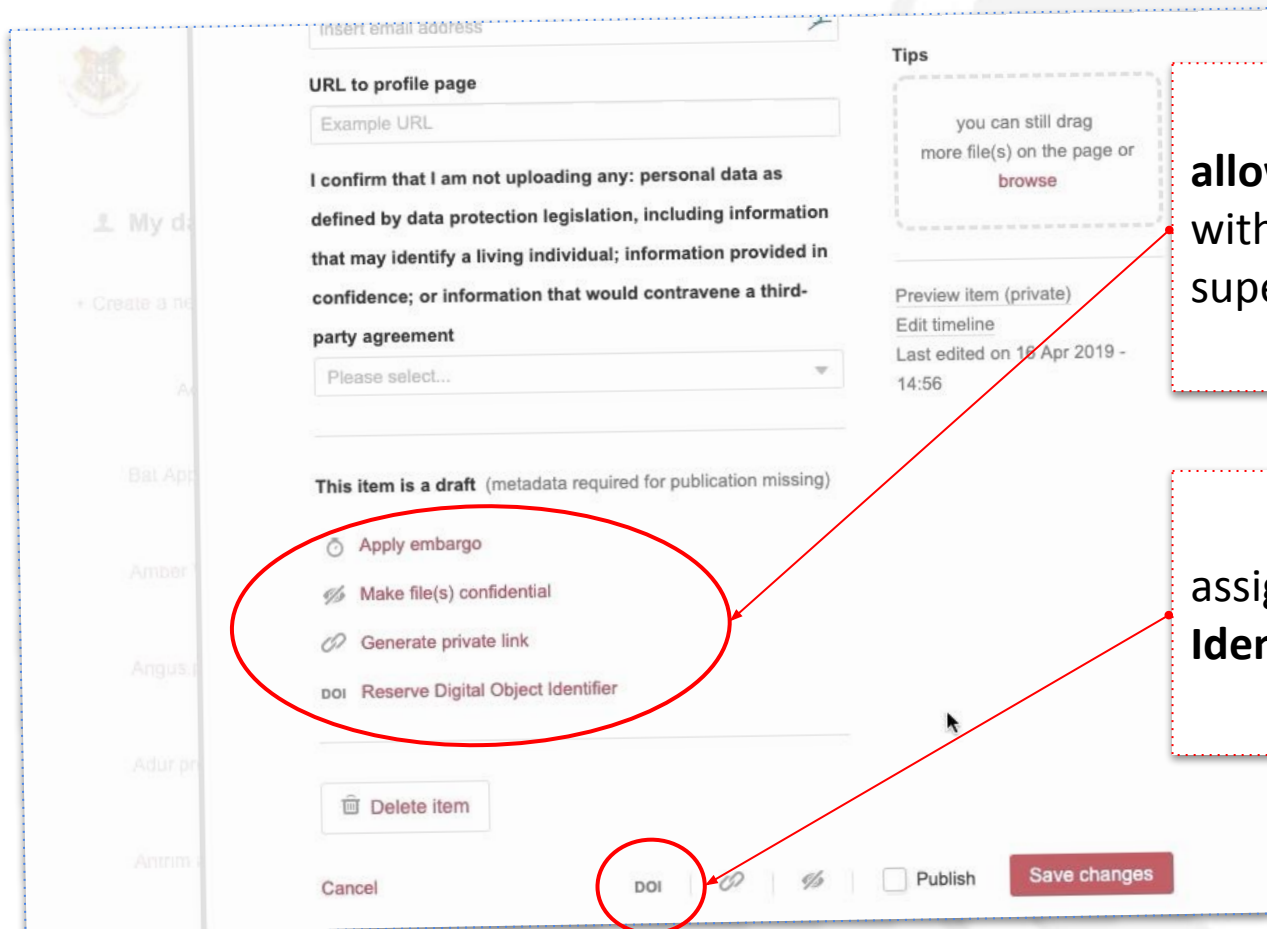
Publicly Available CC BY ND

<https://ufal.github.io/public-license-selector/>

Source: License Selector. Available at: <https://ufal.github.io/public-license-selector/>

ZivaHub | Open Data UCT

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



insert email address

URL to profile page

Example URL

I confirm that I am not uploading any: personal data as defined by data protection legislation, including information that may identify a living individual; information provided in confidence; or information that would contravene a third-party agreement

Please select...

This item is a draft (metadata required for publication missing)

Apply embargo

Make file(s) confidential

Generate private link

DOI Reserve Digital Object Identifier

Delete item

Cancel

DOI

Publish

Save changes

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

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

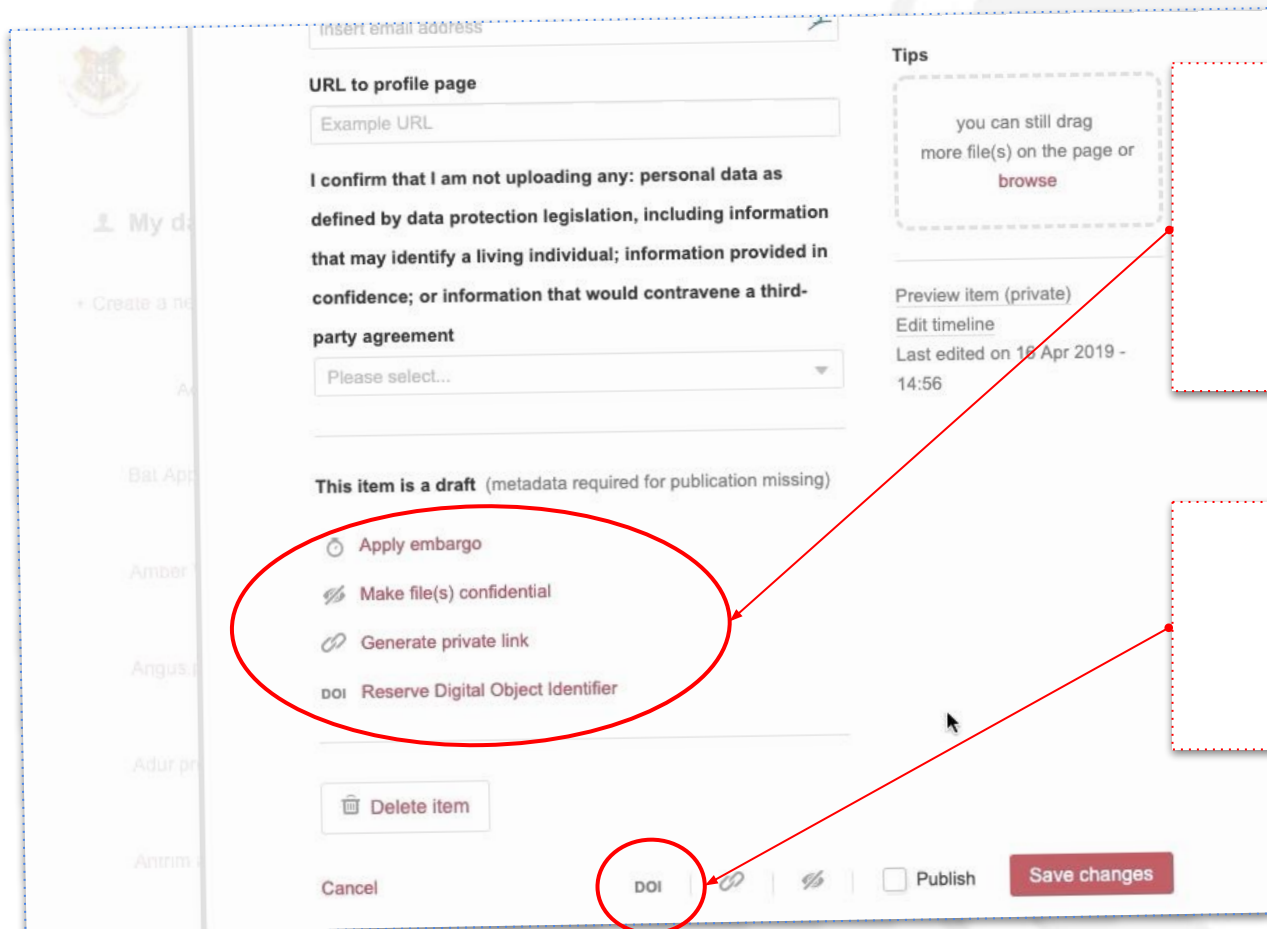
Source: Figshare. End User Guide GIFs. Available: <https://figshare.com/s/b3600c85f576d88d067b>

Open / Shared / Closed data

- **Open data** - the Open Data Institute (ODI) defines Open Data as those that anyone can access, use and share. According to the ODI, *open data must be licensed to make clear that anyone can use the data in any way they want, including transforming, combining, and sharing it with others, even for commercial purposes.*
- **Shared data** - similar to Open data, shared data may be made widely accessible but could have some conditions such as non-commercial reuse or reuse with attribution. It is important to note that *not all shared data has to be available to anyone.* Sometimes shared data is only made available to specific groups such as peers from another university.
- **Closed data** - if researchers are dealing with highly sensitive data - such as sensitive personal data or commercially sensitive data - it may not be possible to share the data at all. However, *even in such cases a metadata description of the research data should be shared.* Sharing of sensitive data can also be supported by making use of safe havens where only authorised users are given controlled access.

ZivaHub | Open Data UCT

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



insert email address

URL to profile page

Example URL

I confirm that I am not uploading any: personal data as defined by data protection legislation, including information that may identify a living individual; information provided in confidence; or information that would contravene a third-party agreement

Please select...

This item is a draft (metadata required for publication missing)

Apply embargo

Make file(s) confidential

Generate private link

DOI Reserve Digital Object Identifier

Delete item

Cancel

DOI

Publish

Save changes

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

all item records get assigned a **Digital Object Identifier (DOI)**

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

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.

Collecting & collaborating on ZivaHub

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

Collections	<p>gather interesting content from other Figshare users according to whatever principle you desire, and share the collection with others.</p> <p>Can be used to create course readers, examples of existing datasets in your field, etc.</p>
Projects	<p>Create a space for private uploads and private sharing with specific individuals (collaborators in other institutions, external reviewers, etc.)</p> <p>Default private; can be published later.</p>

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 Institutions¹ platform, is ISO-certified² as a **trusted digital repository**. As such, it provides a **GDPR-compliant³** 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 type⁴:

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](https://figshare.com/institutions)

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 quota is assigned per faculty and by DLS. Project owners should contact [dls](mailto:dls@uct.ac.za) for further details. For projects working with large datasets (beyond 5GB per item) please contact ICTS Data Storage.

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>



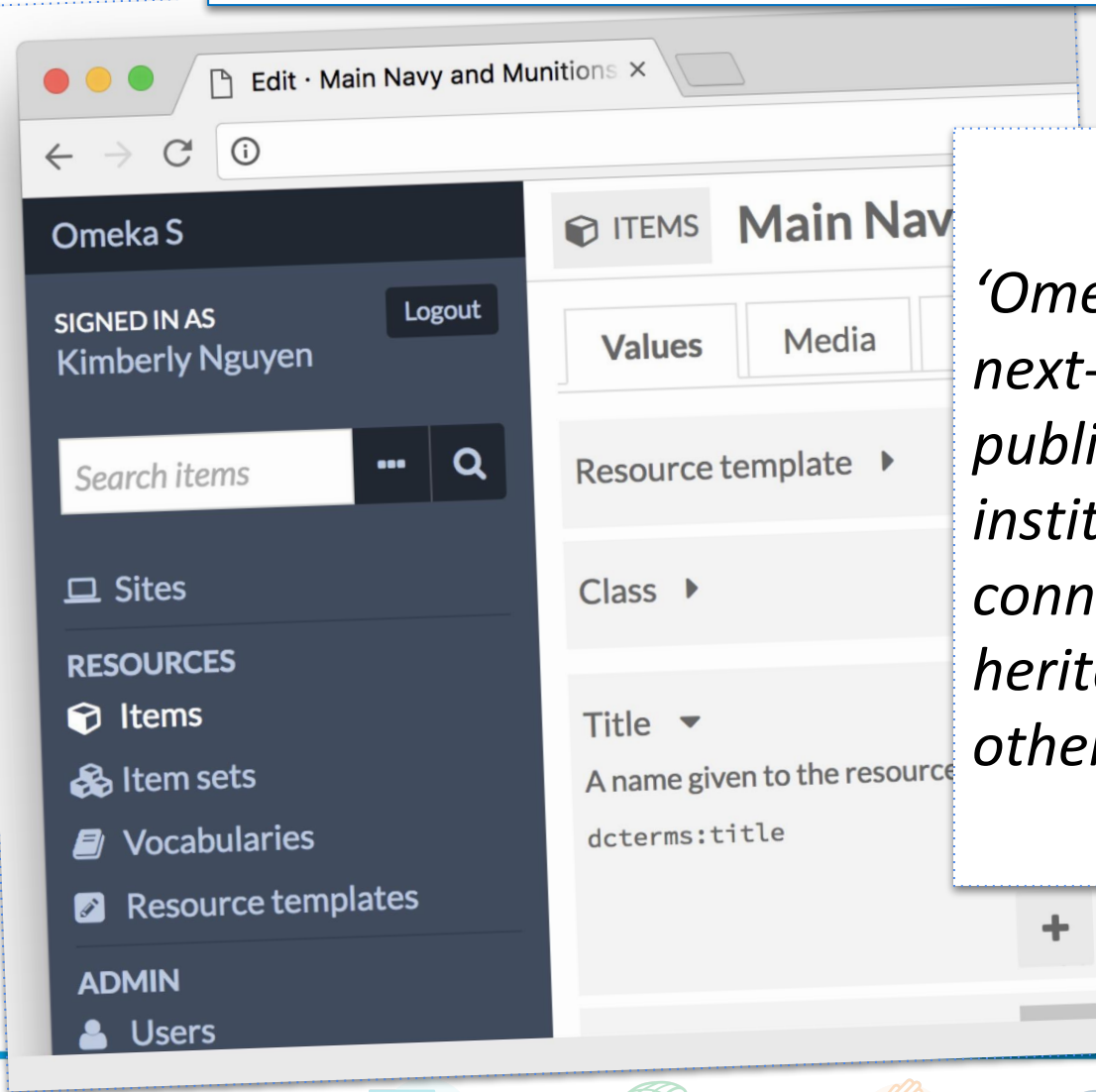
Other forms of digital collections:
building community-driven, institutional showcases
using a **iiif server** and **Omeka S** (both installed onsite)



omeka s

Omeka S

<https://omeka.org/s/>



‘Omeka S is a next-generation web publishing platform for institutions interested in connecting digital cultural heritage collections with other resources online.’

Source: <https://omeka.org/s/>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



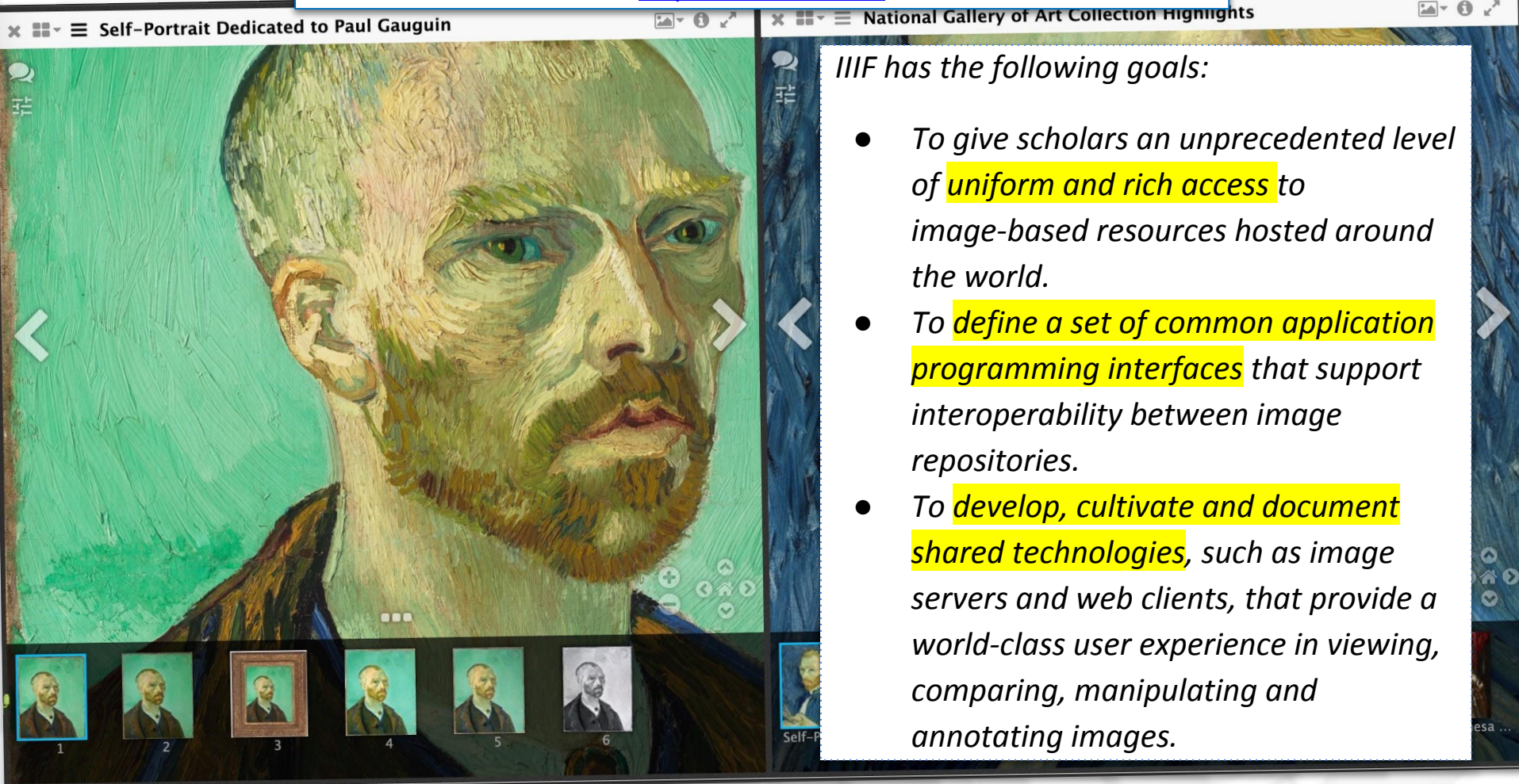
SHARE & PUBLISH



MANAGE, STORE, PRESERVE

iiif server

<https://iiif.io/about/>

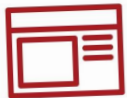


Source: <https://projectmirador.org/demo/>



Omeka S & iiif server

<https://omeka.org/s/>



INSTALL ONCE

Create and manage many sites with a streamlined install



CONNECT TO THE SEMANTIC WEB

Publish items with linked open data.



SHARE WITH DPLA

Describe items with DPLA-ready resource templates.



EXTEND AND BUILD

Extend functionality of Omeka S sites with modules to



DESIGN WITH EASE

Style each Omeka S site with a different fully-respons

*'The **International Image Interoperability Framework (IIIF)** defines several application programming interfaces that provide a standardised method of describing and delivering images over the web, as well as "presentation based metadata"^[1] (that is, structural metadata) about structured sequences of images. If institutions holding artworks, books, newspapers, manuscripts, maps, scrolls, single sheet collections, and archival materials provide IIIF endpoints for their content, any IIIF-compliant viewer or application can consume and display both the images and their structural and presentation metadata.'*

Source: https://en.wikipedia.org/wiki/International_Image_Interoperability_Framework



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



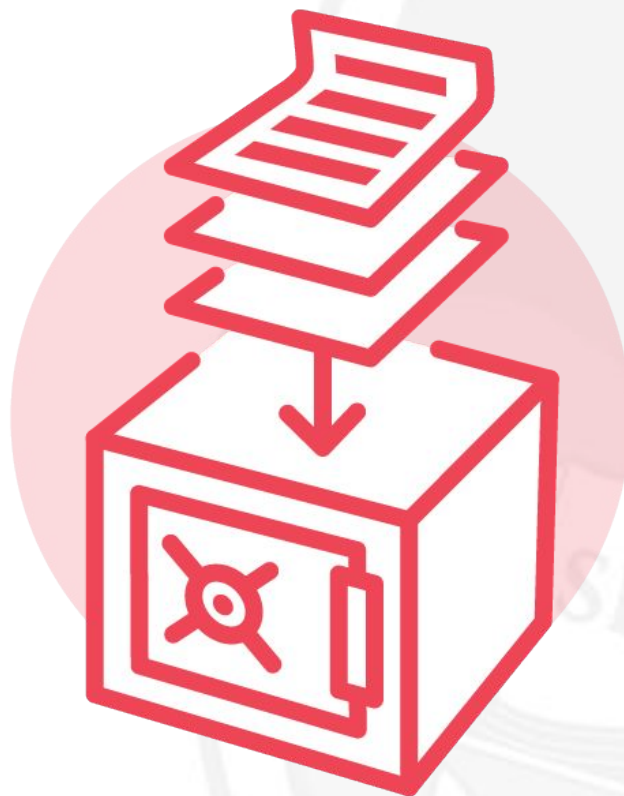
DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE



MANAGE, STORE, PRESERVE



arkivum
Bringing archived data to life



DIGITAL LIBRARY
SERVICES

Manage, Store, Preserve

Data needs to be **findable**, **accessible**, **interoperable** and **reusable** (FAIR), which is largely a question of machine-readable metadata. Only well-curated data can become a meaningful future resource.

- ★ create metadata and documentation
- ★ migrate data to best format
- ★ migrate data to suitable medium
- ★ back-up and store data
- ★ archive data



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH

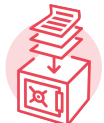


MANAGE, STORE, PRESERVE

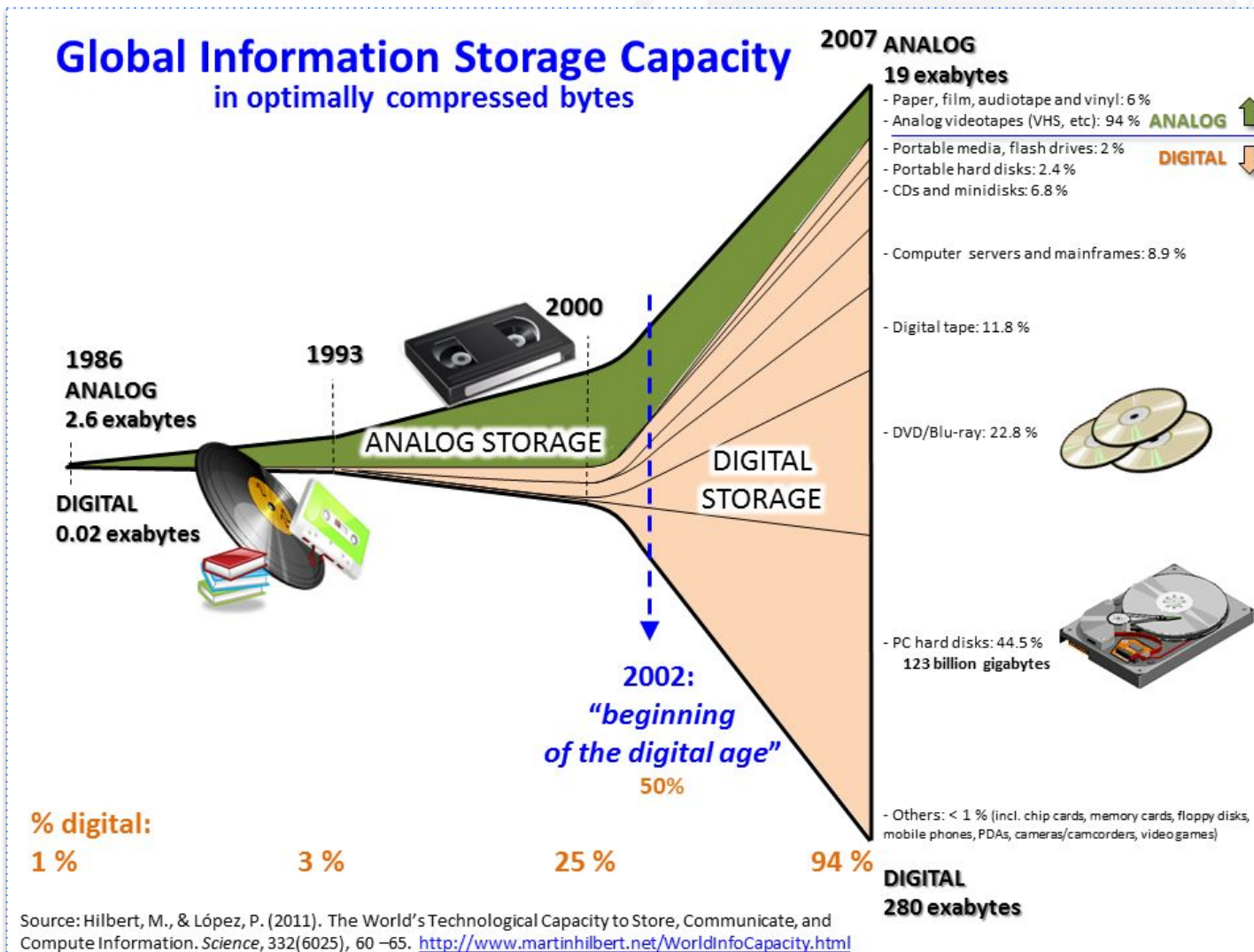
What do we mean by Digital Preservation (DP)?

- Digital Preservation is concerned with providing long-term access to digital objects, preserving continuity in form as well as functionality.
- It is not simply a backup of data, because long-term digital preservation must consider format, software and hardware obsolescence, among other issues.
- Although it is possible for anyone to read a page from a book written 100 years ago, the same is not true of (e.g.) a floppy disk containing WordPerfect files from twenty years ago.

Source: Preservica: A Guide to Making the Case for Digital Preservation. Presentation. 2014. (Online), Available: <https://preservica.com/uploads/legacy/2014/04/A-Guide-to-Making-the-Business-Case-for-Digital-Preservation-2014.pdf>



A brief history of the 'data deluge'



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).



World Digital
Preservation Day
7 November 2019



DIGITAL LIBRARY
SERVICES

Drivers for Digital Preservation

- *Business Owners*
- *Value Propositions*
- *Policies & Frameworks*



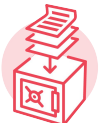
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019

Business owners

'[Digital Preservation] forms part of the integrated management of digital content and assets in such infrastructures [...]. In a business environment therefore, DP is the responsibility of data and asset management or legal compliance. In other words, it is driven by IT and legal departments, and is not part of the corporate mission as is the case for Memory Institutions (MIs). The common feature between MIs and Business Corporate Enterprises (BCEs) is the record-keeping aspect. BCEs are safekeeping their organization's records under the broad definitions of a Unified Information Management. [...] it can be recognized that in recent years, BCEs have taken on a more specific, new, but still evolving meaning that refers to the storage and preservation of the organization's digital information or knowledge assets, either for compliance or as a source of revenue.'

Source: Matthias L. Hemmje (2010) **Drivers for Digital Preservation**. (Online) Accessible: <https://ercim-news.ercim.eu/en80/special/drivers-for-digital-preservation>

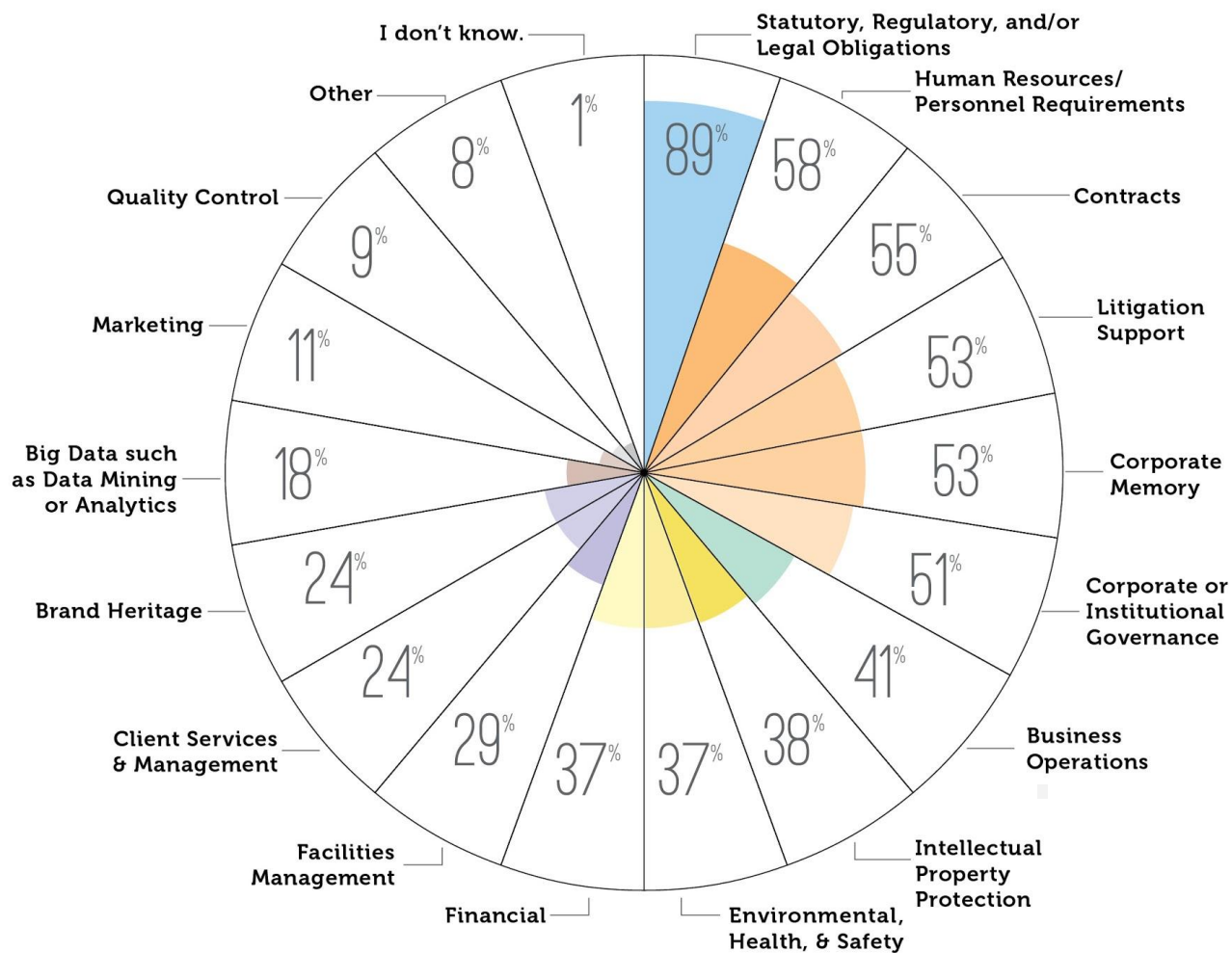


Value propositions

Business Driver	Benefit
Collection development	Organize digital content as an institutional asset
Corporate memory	Maintain accumulated knowledge of the institution
User access	Ensure access to specific users over long term
Information re-use	Enable re-purposing and added-value use of digital data
Reputational protection	Safeguard institution's standing in the community
Legal & regulatory compliance	Comply with freedom of information, privacy, financial, health & safety
Business continuity	Eliminate data loss leading to catastrophic business disruption
Efficiencies & savings	Support a streamlined information management strategy
Protecting investment	Safeguard against commercial loss of digital content
Supporting digital ways of working	Future-proof and enable online-only ways of working with staff, customers and partners

Source: Adrian Brown (2013) *Practical Digital Preservation: a how-to guide for organizations of any size*. See: <https://www.alastore.ala.org/content/practical-digital-preservation-how-guide-organizations-any-size>

Why do we keep (digital) information?



Source: (detail from) Digital Preservation Coalition: **Executive Guide on Digital Preservation: Facts and Figures**. (Online), Available: <https://dpconline.org/our-work/dpeg-home/dpeg-facts-and-figures>

Policy and strategy frameworks

- *International:*
[UNESCO Recommendation concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form](#) (2016); [G20 Anti-Corruption Open Data Principles](#) (2015); [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#) (2003); [The FAIR Guiding Principles for Scientific Data Management and Stewardship](#) (2016); [...]
- *National:*
[National Integrated ICT Policy White Paper](#) (2016); [National Archives and Records Service Digitisation Strategy](#) (2013); [SAHRA / National Heritage Resources: Legislations and Regulations](#) (1962 - 1999); [Intellectual Property Rights from Publicly Financed Research and Development Act](#) (2008); [...]
- *Institutional:*
[UCT Intellectual Property \(IP\) policy](#) (2011); [UCT Open Access \(OA\) policy](#) (2014); [UCT Research Data Management \(RDM\) policy](#) (2017); [...]
- *Departmental:*
UCT Libraries Digital Preservation Strategy (draft, 2019); [UCT Libraries 'Horizon 2019' strategic plan](#) (2014); [...]



World Digital
Preservation Day
7 November 2019



DIGITAL LIBRARY
SERVICES

Practicalities of Digital Preservation

- *Best Practices*
- *Tools & Systems*
- *Vendors*
- *Roles*
- *Activities*



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019

Digital Preservation **best practices**

U.S. National Archives (NARA) **Digital Preservation Framework** on Github

news!

The framework comprises an instrument for assessing file format risk, the results of NARA's assessment of over 350 file format variants, and draft preservation plans for those formats.

We are inviting comments and discussion on the plans using the Github Issues feature:

- What revisions can you suggest to the proposed processing and preservation actions for the formats?*
- Are the Essential Characteristics for each record type comprehensive enough for digital preservation?*
- Are the proposed preservation actions for the formats technically appropriate?*
- Are there appropriate tools for processing and preservation of specific formats that we do not have listed?*
- Are there other high priority formats that we need plans for?*
- What can you suggest in terms of appropriate public access versions of the formats?*

Source: <https://github.com/usnationalarchives/digital-preservation>

Digital Preservation tools & systems

UK National Archives > Digital preservation tools and systems

Your approach to Digital Preservation should be modular and flexible, to ensure it is sustainable. A combination of tools and technology that are currently available may be the most cost-effective means of achieving this flexibility:

- file format identification from The National Archives: [DROID](#) (freeware)
- file registry database from The National Archives: [PRONOM](#) (freeware)
- Community Owned Digital Preservation tool registry: [COPTR](#)
- SPRUCE project, including a Digital Preservation Business Case toolkit: [SPRUCE](#)

There are a growing number of options available on the market that address the digital preservation needs of archives. We have compiled a list of digital preservation solutions for archives, including commercial and open source. Product descriptions have been taken in part or in whole from the product web pages. For each option we have included information such as:

- the features of the product
- whether it is OAIS compliant
- whether it is open source or commercial
- whether it is software or a service
- whether there is easy to access documentation and an active user community (NB for open source products)
- whether storage is included
- what level (if any) of file encryption is offered

Source: <https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/preserving-digital-collections/digital-preservation-tools-systems/>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



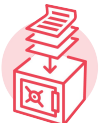
COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Digital Preservation vendors

An **extensive evaluation process** of digital preservation platforms was conducted by DLS for UCT Libraries, which included (amongst others), **Archivematica & AtoM** (Open Source as well as hosted versions) (Artefactual Systems), **Rosetta** (ExLibris), **Preservica** (Arkivum), and **Preservica**.







The evaluation compared **open source** and **licensed options**, and took into close consideration both **data sovereignty** (i.e. requirements for on-premise storage) and **local infrastructure support staffing costs**.

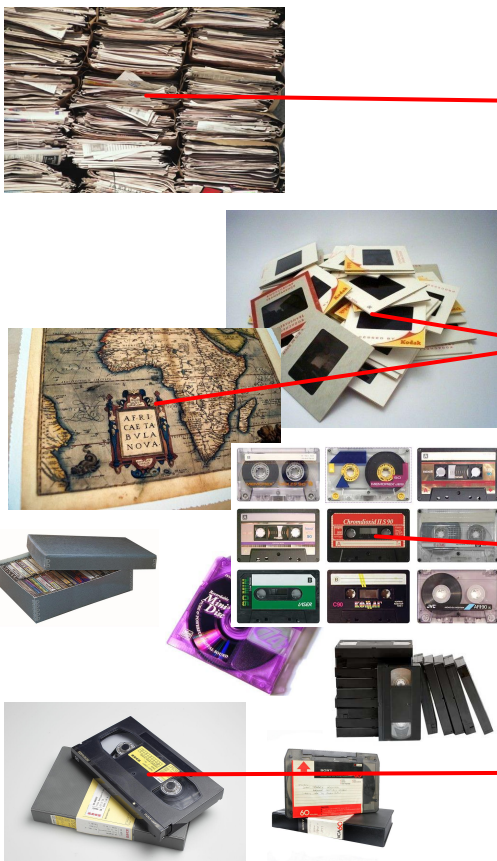
Digital Preservation roles

- **Budgets are limited** (austerity measures), and there is **not enough IT staff** at most SA HEIs (Faculties, Libraries, IT services) **to host independent, local solutions, including so-called ‘free’ Open Source** platforms.
- **Open Source requires ongoing local, specialist IT support**, including: needs analysis, liaison, implementation, customisation, development, upgrades, migration, etc.
- **Federated licensing** (TENET?) would be preferable to **individual institutional licenses**.
- Organisational structures require urgent review to embed new, **specialised skills**:
 - **Data Analysis, Mining, and Visualisation** (Digital Humanities; GIS; ...)
 - **Data Archiving** (File format conversion; Bit curation; DP; ...)
 - **Data Curation** (Data access, archiving and publishing; Metadata schemas; ...)
 - **Data Librarianship** (Data acquisition, citation, location and re-use; ...)
 - **Digital Scholarship** (Liaison; Open Scholarship; R&D; Digital Humanities; ...)
 - **Repository management** (Coding; Digital library infrastructure; Semantic web; ...)
 - **Research Data Management** (Funder mandates; Advocacy; Support services, ...)

Example: Digitisation for digital preservation

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

legacy formats



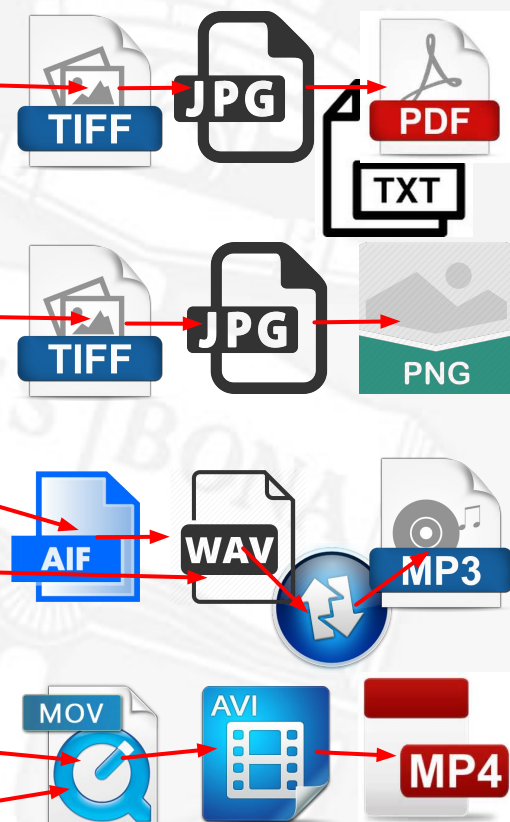
hardware



software



digital files



Example: Digitisation for digital preservation

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

'legacy' media	hardware	software	digital files		
			master (preservation)	service (working)	access (access)
<u>Documents</u> : manuscripts; theses; ...	flatbed scanner; feeder scanner; ...	Acrobat Pro;tif	.jpg	.jpg .txt .pdf
<u>Images</u> : photographic prints; positives (slides); negatives; maps; ...	virtual drum scanner; digital camera & lighting equipment; map scanner, ...	Silverfast Studio; Nextimage; Photoshop; Lightroom;tif .fff .dng	.jpg	.jpg .png
<u>Audio</u> : ¼-inch reel-to-reel; cassette; DAT; MD; ...	reel-to-reel, cassette, DAT and MD recorders; DAC; mixer; ...	Logic Pro; Waves Restoration Suite plugins; MediaHuman Audio converter;aif	.wav	.mp3
<u>Video</u> : Umatic; Betacam; VHS; MiniDV; ...	Umatic, Betacam and VHS cleaners and recorders; MiniDV, DVCam descks; ...	MediaExpress; FinalCut Studio; Premiere Pro;mov .mpg2	.avi	.mp4





arkivum
Bringing archived data to life

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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE



arkivum
Bringing archived data to life



DIGITAL LIBRARY
SERVICES

Digital preservation technologies & processes

File format
normalisation

ESCROW*

Migration
paths

Multiple
copies

Automatic
metadata
capture

Fixity
checking
& virus
scans

Search
platform

Data
under
managem
ent

* **'Source code escrow'** is the deposit of the [source code](#) of [software](#) with a third-party [escrow](#) agent. Escrow is typically requested by a party licensing software (the licensee), to ensure maintenance of the software instead of [abandonment](#) or [orphaning](#).⁹ Online. Available: https://en.wikipedia.org/wiki/Source_code_escrow

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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE



arkivum
Bringing archived data to life



DIGITAL LIBRARY
SERVICES

Planning for our own Data Management

Plan

Process

Enhance

Preserve

Publish

Reuse

Proposal

MoU

DMP
(proposal)

"We **will do** XYZ
to manage our
data"

resourcing

DMP
(active)

"We **are doing**
XYZ to manage
our data"

documentation /
review

DMP
(final)

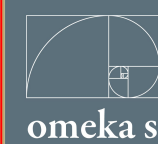
"We **have done**
XYZ to manage
our data"



UCT DMP



arkivum
Bringing archived data to life



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE

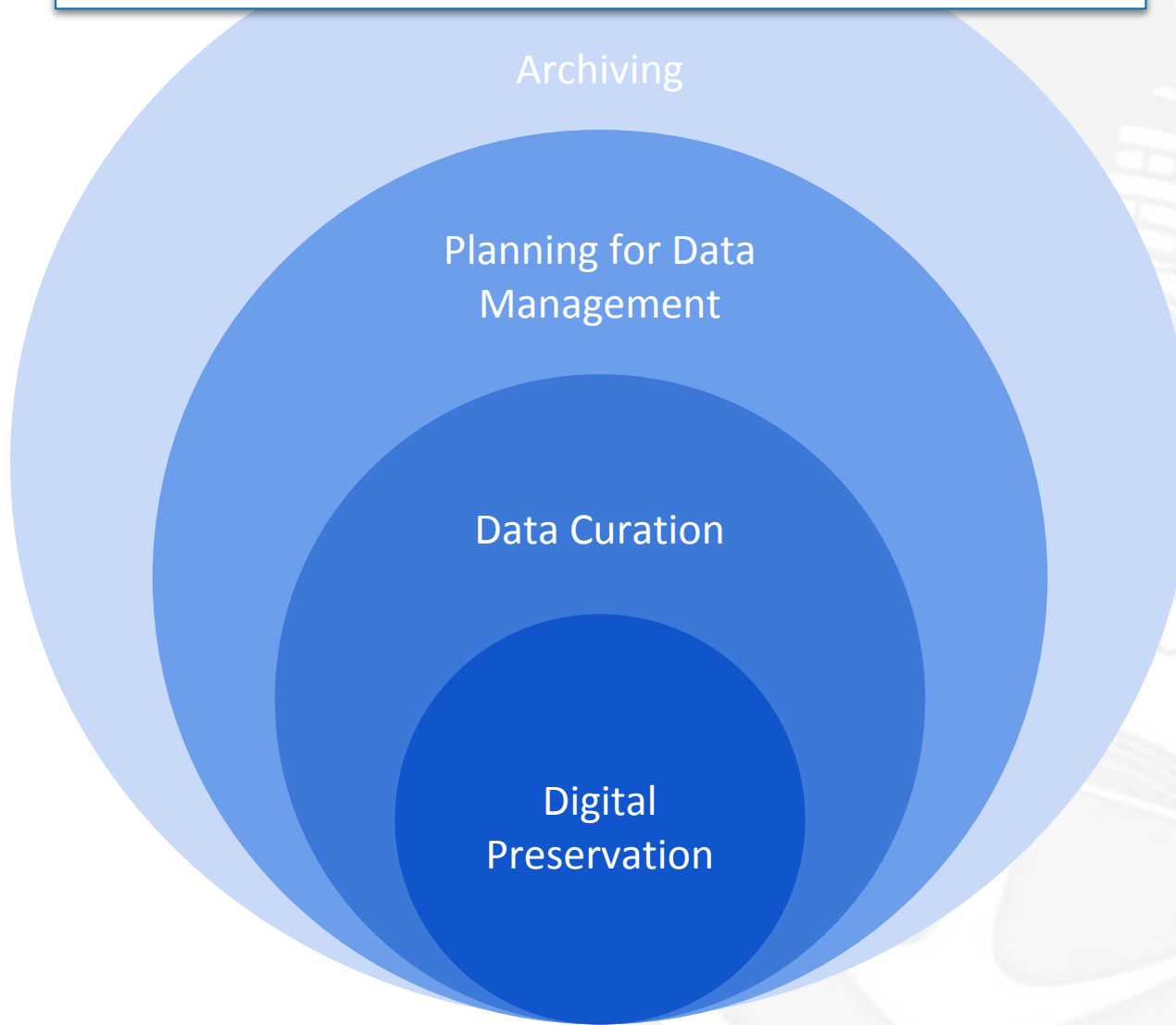


SHARE & PUBLISH



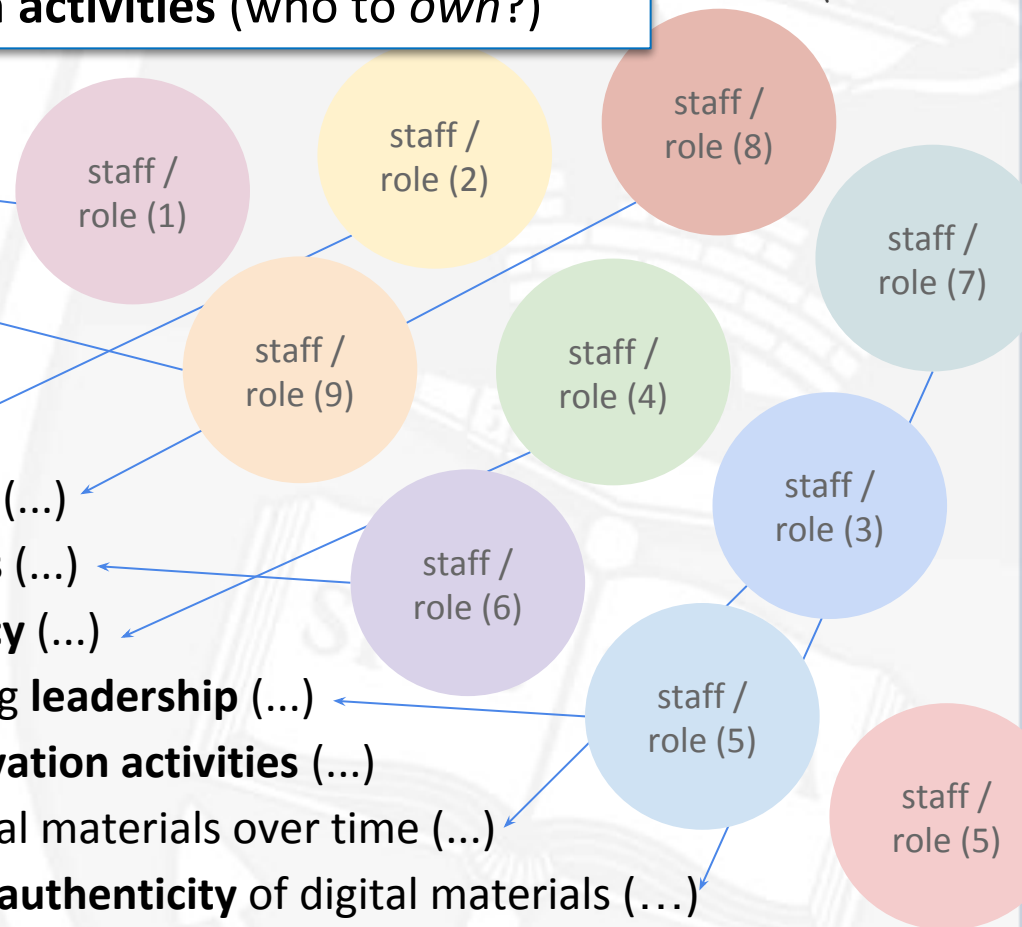
MANAGE, STORE, PRESERVE

Archiving > DMP >> Curation >>> Digital Preservation



Digital Preservation **activities** (who to *own*?)

- Capture **metadata** (...)
- **Liaise** with **stakeholders** (...)
- Use appropriate **standards** (...)
- Provide appropriate **access** (...)
- Carefully **appraise** and **select** (...)
- Provide supporting **documentation** (...)
- Keep up with technological **changes** (...)
- Plan and develop **strategy** and **policy** (...)
- **Work together** with strong, enabling **leadership** (...)
- Assign appropriate **levels of preservation activities** (...)
- **Add value** to an organization's digital materials over time (...)
- Ensure the continued **integrity** and **authenticity** of digital materials (...)
- Actively monitor, plan, and manage digital materials, systems and workflows (...)
- Help make digital preservation be '**business as usual**' across your organisation (...)



Adapted from: Digital Preservation Coalition: **Executive Guide on Digital Preservation for all organizations: All organisations.** (Online), Available: <https://dpconline.org/our-work/dpeg-home/dpeg-organisation-type/dpeg-all-orgs>



World Digital Preservation Day

7 November 2019



DIGITAL LIBRARY
SERVICES

Digital Preservation: an all-inclusive conversation

- *DPC | World Digital Preservation Day (07.11.2019)*
- *Australasia Preserves | A digital preservation CoP*
- *NeDICC | Network of Data and Information Curation Communities*
- *RDM at UCT Slack Workspace & DLS Training*



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019

Digital Preservation Coalition

Strategic Plan 2018 - 2022

- **Community Engagement:** *enabling a growing number of agencies and individuals in all sectors and in all countries to participate in a dynamic and mutually supportive digital preservation community.*
- **Advocacy:** *campaigning for a political and institutional climate more responsive and better informed about the digital preservation challenge; raising awareness about the new opportunities that resilient digital assets create.*
- **Workforce Development:** *providing opportunities for our members to acquire, develop and retain competent and responsive workforces that are ready to address the challenges of digital preservation.*
- **Capacity Building:** *supporting and assuring our members in the delivery and maintenance of high quality and sustainable digital preservation services through knowledge exchange, technology watch, research and development.*
- **Good Practice and Standards:** *identifying and developing good practice and standards that make digital preservation achievable, supporting efforts to ensure services are tightly matched to shifting requirements.*
- **Management and Governance:** *ensuring the DPC is a sustainable, competent organization focussed on member needs, providing a robust and trusted platform for collaboration within and beyond the Coalition.*



World Digital
Preservation Day
7 November 2019



DIGITAL LIBRARY
SERVICES

World Digital Preservation Day

Home > Events > World Digital Preservation Day

World Digital Preservation Day

7 November 2019 00:00 - 24:00 Worldwide

Advocacy



Dünya Sayısal Koruma Günü

At-Risk Digital Materials

About World Digital Preservation Day

World Digital Preservation Day is held on the first Thursday of every November. This year we will celebrate all things digital preservation on 7th November 2019!

#WDPD2019 
n Twitter

**Tim
Gollins @timgollins**

RT @Sarah_DPC: With the
INTERNATIONAL Digital
Preservation Conference
#iPRES2019 in Amsterdam this
week ... what better time to
invite #dig...

VIEW TWEET

  Sep 20

**Tim
Gollins @timgollins**



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019



Australasia Preserves

A digital preservation community of practice

AUSTRALASIA PRESERVES

A Digital Preservation Community Of Practice

[HOME](#) [ABOUT](#) [CONTACT](#) [CODE OF CONDUCT](#) [JOIN](#) [BLOG POSTS](#)



Australasia Preserves is a digital preservation community of practice (CoP) for Australia, New Zealand, and the broader Australasian region.



A **community of practice** is a group of people who share a craft or a profession



Being part of a community of practice provides opportunities to develop personally and professionally



The **Australasia Preserves community of practice** is a group of people who share an interest in the field of digital preservation



Australasia Preserves helps build local and global relationships, enabling sharing of digital preservation knowledge and experience



‘South Africa Preserves’ [?]

See: ***Australasia Preserves - a digital preservation community of practice***

***Australasia Preserves** is a digital preservation community of practice for the Australasian region, with members from a wide range of domains across Australia and New Zealand. Community activities cover varied interests and areas of working with digitised and born-digital materials, encouraging participation and input from a wide variety of people and organisations, including libraries, archives, records, IT, government, and universities.*

[...] This community of practitioners, managers, educators, students, and enthusiasts enables sharing of digital preservation knowledge and expertise, and fosters opportunities to develop personally and professionally through an active online forum, monthly virtual meet-ups, locally organised events and networking opportunities, and working groups.

The community welcomes anyone interested in learning and sharing good practice for ongoing care of and access for digital materials to get involved, to connect, and to collaborate, in order to enhance digital preservation capability for a wide range of people and organisations in the Australasian region.

- 2018 highlights for the community: <https://bit.ly/2FJrSUB>
- The Australasia Preserves 2019 Briefing Pack: <https://bit.ly/2GM82dD>
- Join the online forum: <https://groups.google.com/forum/#!forum/australasia-preserves>

Source: ‘*Australasia preserves*’ (2019). (Online) Accessible: <https://www.australasiapreserves.org/p/about.html>



NeDICC

Network of Data and Information Curation Communities

NeDICC aims to promote the development and use of research data and information curation standards and practices to ensure the long term preservation and accessibility of digital research outputs in support of e-Research.

Specific aims and functions of the Network include:

- The provision of a forum, for practitioners and managers involved in digital object management practices, to exchange experience and express alternative views.
- Activities aimed at promoting communication and co-operation between members of the Network include meetings, seminars, workshops and conferences to:
 - address issues of interest/concern.
 - expose the community to new developments and trends, provide opportunities to engage with a wider audience, as well as showcase work and initiatives.
 - develop the knowledge and skills of members.
 - promote awareness and best practices relating to digital preservation, dissemination and use of research outputs.
 - collaborate on projects in support of shared objectives.

Source: <https://nedicc.com/about-nedicc/>



Closing remarks, Upcoming workshops & Staying in touch with DLS



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**: Findable, Accessible, Interoperable and Reusable).

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

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

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>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD



PLAN & DESIGN



COLLECT & CAPTURE



COLLABORATE & ANALYSE



DISCOVER, REUSE & CITE



SHARE & PUBLISH



MANAGE, STORE, PRESERVE

Upcoming Workshops

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

Data Drop-In: Deposit your Data with ZivaHub as part of Open Access Week
16 Oct @ 10 AM



Open Access Week 2019: 21 - 27 October

To help promote open access to data as part of this commemoration, Figshare is running a data upload competition during the month of October. As you might be aware, **ZivaHub**, UCT's open access data repository, runs on Figshare for Institutions software. As custodians of the data on this platform, we are keen to participate and win with your help.

DATA DROP-IN

For our **ZivaHub session on the 9th of October**, We are offering an exclusive bring and share 'data drop in' session to researchers at UCT who want to publish their data openly.

Each uploader will receive Figshare swag (enamel pins, stickers and more!). The competition runs throughout the month of October. To qualify, all data must be submitted to ZivaHub between the **1st of October 2019** and **31st October 2019** to count towards the final numbers.

**COME AND DROP IN YOUR DATA
AND STAND A CHANCE TO
WIN!!!!!!**



DIGITAL LIBRARY
SERVICES



'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 – public and private – as well as genuine equal partnership with stakeholders in all so

<https://content.iospress.com/articles/inform>

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 DMP

OneDrive / Google Drive etc.

UCT Open Science Framework (OSF)

Digital preservation

ZivaHub | Open Data UCT





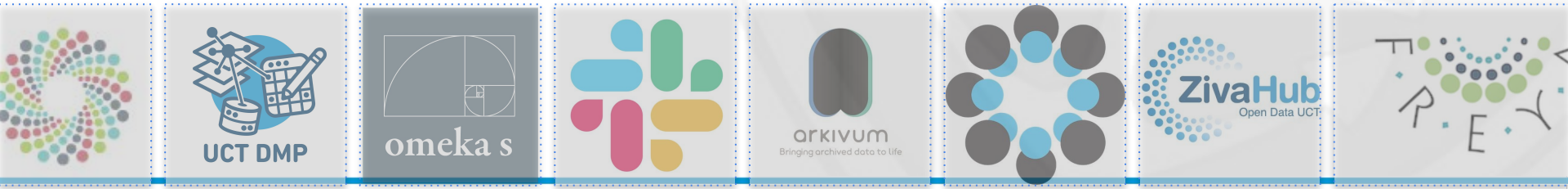
**World Digital
Preservation Day**
7 November 2019

Let's change
what we value
in research.



DIGITAL LIBRARY
SERVICES

THANK YOU! :)



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019



World Digital Preservation Day

7 November 2019

Let's change
what we value
in research.



Sign
DORA



DIGITAL LIBRARY
SERVICES



dls@uct.ac.za



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



rdm-at-uct.slack.com

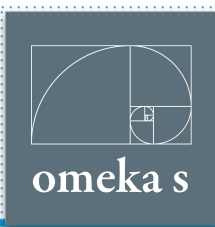


DIGITAL LIBRARY
SERVICES

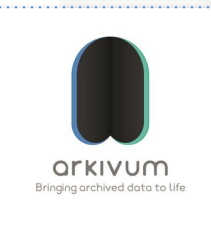
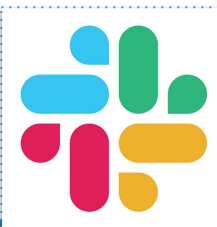
<http://www.digitalservices.lib.uct.ac.za/>



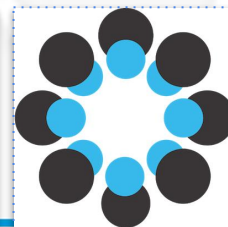
UCT DMP



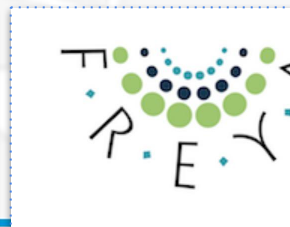
omeka s



arkivum
Bringing archived data to life



ZivaHub
Open Data UCT



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA - UNIVERSITEIT VAN KAAPSTAD

Wednesday, 16th October 2019