Planning for FAIR data

Joy Davidson, DCC

With contributions from DCC colleagues

What is Digital Curation Centre?

a centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management and open science

DCC

www.dcc.ac.uk

Training | Events | Tools | Advocacy | Consultancy | Guidance | Publications | Projects

Learning objectives

This session will introduce participants to research data management, open and FAIR data and walk you through the kinds of information that should be considered when developing a data management plan. You'll also see a demonstration of DMPTool and have a chance to give it a try.

After this session, participants will:

- Understand the difference between FAIR and Open Data
- Be aware of what should be included in a DMP
- Know how to use fee tools to help you write DMPs

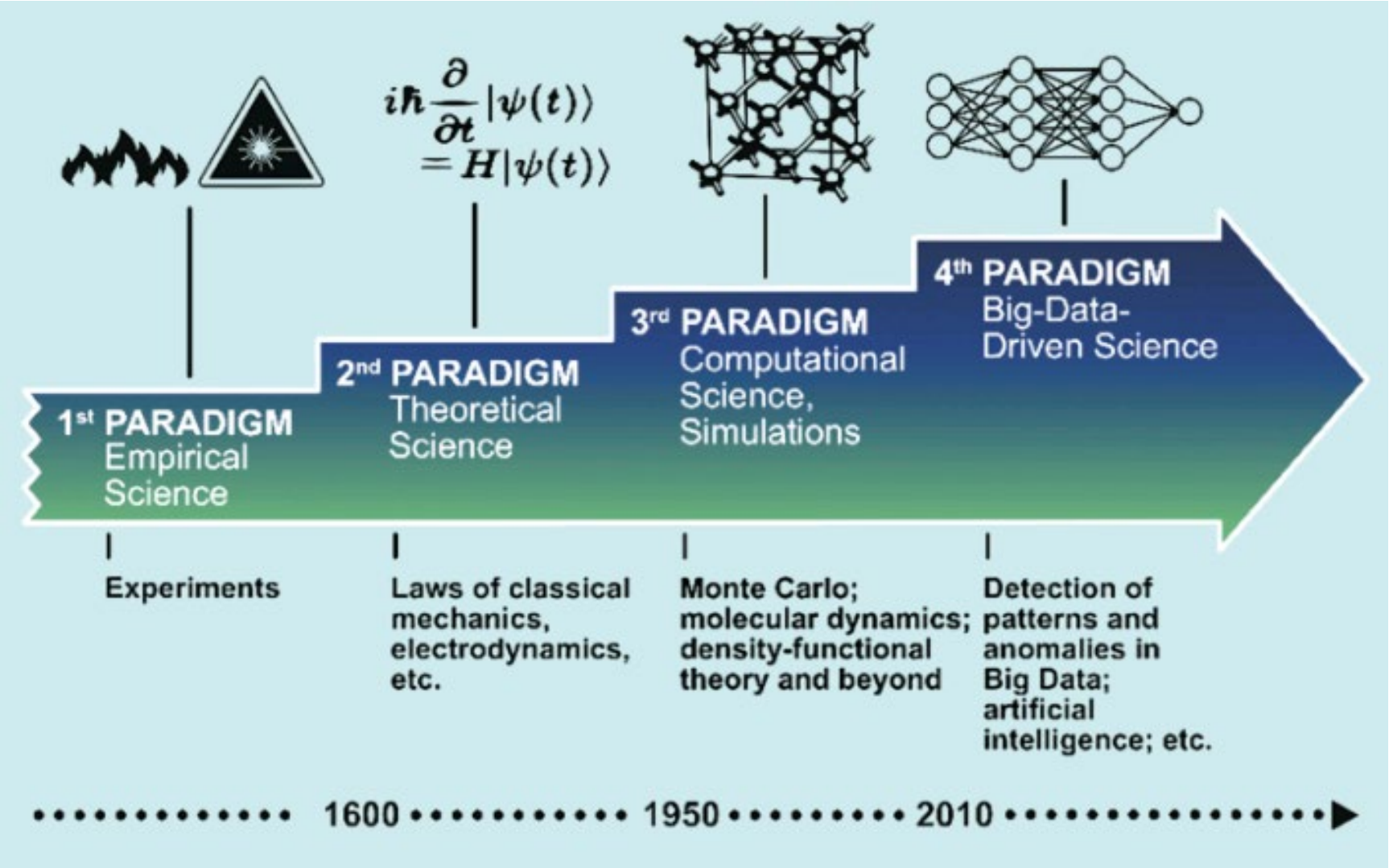
Agenda – two part workshop

- Introduce RDM, open and FAIR data and data management planning
- Review a sample data management plan (DMP)
- Coffee break
- Demonstration of DMPTool - a data management planning tool
- DMPTool hands-on session

But first, a bit about you!

Please go to menti.com and enter code 8758 0867

Data driven research needs good data!



The Four Research Paradigms of Material Sciences
Draxl C., Scheffler M. (2020) Big Data-Driven Materials Science and Its FAIR Data Infrastructure. In: Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer, Cham. https://doi.org/10.1007/978-3-319-44677-6_104

What is Research Data Management?

- Data Management Planning
- Creating data
- Documenting data
- Accessing / using data
- Storage and backup
- Selecting what to keep
- Sharing data
- Data licensing and citation
- Preserving data

Funders have expectations about DMPs

UGC 大學教育資助委員會 University Grants Committee

Visitor | Text Size 繁 簡

The RGC | Funding Opportunities | Guidelines & Forms | Highlights of Achievements | Statistics | The Ins and Outs

Home > The RGC > Publications

Publications

Annual Reports

Major Reports

- Consultancy Report of the Review of Research Grants Council Documents
- Open Access Plan of the Research Grants Council (6.1.2021)
- Review Report of the Working Group on the Review of the Research Grants
- Interim Report of the Review of the Research Grants Council (Phase II) (4.
- Review Report of the Task Force on Review of Research Policy and Funding
- Interim Report for Consultation by Task Force on Review of Research Policy
- Review of the Research Grants Council (Phase I) (21.9.2017)

4. Phase II: Target Enforcement and Monitoring

4.1 Progressing to Phase II (from February 2024 to October 2026), with the preparatory work carried out in Phase I, the main theme would be target enforcement and monitoring. Subject to a review on the effectiveness of Phase I initiatives as well as further consultation with the universities, progressive percentage targets for peer-reviewed journal articles arising from RGC-funded projects to be publicly available would be set, with funding support on open access provided for the universities in parallel. As far as data management is concerned, DMPs would be a mandatory requirement for applications for the RGC's research funding schemes. Subject to the readiness of the research sector, the RGC would also consider launching a pilot scheme for data sharing at the end of this Phase.

Open Access Plan of the Research Grants Council

https://www.ugc.edu.hk/eng/rgc/about/publications/

FAIR principles

Principles apply equally to data and associated metadata (i.e., the description that provides context about the data).

Findable

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

Accessible

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

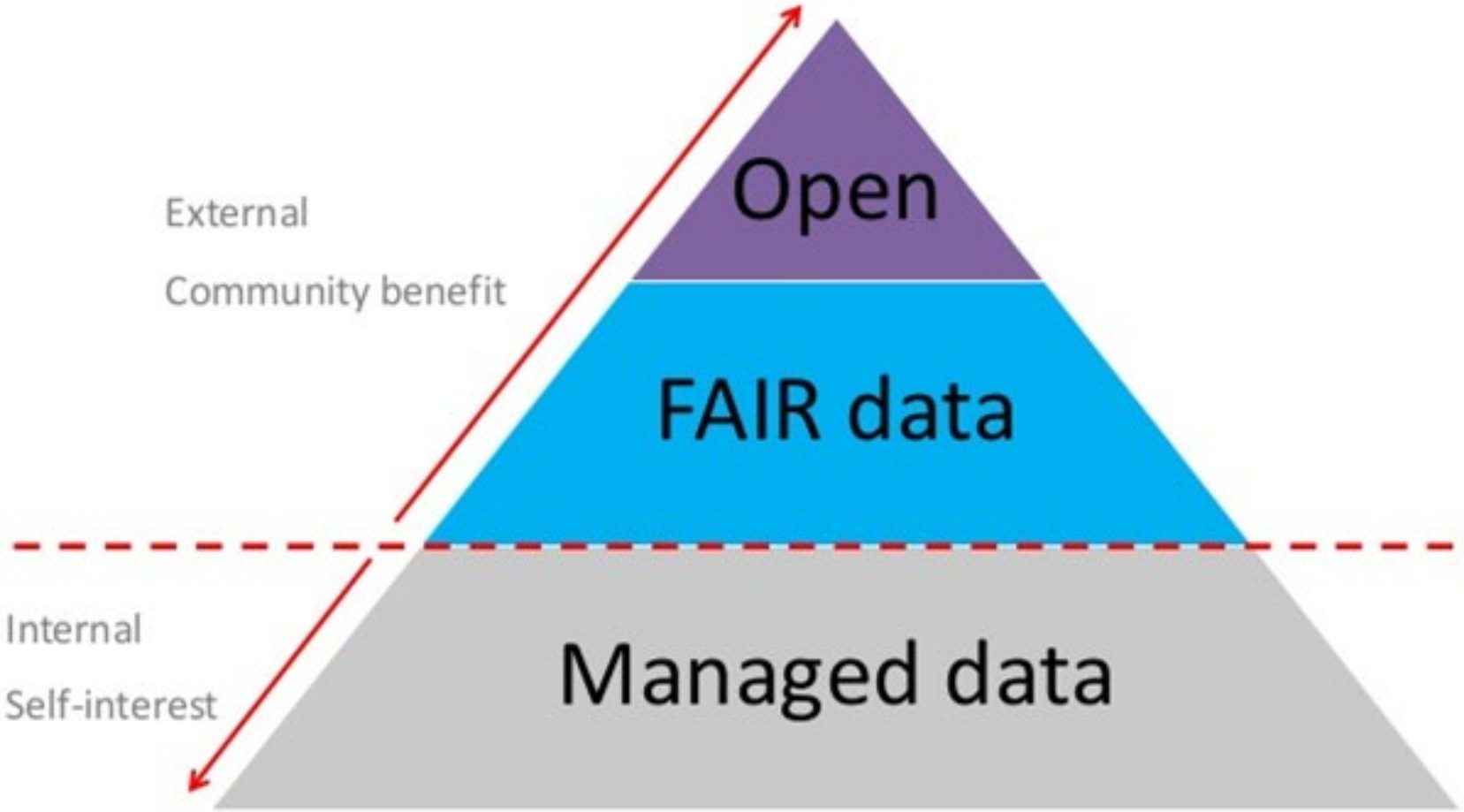
Interoperable

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

Reusable

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

How do Open, FAIR & RDM intersect?



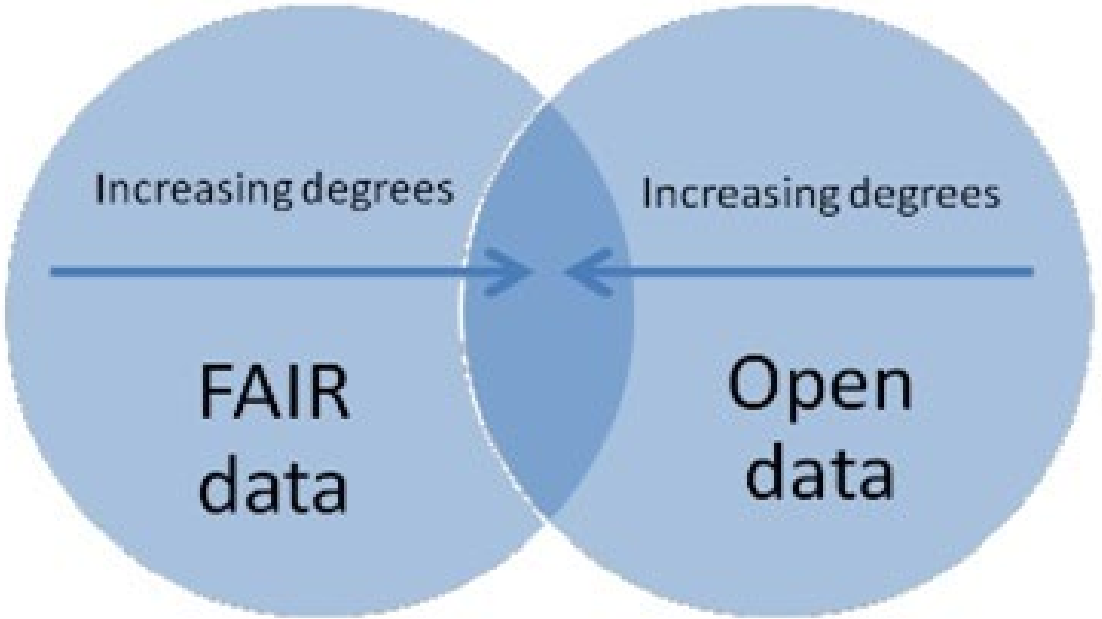
Slide from ‘What it means to be FAIR’, Sarah Jones https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?

FAIR and Open

Concepts of FAIR and Open should not be conflated.

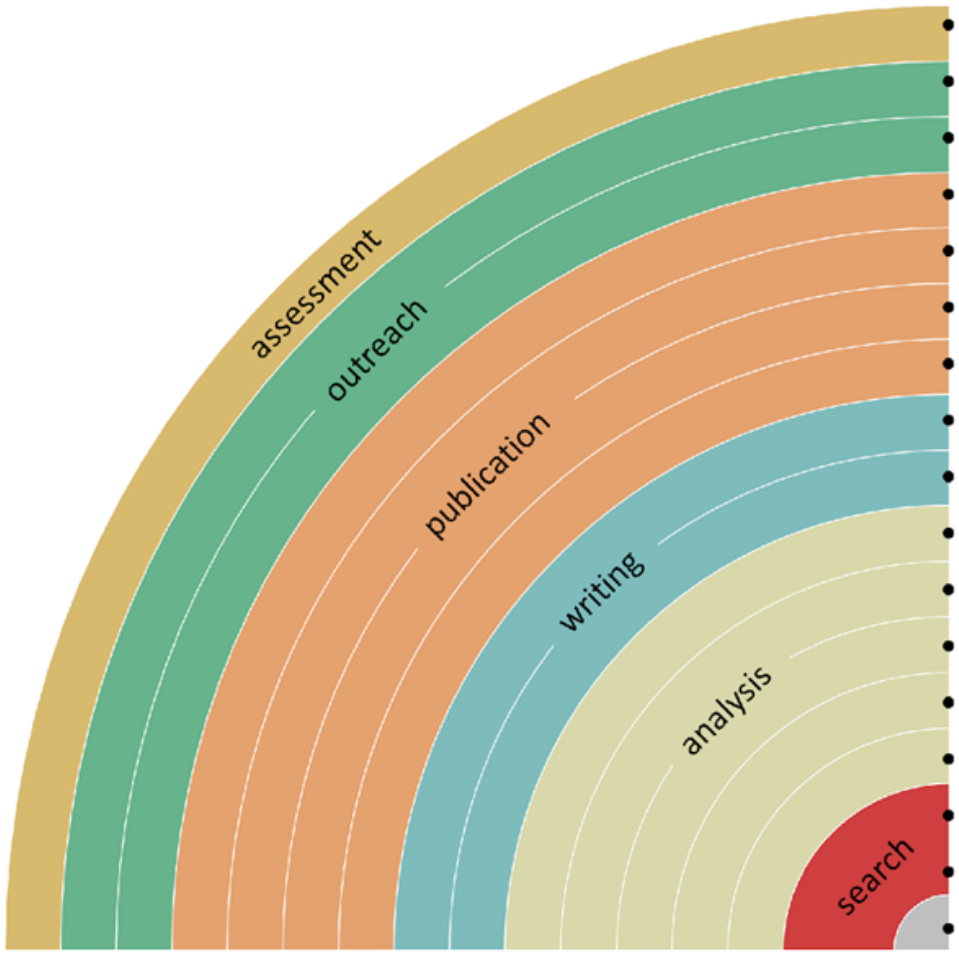
Data can be FAIR or Open, both or neither

- The greatest potential reuse comes when data are both FAIR and Open
- Align and harmonise FAIR and Open data policy



Slide from ‘What it means to be FAIR’, Sarah Jones https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?

You can make your workflow more open by …

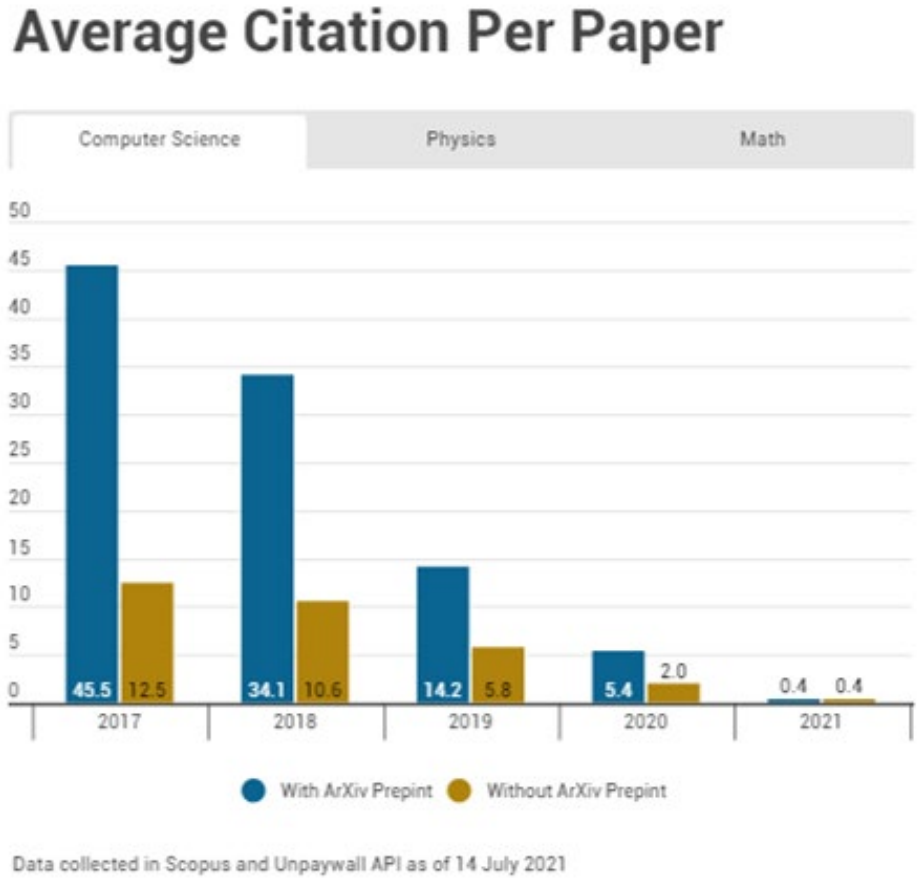


- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing 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

DOI: 10.5281/zenodo.1147025

Good RDM and sharing can lead to…

- More visible research and increased impact
- Easier outputs reporting
- More opportunities to collaborate
- Better and more reproducible research!

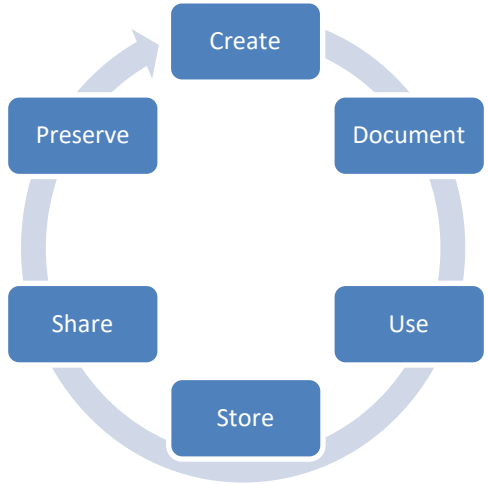


ArXiv Preprints by HKUST and Their Citation Advantage
https://library.ust.hk/sc/arxiv-hkust/

Data Management Plans (DMPs)

What should be addressed in a DMP?

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation



A DMP is a plan to share.

Remember - you could be sharing with your future self!

Consider: where will you store the data during your research?

- Your own laptop
- University systems
- Cloud storage
- Combination

Your decision will be based on how sensitive your data are, how robust you need the storage to be, who needs access to the data, and when they need access to the data!

Consider: How will you name your files?

- Keep it simple!
- Order the elements logically
- Include dates and times is necessary
- Avoid special characters
- Use hyphens or underscores not spaces
- Make sure you agree approach with your research partners

Workshop_report_200820_final.doc

Workshop_report_200820_final_jdedits.doc

Workshop_report_200820_final_FINAL.doc

Workshop_report_200820_final_FINAL_210820. doc

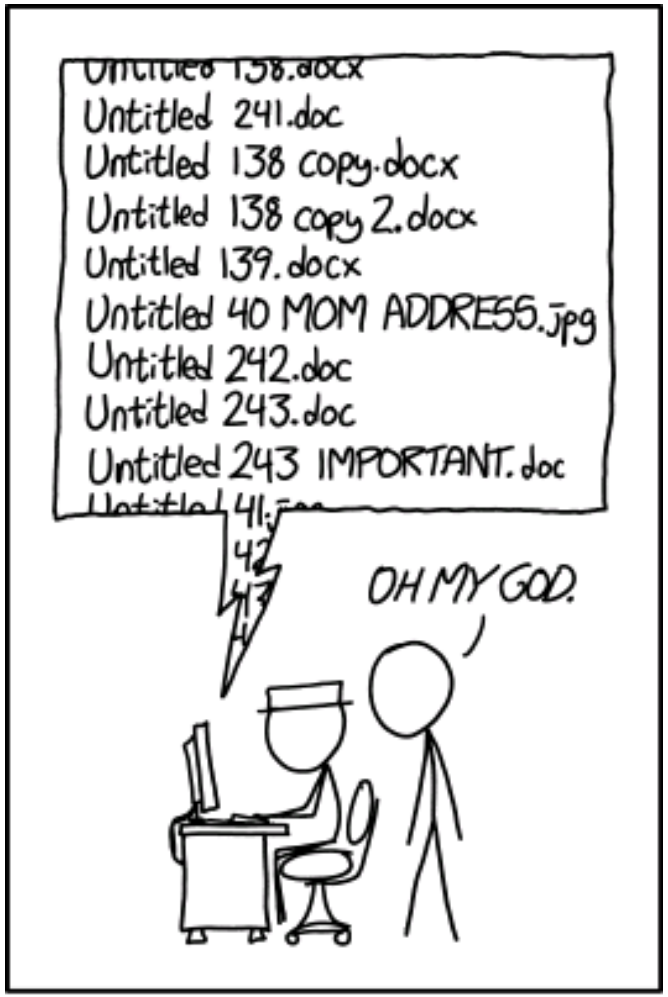
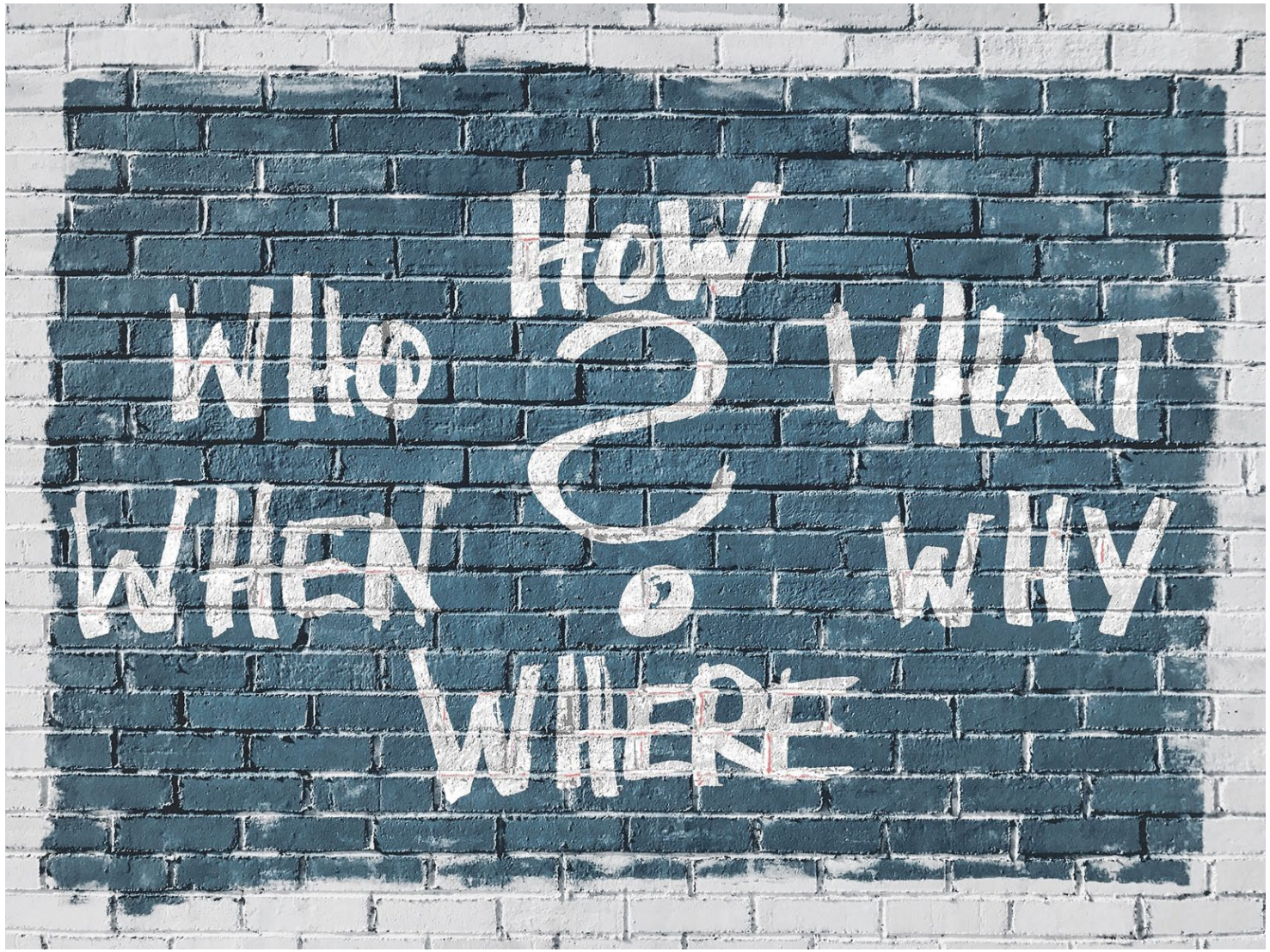


Image source: http://sxkcd.com/1459/

Consider:
How will you describe your data to provide context – for you and for others?



Consider: What documentation is needed?

Documentation

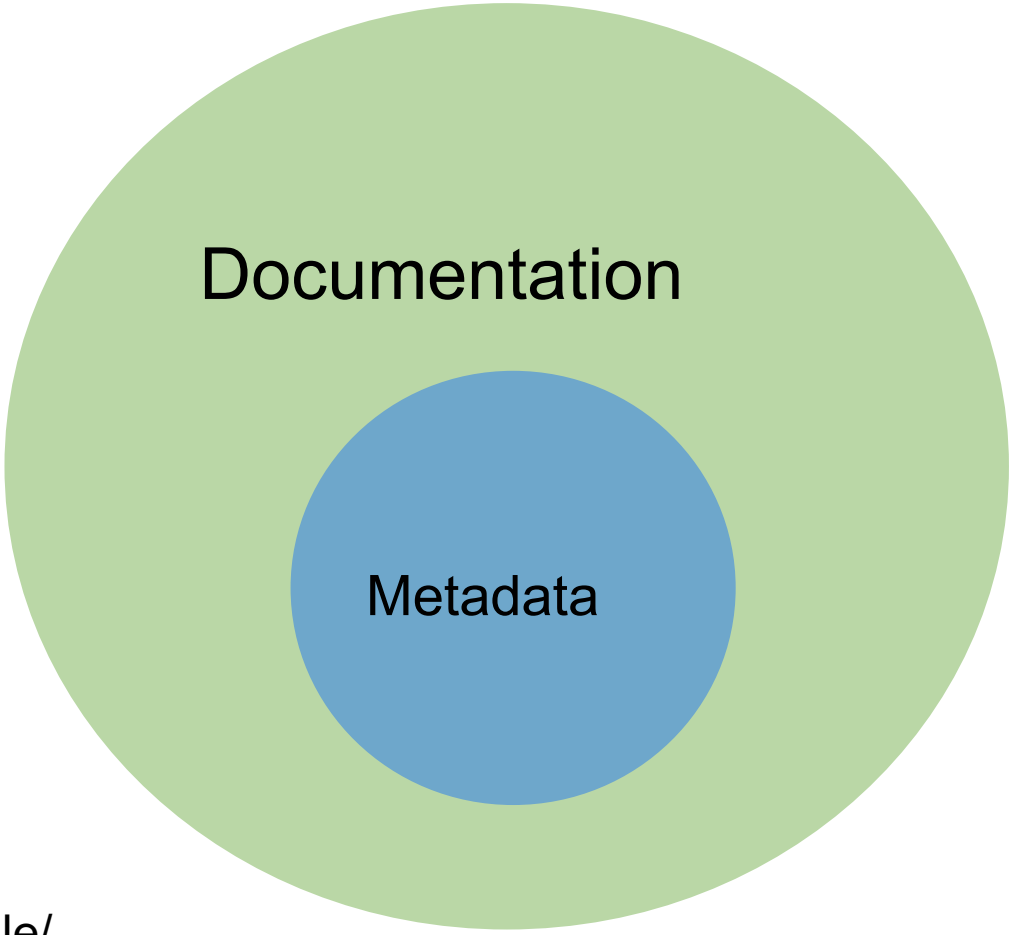
Think about what is needed in order to evaluate, understand, and reuse the data.

- Why was the data created?
- Have you documented what you did and how?
- Did you develop code to run analyses? If so, this should be kept and shared too.
- Important to provide wider context for trust

Slide from ‘An Introduction to Research Data Management, FAIR and Open Data’, S. Venkataraman.
https://drive.google.com/drive/folders/1_MXFhrzKVuKjoytVf7wh5Pndp-BAWAA1

Descriptions should humans and machine readable

- Metadata
  - Standardised
  - Structured
  - **Machine** and human readable



Find useful tips in ‘Your Dataset Deserves Good Documentation’ https://library.ust.hk/sc/readme-file/

Consider: How much metadata will you provide?

- DataCite metadata
- Citation/disambiguation
  - Identifier e.g. DOI
  - Creator
  - Title
  - Publisher
  - Publication Year
- Licencing/access conditions

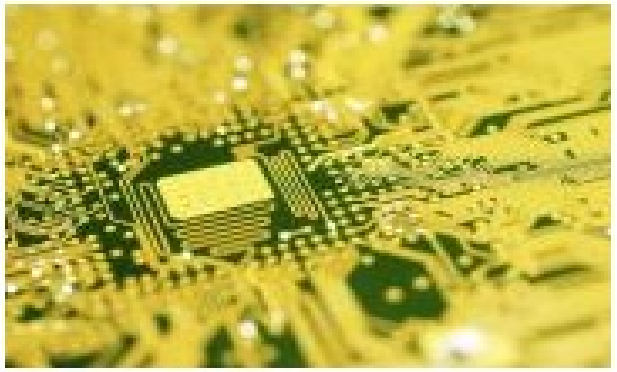
Why should you go beyond the minimum required?

Consider: Using discipline specific metadata standards

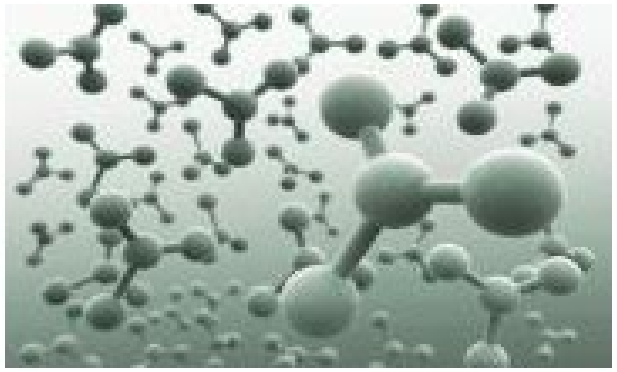
Search by Discipline

Biology

Earth Science



General Research Data



Physical Science

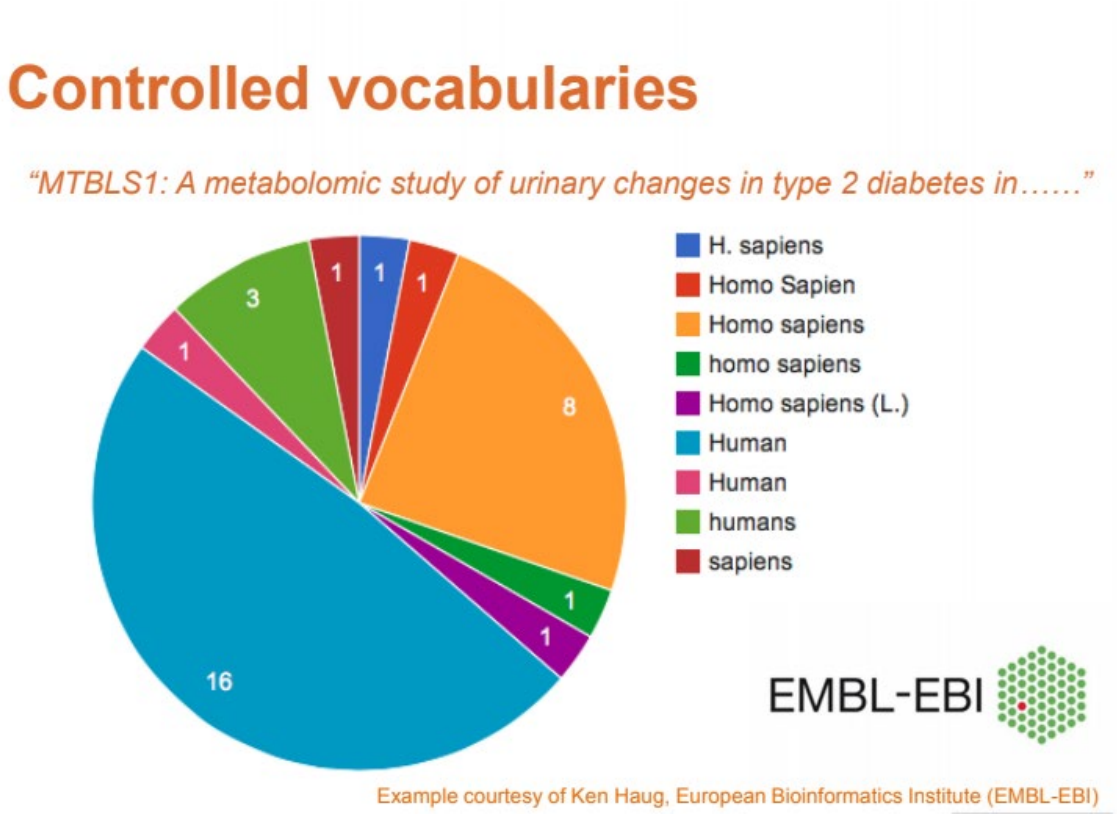
Social Science & Humanities

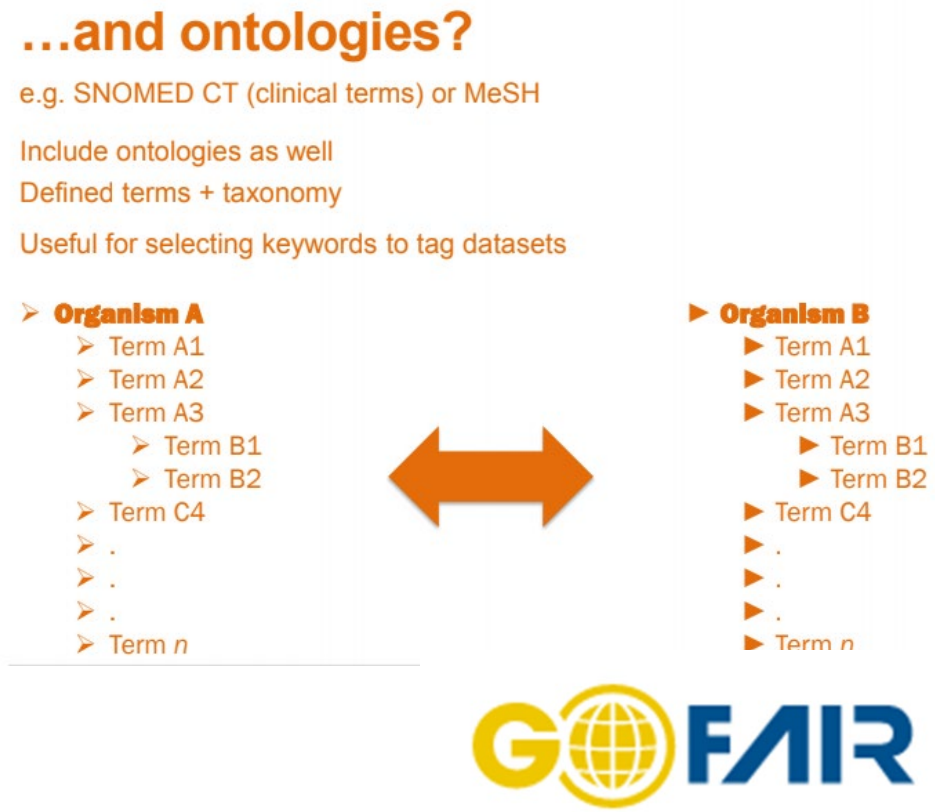
RDA
RESEARCH DATA ALLIANCE

http://rd-alliance.github.io/metadata-directory/

Consider: Using controlled vocabularies

Ontologies and thesauri support interoperability





Slides from 'An Introduction to Research Data Management, FAIR and Open Data', S. Venkataraman. https://drive.google.com/drive/folders/1_MXFhrzKVuKjoytVf7wh5Pndp-BAWAA1

https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/

Consider: Where will you store your data after your project?

Preferred repositories:

1. Domain specific

2. Institutional (DataSpace@HKUST)

3. Generalist (Zenodo, figshare)

Try to choose a FAIR aligned repository

DataSpace@HKUST

DataSpace@HKUST

HOW CAN DATASPACE@HKUST HELP?

The service is supported by a special team, which consists of professional librarians with various expertise. Members of the DataSpace Team will assist you in depositing and publishing datasets; they will also create standardized metadata to describe your datasets and tidy up any formatting issues with you.

Once a dataset is published in DataSpace@HKUST, it will receive a permanent Digital Object Identifier (DOI) that makes the dataset easily discoverable and persistently accessible. Meanwhile, a ready-made standard citation is created for users to cite your work, which can lead to more accurate citation count and help you get the proper credit.

https://dataspace.ust.hk/

- **Metadata support**
- **Formatting support**
- **DOI to support discovery and access**
- **Ready made citation to get you more credit for your work**

Finding a suitable domain specific repository

https://www.re3data.org/

University of Glasgow - MyGlasg | Google Calendar - October 2021 | KHUST training - Google Drive | Planning for FAIR data_HKUST_0 | Search | re3data.org

re3data.org/search?query=&countries[]=HKG

Apps | Reading list

re3data.org

Search | Browse | Suggest | Resources | Contact | DataCite

Countries

China (1)

Hong Kong (3)

AID systems

API

Data access

Data access restrictions

Database access

Data licenses

Data upload

Data upload restrictions

Enhanced publication

Institution responsibility type

Institution type

Keywords

Metadata standards

PID systems

Provider types

Quality management

Repository languages

Software

Syndications

Repository types

Versioning

Found 3 result(s)

DataSpace@HKUST

Subject(s): Natural Sciences | Life Sciences | Fine Arts, Music, Theatre and Media Studies | Humanities | Social and Behavioural Sciences | Social Sciences | Humanities and Social Sciences

Content type(s): Standard office documents | Images | Databases | Standard office documents | Scientific and statistical data formats | Plain text

Country: Hong Kong

Created and managed by the Library, DataSpace@HKUST is the data repository and workspace service for postgraduate students can use the platform to store, share, organize, preserve and publish research data. It Harvard's Institute for Quantitative Social Science. Using Dataverse architecture, the repository hosts multipl may contain multiple data files and the corresponding descriptive metadata.

DataHub Figshare

University of Hong Kong Data Repository

Subject(s): Humanities | Biology | Social and Behavioural Sciences | Engineering Sc | Life Sciences | Social Sciences | Education Sciences | Computer Scienc

Content type(s): Source code | Software applications | Standard office documents | Struc | Plain text | Images

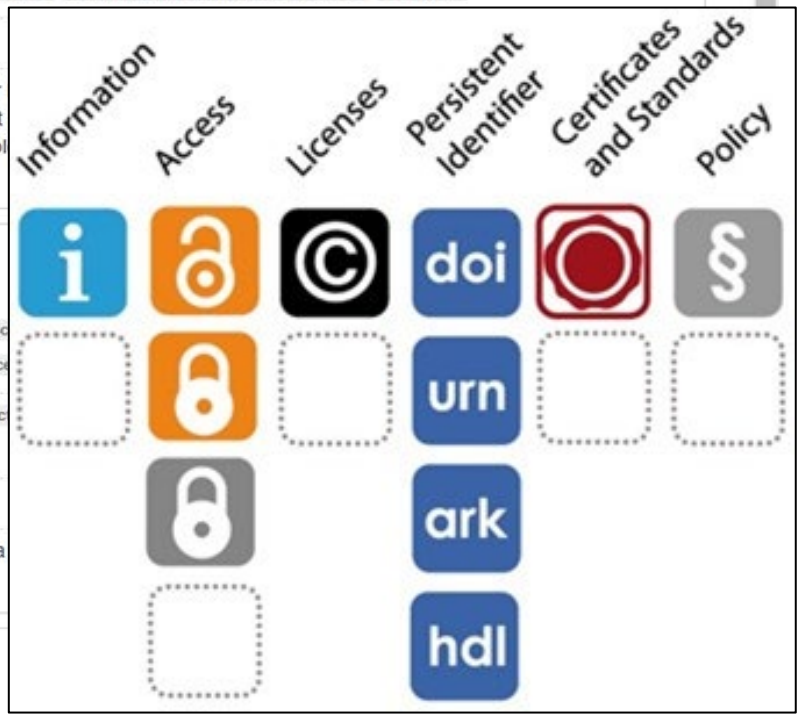
Country: Hong Kong

Provided by the University Libraries, DataHub is the comprehensive institutional repository for research data the University of Hong Kong and their collaborators.

GigaDB

GigaScience Database

https://www.re3data.org/search?contentTypes[]=Standard office documents



Browse by subject

Domains and sub-domains

Graphical | Text

- A. Humanities and Social Sciences
  - a. Humanities
    - I. Ancient Cultures
      - 1. Prehistory
      - 2. Classical Philology
      - 3. Ancient History
      - 4. Classical Archaeology
      - 5. Egyptology and Ancient Near Eastern Studies
    - II. History
      - 1. Medieval History
      - 2. Early Modern History
      - 3. Modern and Current History
      - 4. History of Science
    - III. Fine Arts, Music, Theatre and Media Studies
      - 1. Art History
      - 2. Musicology
      - 3. Theatre and Media Studies
    - IV. Linguistics
      - 1. General and Applied Linguistics
      - 2. Individual Linguistics
      - 3. Typology, Non-European Languages, Historical Linguistics
    - V. Literary Studies
      - 1. Medieval German Literature
      - 2. Modern German Literature
      - 3. European and American Literature
      - 4. General and Comparative Literature and Cultural Studies
    - VI. Non-European Languages and Cultures, Social and Cultural Anthropology, Jewish Stud
      Religious Studies
- D. Engineering Sciences
  - a. Mechanical and industrial Engineering
    - I. Production Technology
      - 1. Metal-Cutting Manufacturing Engineering
      - 2. Primary Shaping and Reshaping Technology
      - 3. Micro-, Precision, Mounting, Joining, Separation Technology
      - 4. Plastics Engineering
      - 5. Production Automation, Factory Operation, Operations Manangement
    - II. Mechanics and Constructive Mechanical Engineering
      - 1. Construction, Machine Elements
      - 2. Mechanics
      - 3. Lightweight Construction, Textile Technology
      - 4. Acoustics
  - b. Thermal Engineering/Process Engineering
    - I. Process Engineering, Technical Chemistry
      - 1. Chemical and Thermal Process Engineering
      - 2. Technical Chemistry
      - 3. Mechanical Process Engineering
      - 4. Biological Process Engineering
    - II. Heat Energy Technology, Thermal Machines, Fluid Mechanics
      - 1. Energy Process Engineering
      - 2. Technical Thermodynamics
      - 3. Fluid Mechanics
      - 4. Hydraulic and Turbo Engines and Piston Engines
  - c. Materials Science and Engineering
    - I. Materials Engineering
      - 1. Metallurgical and Thermal Processes, Thermomechanical Treatment of Materials
      - 2. Sintered Metallic and Ceramic Materials
      - 3. Composite Materials
      - 4. Mechanical Behaviour of Construction Materials
      - 5. Coating and Surface Technology
    - II. Materials Science
      - 1. Thermodynamics and Kinetics of Materials
      - 2. Synthesis and Properties of Functional Materials
      - 3. Microstructural Mechanical Properties of Materials

https://www.re3data.org/browse/by-subject/

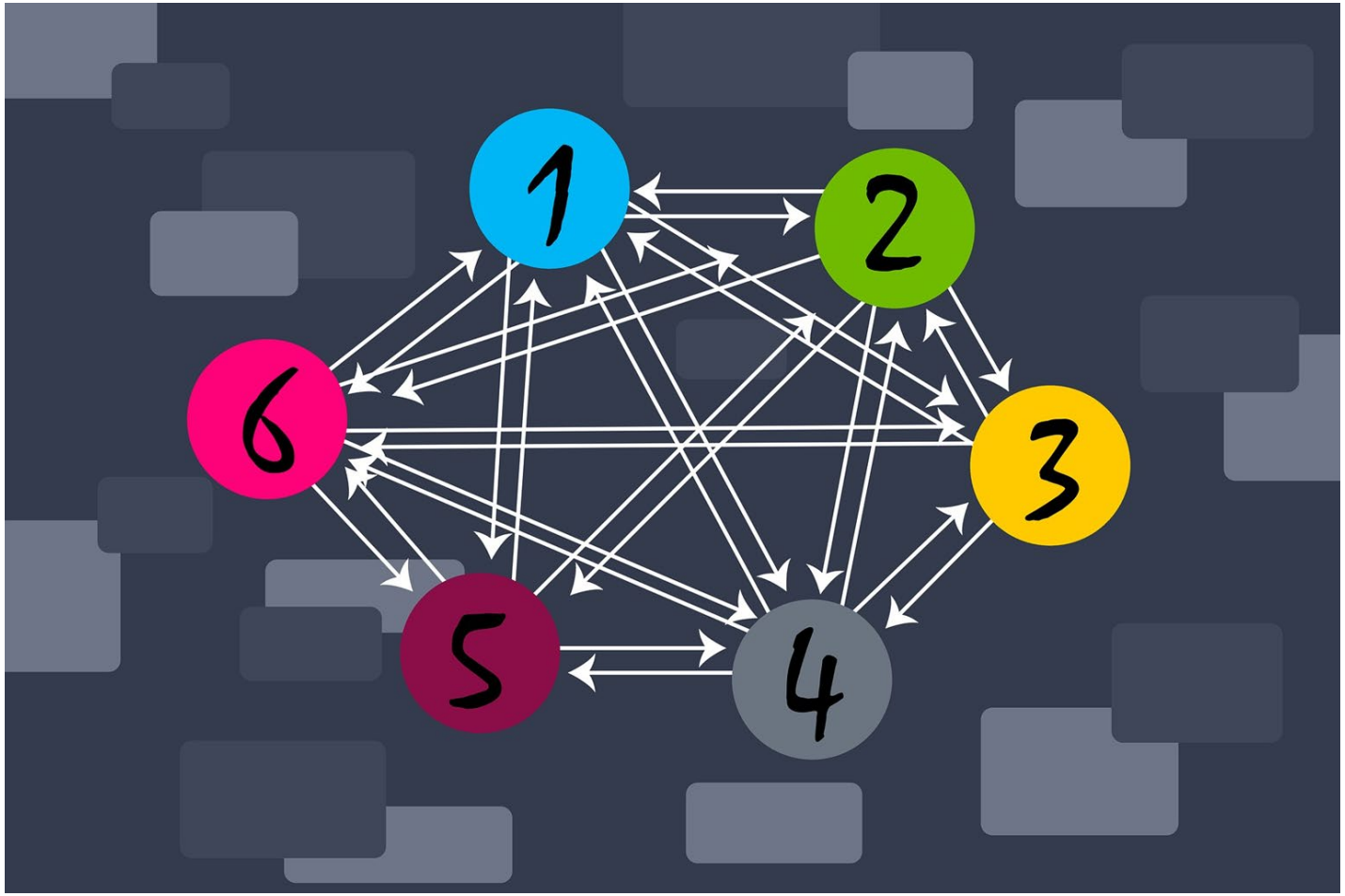
Browse by content type

Archived data
Audiovisual data
Configuration data
Databases
Images
Networkbased data
Plain text
Raw data
Scientific and statistical data formats
Software applications
Source code
Standard office documents
Structured graphics
Structured text
other

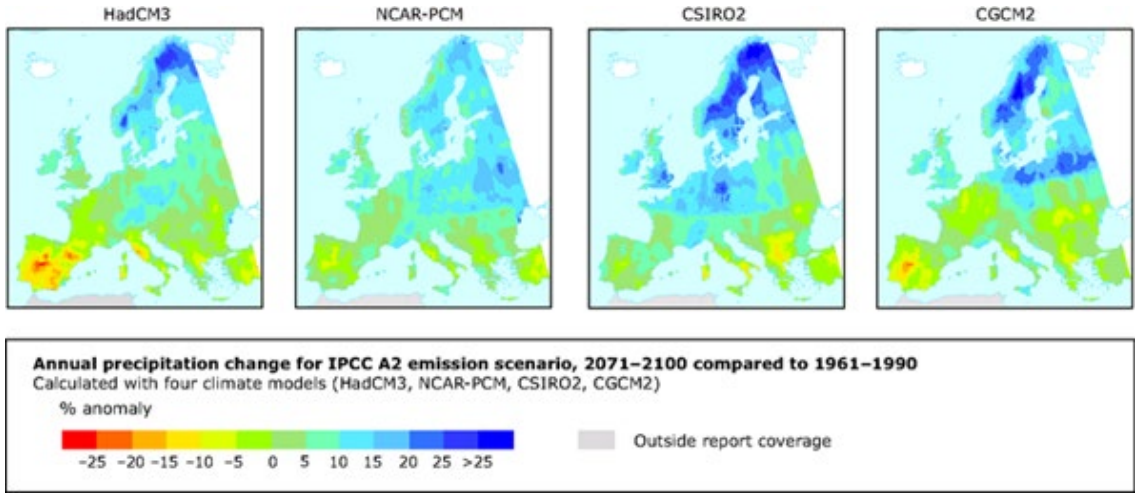
Sometimes it isn't the domain you are interested in, but the kind of data the repository can accept.

https://www.re3data.org/browse/by-content-type/

Consider: how will you link to related outputs (e.g., ORCID, DOIs)



Remember to also consider links to physical data, software and models



https://www.eea.europa.eu/data-and-maps/figures/changes-in-annual-precipitation-for-the-ipcc-a2-scenario-2071-2100-compared-with-1961-1990-for-four-different-climate-models/chapter-3-map-3-1-belgrade-precipitation.eps/image_large

http://www.ukcrcexpmed.org.uk/Coventry_Warwick_CRF/PublishingImages/Tissue%20Bank%201.jpg

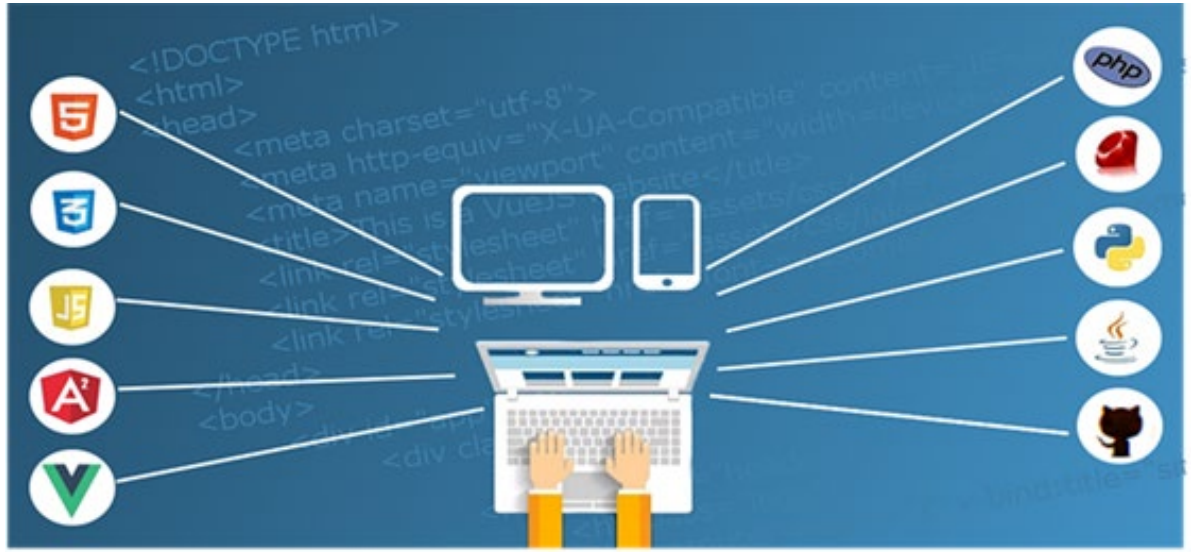
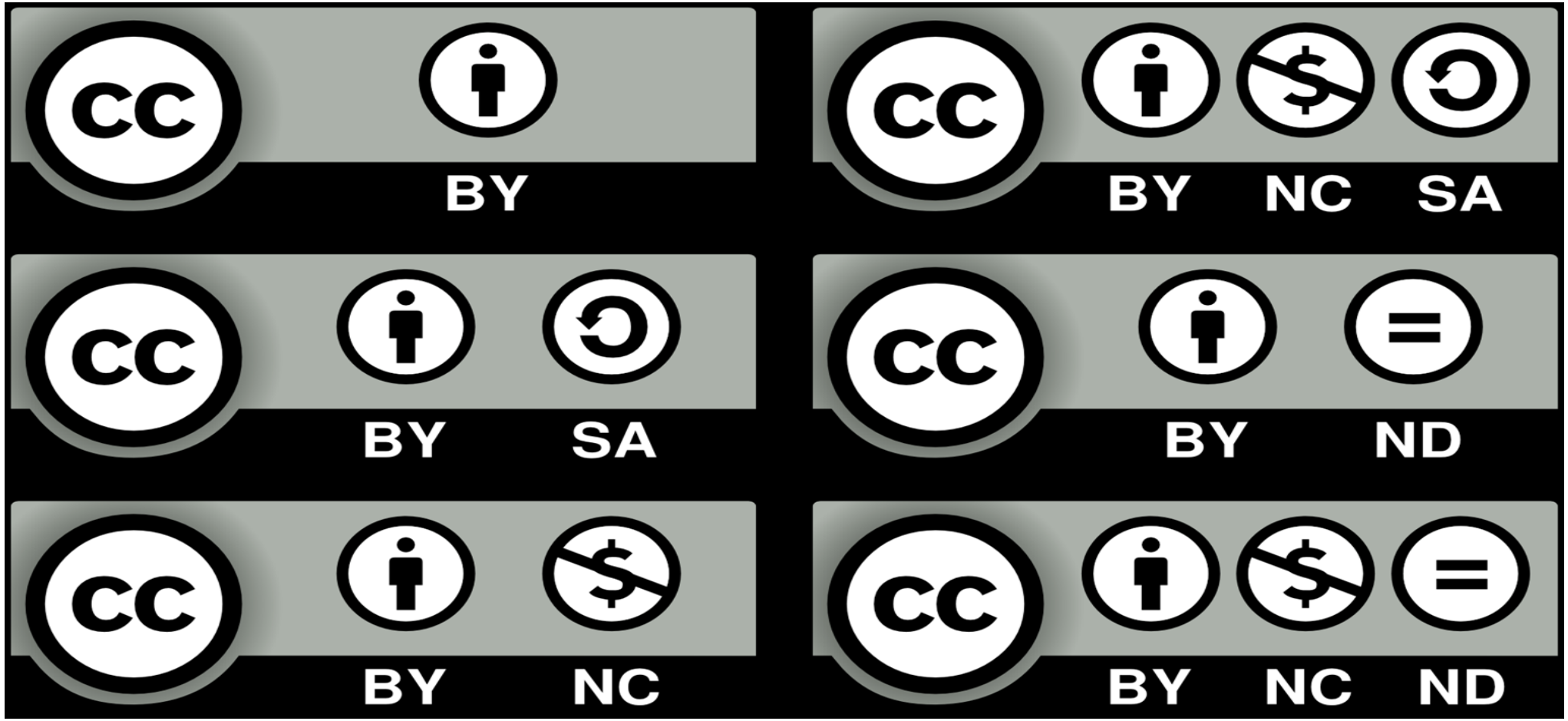


Image by Olalekan Oladipupo from Pixabay

Consider: How will you license your data?



Creative Commons https://creativecommons.org/

Consider: How open can you make your data?

Levels of openness should reflect:

- Funding body requirements
- Personal sensitivities
- Commercial sensitivities

Bear in mind that closed data can still be FAIR data

- Authentication process
- Safe havens or institutional data vault
- Metadata should be FAIR

Consider - Not all data needs to be kept forever

Five steps to follow

1. **Could** this data be re-used
2. **Must** it be kept as evidence or for legal reasons
3. **Should** it be kept for its potential value
4. **Consider costs** – do benefits outweigh cost?
5. **Evaluate criteria** to decide what to keep

To be FAIR, metadata should be kept accessible even if the data no longer exist

5 steps to decide what data to keep

www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

So, now you know a bit about:

- The drive for RDM
- The differences between RDM, open and FAIR data
- What should be addressed when writing your DMP

Now - it’s your turn to review a DMP!

DMP group activity - accessibility and findability

Take 15 minutes to read the three DMP samples and then please go to menti.com and use code 8758 0867

Engineering DMP

2.1. What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?

Selected data will be facilitated by open research data repository, The MOST Wiedzy Open Research Catalog from Gdańsk University of Technology with metadata standards such as DataCite. Metadata description will be stored in JSON-LD format. Contributor will be identified and authorized by an ORCID number.

Natural Science DMP

5. Data Sharing and Long-term Preservation

All research data from this study will be made FAIR and shared as Open Data on Zenodo.

Social Sciences DMP

5. Data Sharing and Long-term Preservation

The data collected over the course of the study will be kept for a period of up to 10 years after the end of the study in a NAS storage facility. Audio recordings will be destroyed once analysis is complete while the pseudo anonymised transcripts will be retained.

Data from this study will not be shared outside of the project consortium.

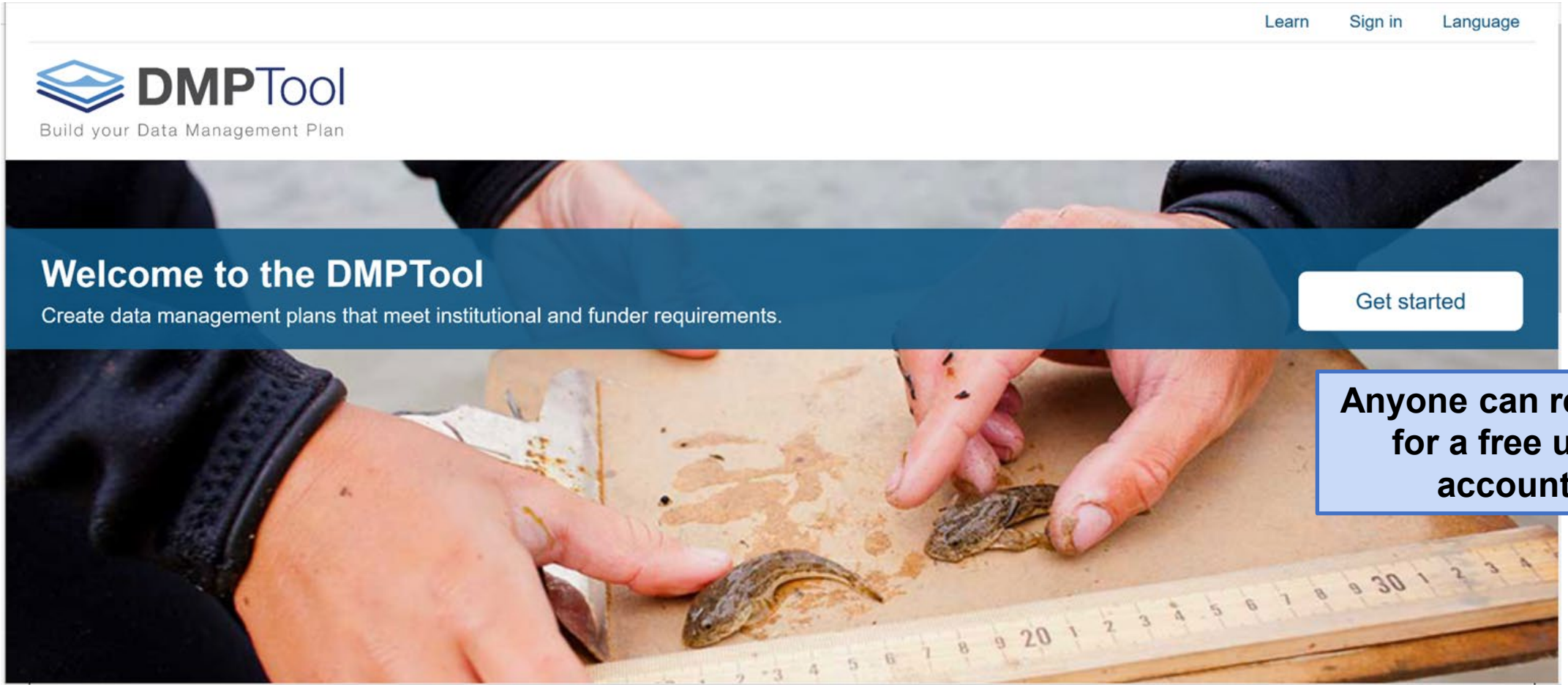
Summary

- DMPs are generally short (2-4 pages)
- Living document - so you might not have all the answers at the start but aim to answer as much as you can
- Be as specific as you can (refer to specific standards where possible)
- Avoid jargon
- Remember to think about what you can share - not always the data but metadata might be shared
- Include some context about the nature of your project but DMP should focus on data
- If your funder offers a DMP template, use it!

Short break – 10 minutes

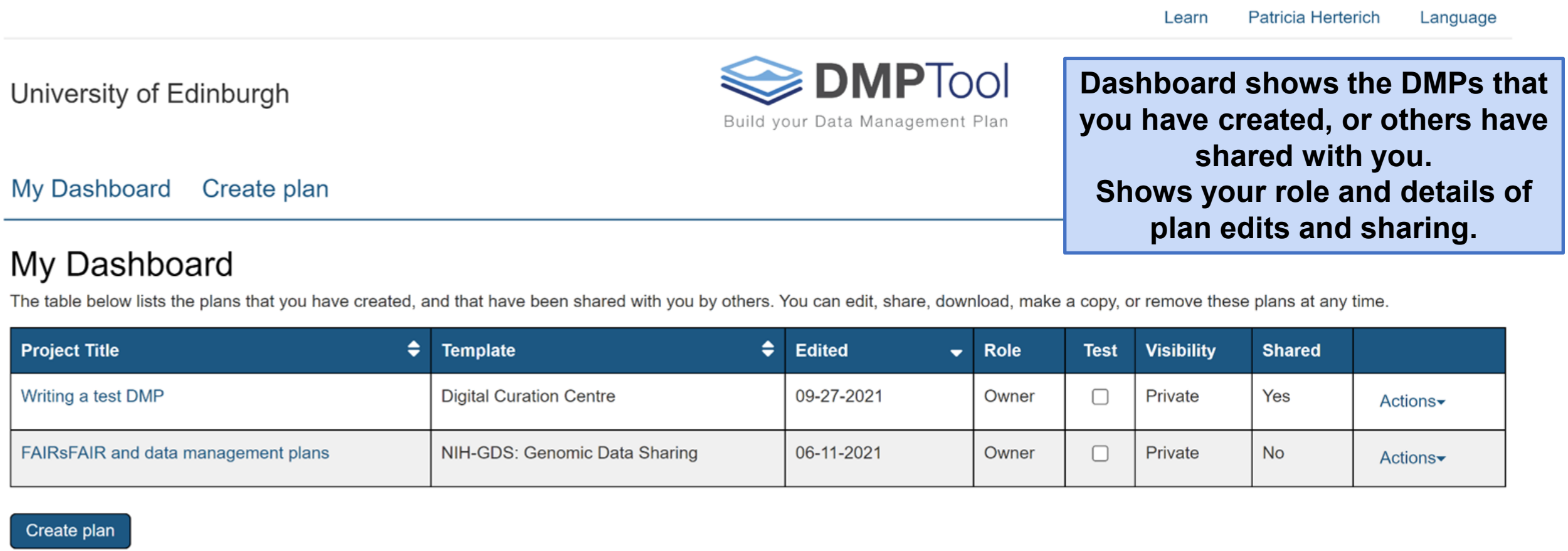
Demonstration of DMPTool

Sign up or sign in https://dmptool.org/



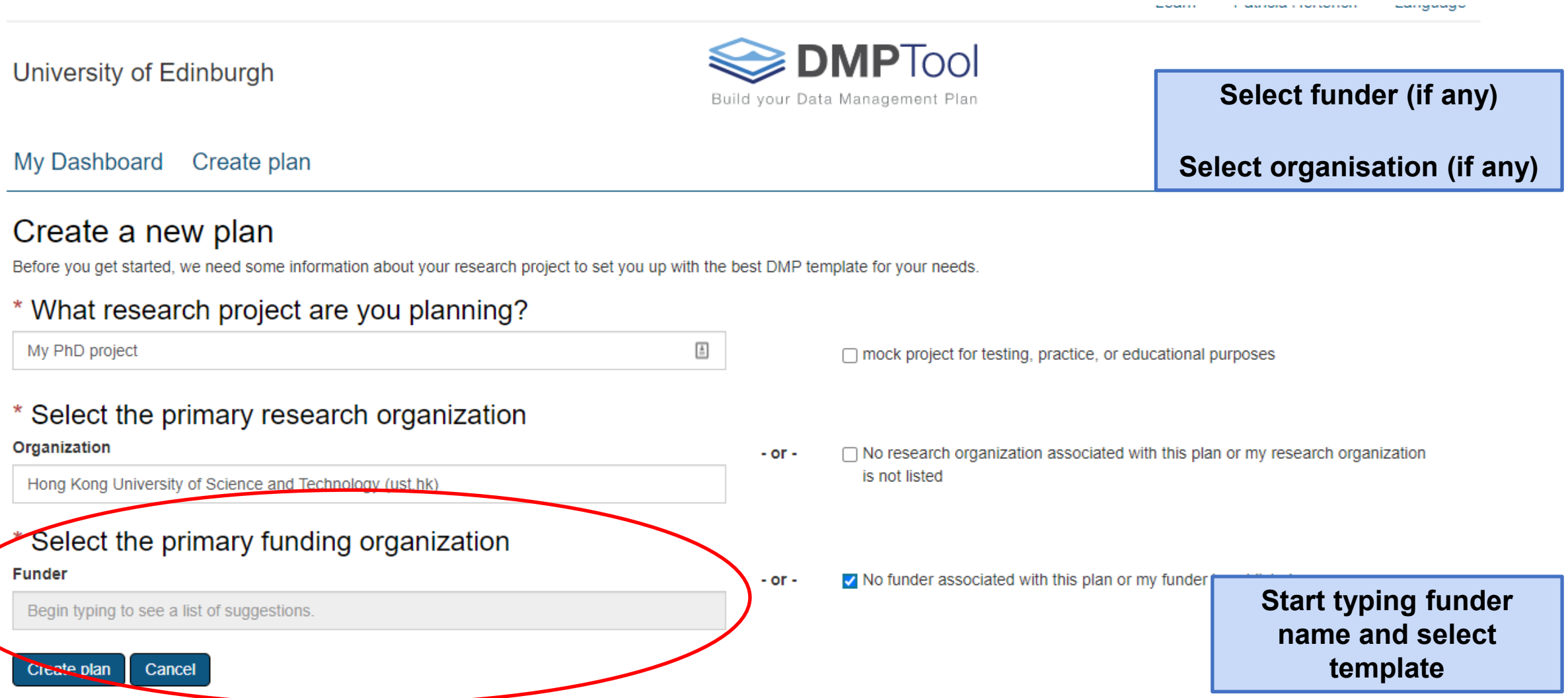
Anyone can register for a free user account.

Returning users see ‘My Dashboard’ page

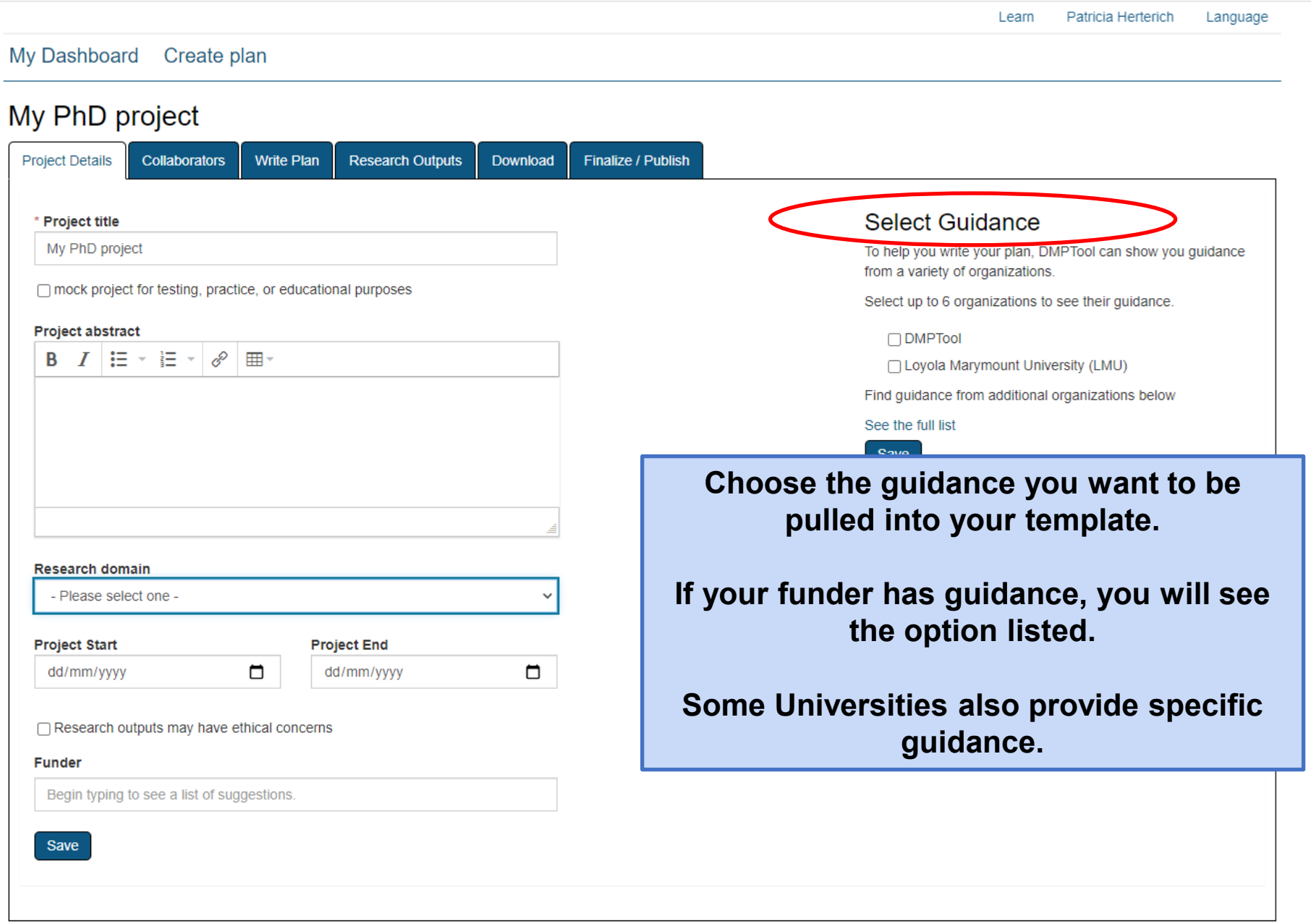


Dashboard shows the DMPs that you have created, or others have shared with you.
Shows your role and details of plan edits and sharing.

Creating a new plan



Project Details

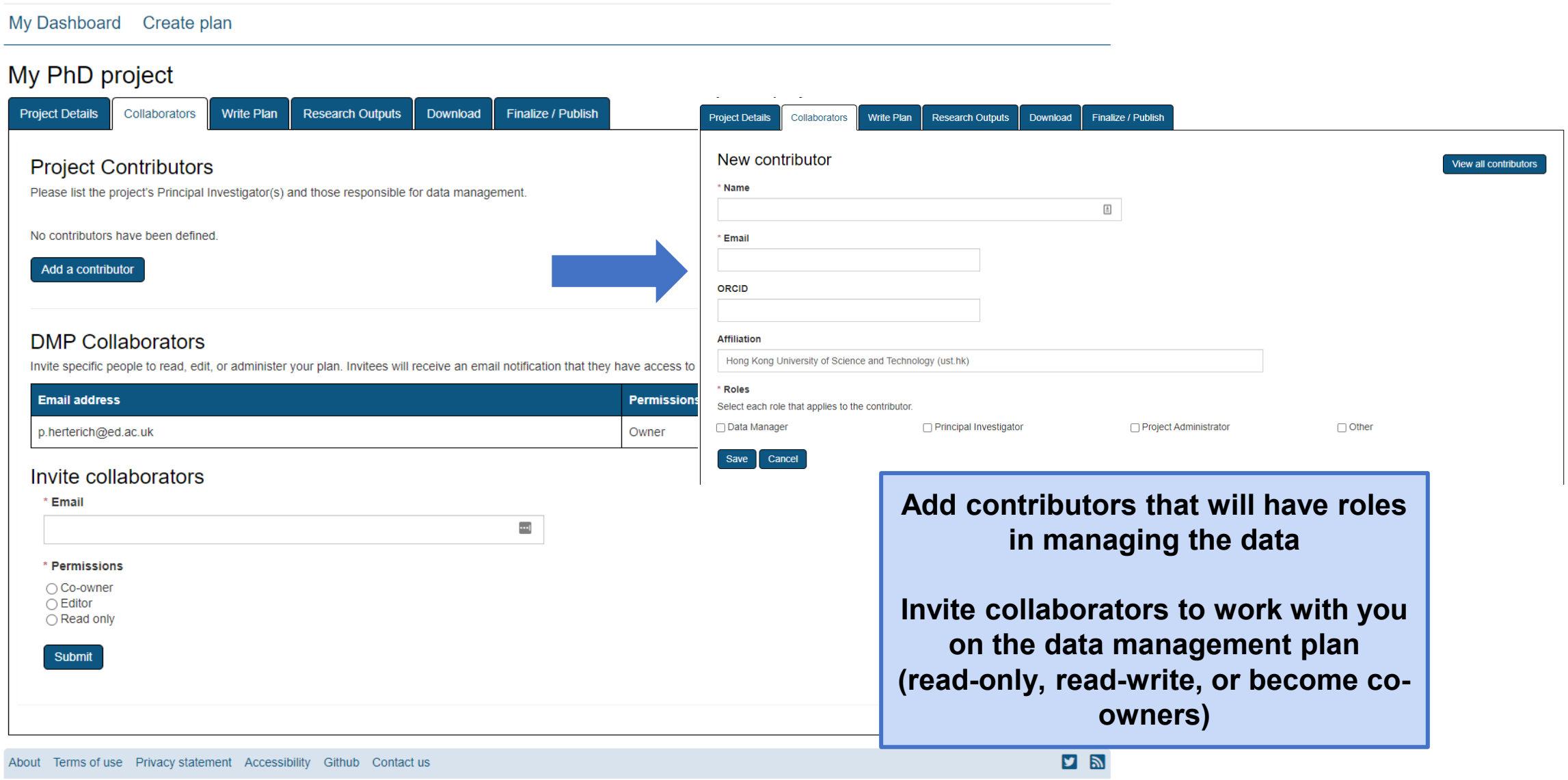


Choose the guidance you want to be pulled into your template.

If your funder has guidance, you will see the option listed.

Some Universities also provide specific guidance.

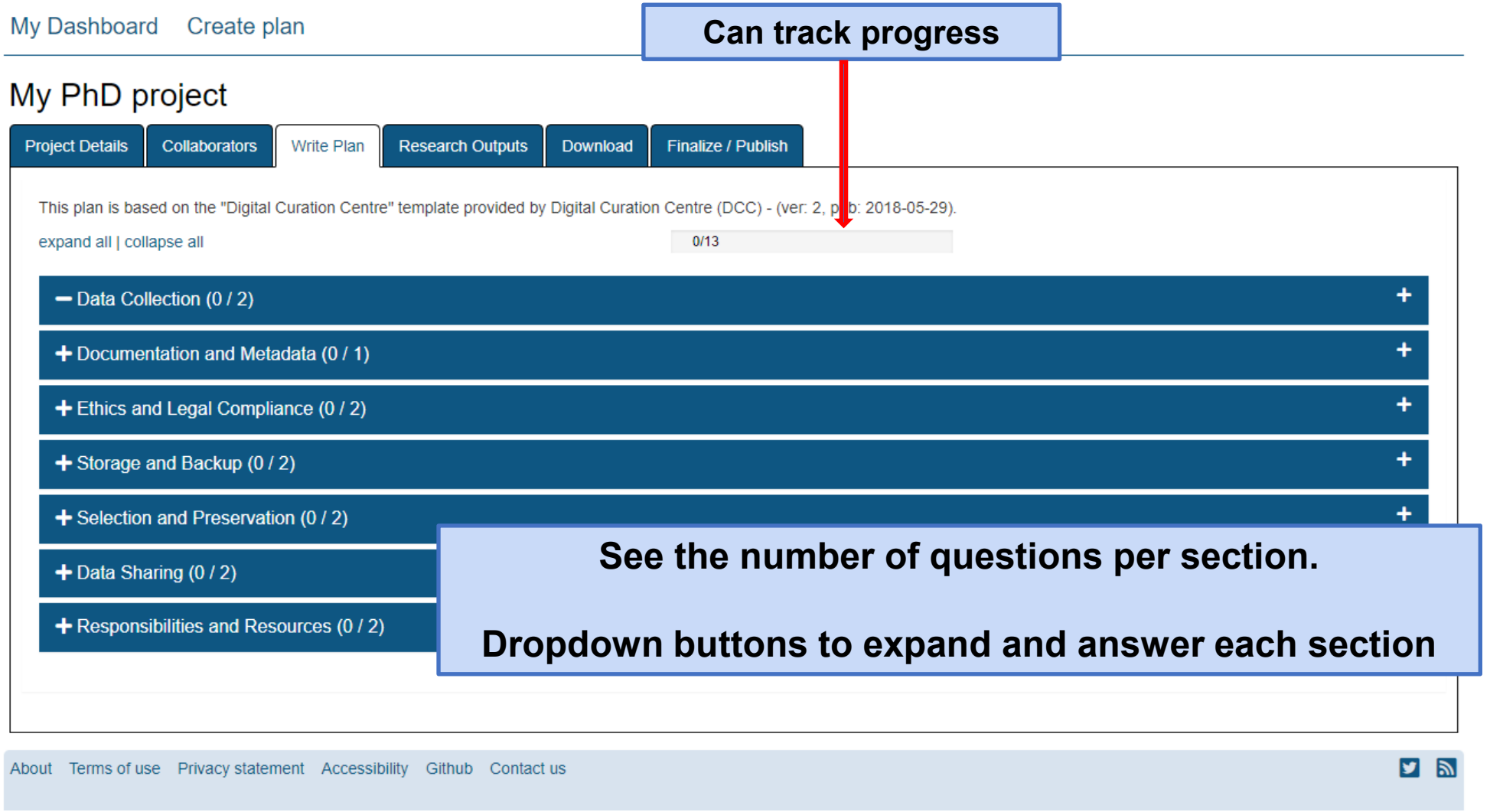
Contributors and collaborators



Add contributors that will have roles in managing the data

Invite collaborators to work with you on the data management plan (read-only, read-write, or become co-owners)

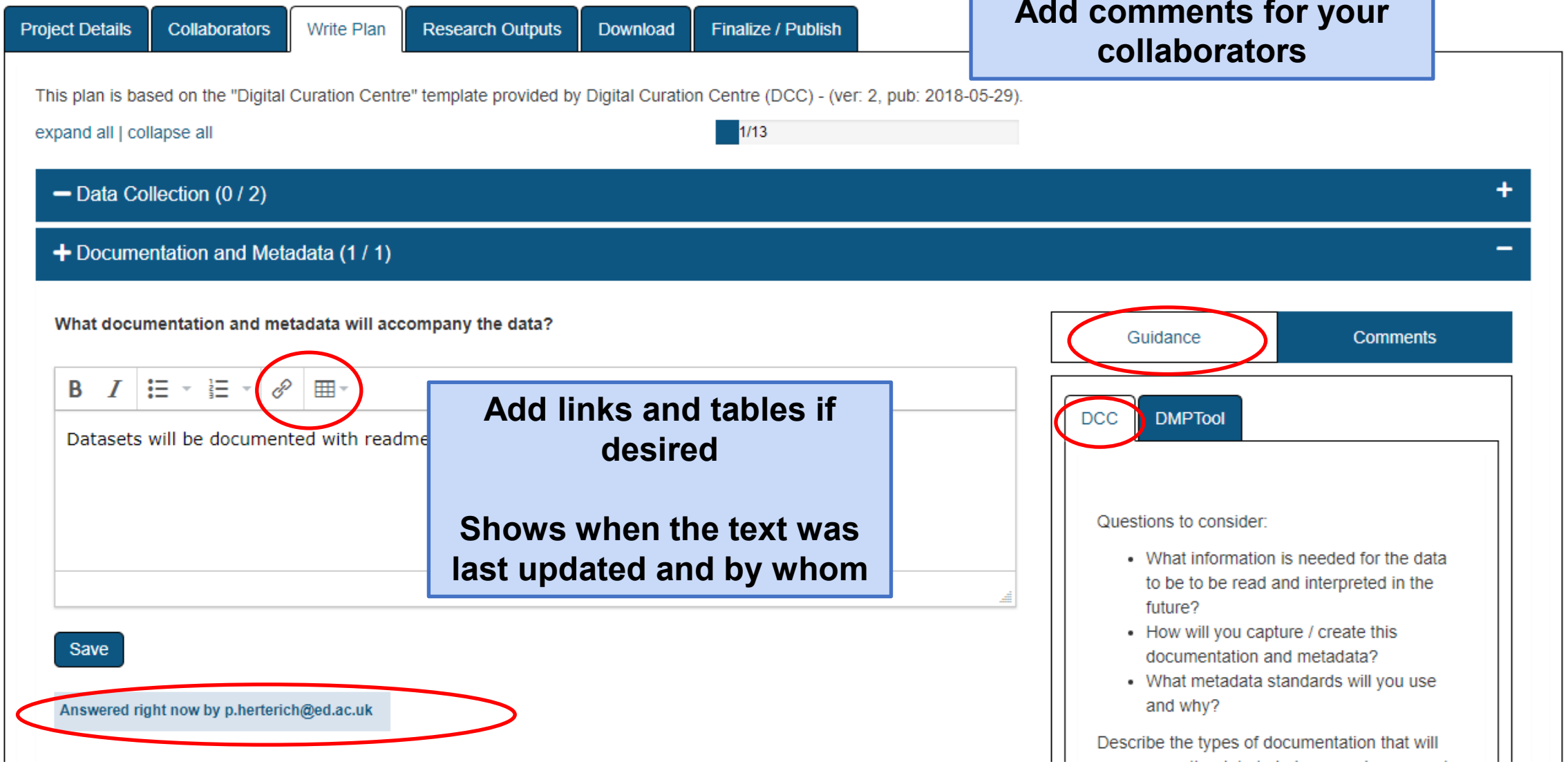
Writing your plan



Answering questions

Toggle between guidance selected

Add comments for your collaborators

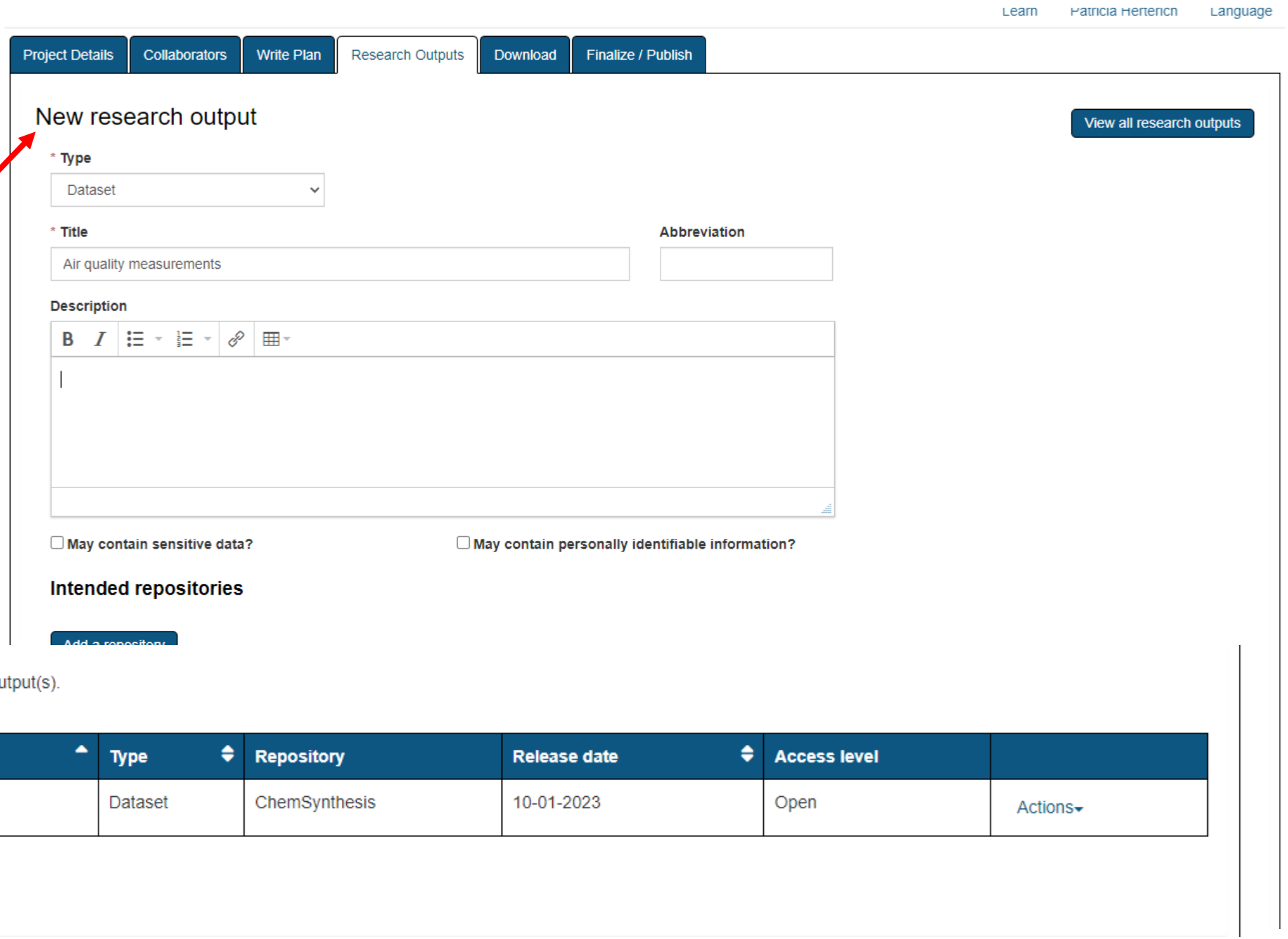


Add links and tables if desired

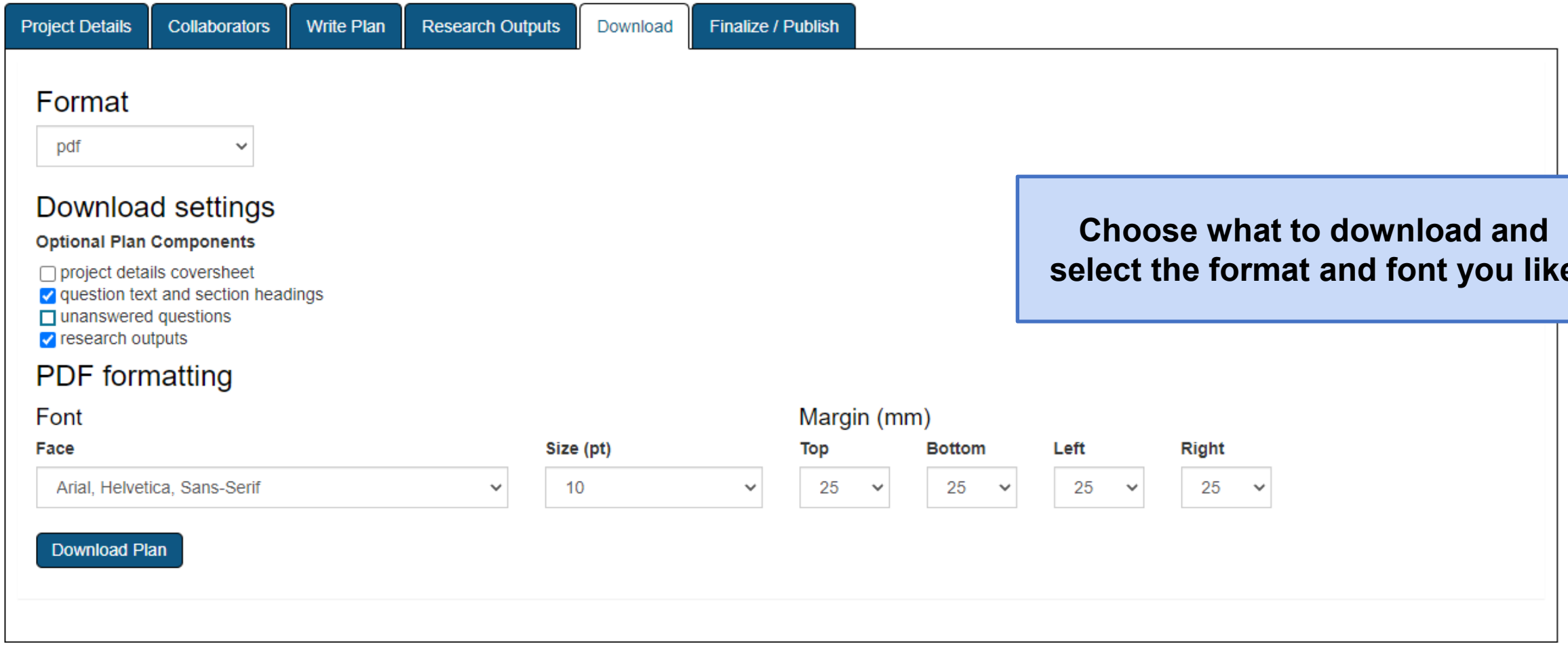
Shows when the text was last updated and by whom

Research outputs

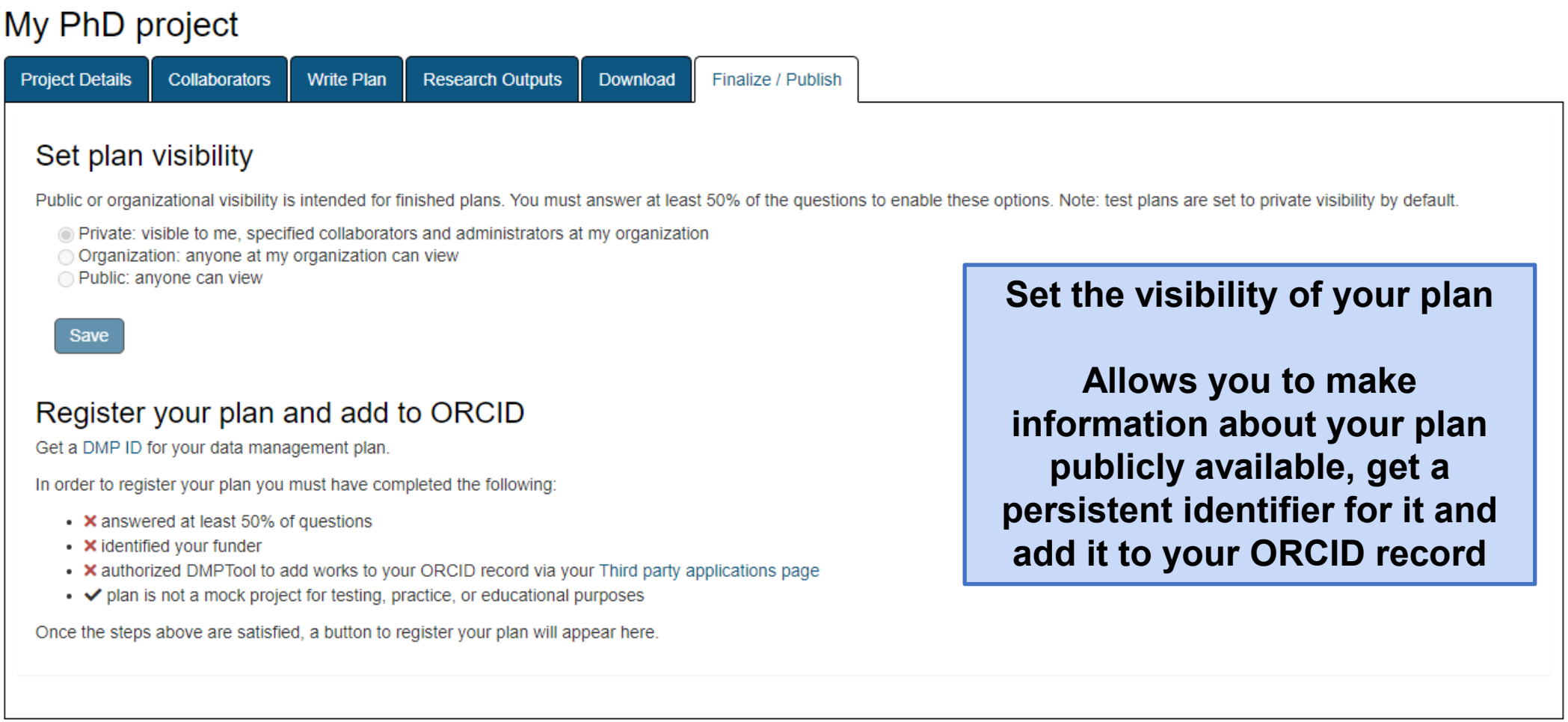
Define repository, licence and metadata details on dataset level in a structured way



Downloading your DMP



Publishing your plan

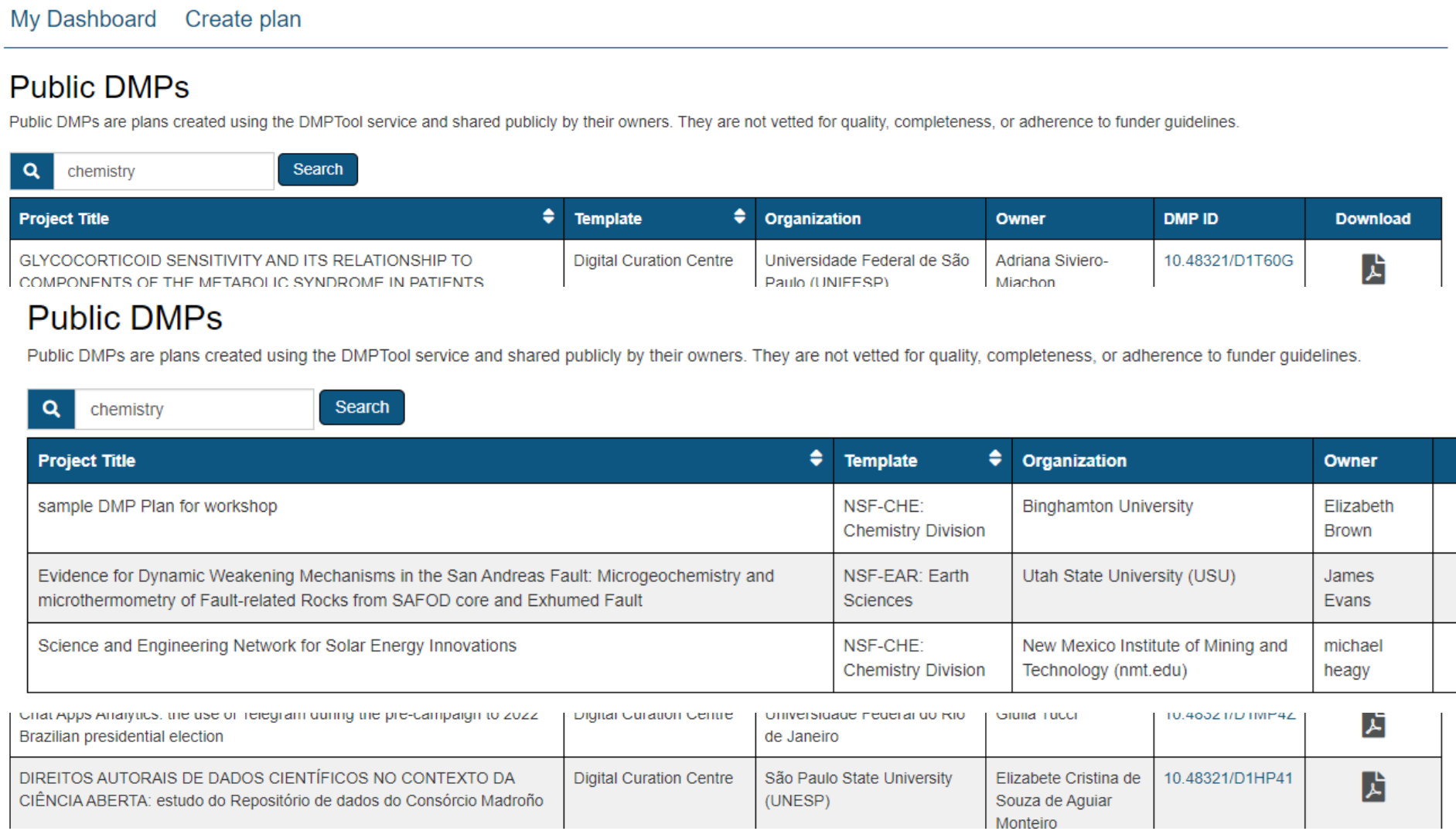


Set the visibility of your plan

Allows you to make information about your plan publicly available, get a persistent identifier for it and add it to your ORCID record

Looking for inspiration?

https://dmptool.org/public_plans



My Dashboard Create plan

Public DMPs

Public DMPs are plans created using the DMPTool service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

chemistry Search

Project Title	Template	Organization	Owner	DMP ID	Download
GLYCOCORTICOID SENSITIVITY AND ITS RELATIONSHIP TO COMPONENTS OF THE METABOLIC SYNDROME IN PATIENTS	Digital Curation Centre	Universidade Federal de São Paulo (UNIFESP)	Adriana Siviero-Miachon	10.48321/D1T60G	

Public DMPs

Public DMPs are plans created using the DMPTool service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

chemistry Search

Project Title	Template	Organization	Owner	Download
sample DMP Plan for workshop	NSF-CHE: Chemistry Division	Binghamton University	Elizabeth Brown	
Evidence for Dynamic Weakening Mechanisms in the San Andreas Fault: Microgeochemistry and microthermometry of Fault-related Rocks from SAFOD core and Exhumed Fault	NSF-EAR: Earth Sciences	Utah State University (USU)	James Evans	
Science and Engineering Network for Solar Energy Innovations	NSF-CHE: Chemistry Division	New Mexico Institute of Mining and Technology (nmt.edu)	michael heagy	

Chat Apps Analytics: the use of Telegram during the pre-campaign to 2022 Brazilian presidential election	Digital Curation Centre	Universidade Federal do Rio de Janeiro	Giulia Tucci	10.48321/D1MP42	
DIREITOS AUTORAIS DE DADOS CIENTÍFICOS NO CONTEXTO DA CIÊNCIA ABERTA: estudo do Repositório de dados do Consórcio Madroño	Digital Curation Centre	São Paulo State University (UNESP)	Elizabete Cristina de Souza de Aguiar Monteiro	10.48321/D1HP41	

DMPTool Hands-on activity

- Please register for a DMPTool account at https://dmptool.org/ (10 minutes)

- Start a new plan. Select HKUST as your organisation and choose ‘No Funder’ to generate your blank plan. Now start working through the DMP themes and individual questions considering one of your own research projects. (20 minutes)
  - Consider which questions you can you answer on your own and where might you need support.

- DMPTool support collaboration during the writing process. Try to share your plan with a colleague. (10 minutes)

How did you get on?

Please go to menti.com and use code 8758 0867

Good practice when creating DMPs

☑ Start early in the research process

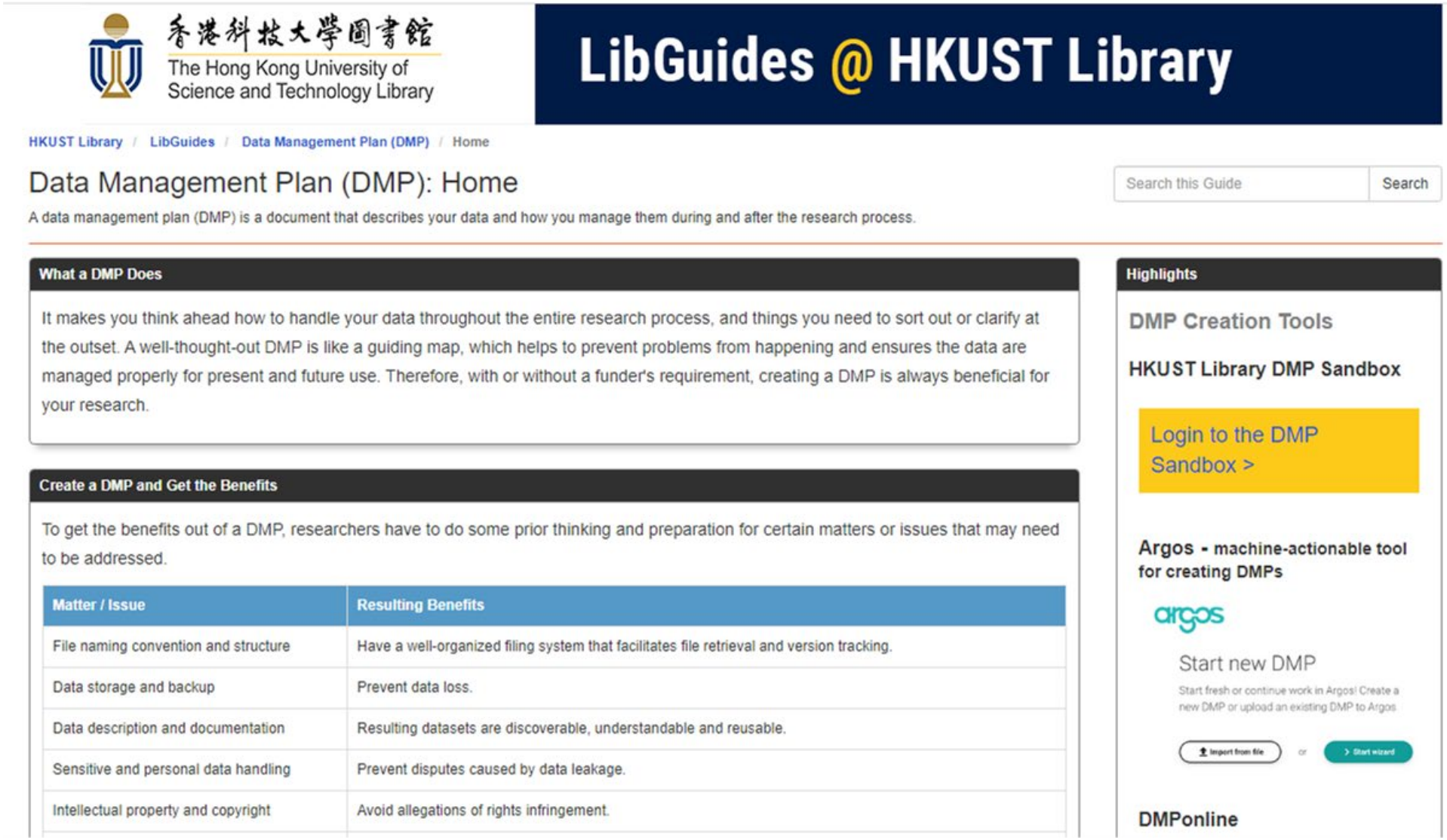
☑ Make use of freely available DMP tools

☑ Get guidance when writing the plan

☑ Be realistic about what you can do (real vs ideal)

☑ Update your DMP as your research progress – things change

Consult your local support team for guidance!



香港科技大學圖書館
The Hong Kong University of Science and Technology Library

LibGuides @ HKUST Library

HKUST Library / LibGuides / Data Management Plan (DMP) / Home

Data Management Plan (DMP): Home

Search this Guide | Search

A data management plan (DMP) is a document that describes your data and how you manage them during and after the research process.

What a DMP Does

It makes you think ahead how to handle your data throughout the entire research process, and things you need to sort out or clarify at the outset. A well-thought-out DMP is like a guiding map, which helps to prevent problems from happening and ensures the data are managed properly for present and future use. Therefore, with or without a funder's requirement, creating a DMP is always beneficial for your research.

Create a DMP and Get the Benefits

To get the benefits out of a DMP, researchers have to do some prior thinking and preparation for certain matters or issues that may need to be addressed.

Matter / Issue	Resulting Benefits
File naming convention and structure	Have a well-organized filing system that facilitates file retrieval and version tracking.
Data storage and backup	Prevent data loss.
Data description and documentation	Resulting datasets are discoverable, understandable and reusable.
Sensitive and personal data handling	Prevent disputes caused by data leakage.
Intellectual property and copyright	Avoid allegations of rights infringement.

Highlights

DMP Creation Tools

HKUST Library DMP Sandbox

Login to the DMP Sandbox >

Argos - machine-actionable tool for creating DMPs

argos

Start new DMP

Start fresh or continue work in Argos! Create a new DMP or upload an existing DMP to Argos

Import from file | or | Start wizard

DMPonline

https://libguides.ust.hk/dmp

Data management planning:

- Helps prevent data loss
- Helps you produce FAIR data and get more impact
- Supports research integrity and enables validation
- Makes it easier to collaborate
- Leads to real world benefits!

Artist's impression of COVID-19 open access data sharing. Credit: Spencer Phillips

Open data sharing accelerates COVID-19 research

19 Oct 2020 - 15:51

Summary

- Open access increases the visibility of research data and information, giving scientists the ability to build upon and react to existing research quickly
- EMBL-EBI launched the European COVID-19 Data Platform to enable rapid access to datasets and results pertaining to the SARS-CoV-2 outbreak
- Open access data sharing has greatly accelerated COVID-19 research and helps further our understanding of the biology, transmission, and spread of the SARS-CoV-2 virus

https://www.ebi.ac.uk/about/news/announcements/open-data-sharing-accelerates-covid-19-research

Thanks very much for your attention and participation!

We are happy to answer any other questions you still have.

Please go back to menti.com and enter code 8758 0867

For more guidance on RDM and DMPs, please see our website

https://dcc.ac.uk/